DISSERTATION

Titel der Dissertation

„Dynamical Models of Biological Networks"

Verfasser

Mag. rer. nat. Lukas Endler

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2012

Studienkennzahl lt. Studienblatt: A 091 419

Dissertationsgebiet lt. Studienblatt: Dr. Studium der Naturwissenschaften Chemie

Betreuerin / Betreuer: Prof. Dr. Peter Schuster

Acknowledgements

There are many people who helped me along the way and to finish my thesis.

I learnt a lot from my supervisors Peter Schuster and Christoph Flamm - from dynamical systems analysis, over evolution to the dark intricacies of Perl programming. Further I want to thank Stefan, Rainer, Stefanie and James for their cooperation, support, and brilliant ideas, Josef Hofbauer for his mathematical expertise, Richard for keeping all systems going, Judith for coffee and an open ear, and Ulli and Raphi for everything.

More thanks go to Bernie, Wash, Svrci, MTW, Hakim, Camille, and to all people at the TBI for creating a great environment, to Bertram, Stephan, Josch, Sandra and the Vogt for providing a rubber chicken to keep me sane, and Nicolas, Nick, Camille, Nico and the rest of Computational Neurobiology group at the EBI for great talks, understanding, and chasing me around the badminton court.

Finally I want to thank Catherine and my parents for their support and love.

Abstract

While experimental research forms the foundation of biological research, mathematical abstractions and models have become essential to understand the observed phenomena underlying complex systems. Particularly in molecular biology mathematical models of reaction and regulatory networks help to extend knowledge of single interactions and entities to a systems-level. Gene regulatory networks are especially good targets for modelling as they are experimentally accessible and easy to manipulate. In this thesis different types of gene regulatory networks are analysed using mathematical models. Further a computational framework of a novel, self-contained *in silico* cell model is described and discussed.

At first the behaviour of two cyclic gene regulatory systems - the classical repressilator and a repressilator with additional auto-activation - are inspected in detail using analytical bifurcation analysis. Both systems are found to exhibit various dynamical behaviours, namely multiple steady states, and limit cycle oscillations. The repressilator with auto-activation additionally can possess stable heteroclinic cycles leading to aperiodic oscillations. Parameter dependencies for the occurrence and stability of equilibria and limit, as well as heteroclinic, cycles are derived for systems with arbitrary gene numbers. To examine the behaviour under random fluctuations, stochastic versions of the systems are created. Using the analytical results sustained oscillations in the stochastic versions are obtained, and the two oscillating systems compared. This shows that the additional auto-activation leads to slightly more uniform oscillations with longer auto-correlation times than in the classical repressilator.

In the second part of the thesis possible implications of gene duplication on a simple gene regulatory system are inspected. A model of a small network formed by GATA-type transcription factors, central in nitrogen catabolite repression in yeast, is created and validated against experimental data to obtain approximate parameter values. Further, topologies of potential gene regulatory networks and modules consisting of GATA-type transcription factors in other fungi are derived using sequence-based approaches and compared. A model of a single autoactivating GATA type transcription factor is used to study the effects of gene duplication. The model predicts profound gene dosage effects. Exemplary mutations relieving the dosage effects are studied, leading to various motifs commonly

found in gene regulatory networks, such as cascades and feed forward loops. One mutation, loss of trans-activation in one paralogue, potentially even leads to a tunable oscillator.

The last part describes `MiniCellSim`, a model of a self-contained *in silico* cell. In this framework a dynamical system describing a protocell with a gene regulatory network, a simple metabolism, and a cell membrane is derived from a string representing a genome. All the relevant parameters required to compute the time evolution of the dynamical system are calculated from within the model, allowing the system to be used in studies of evolution of gene regulatory and metabolic networks.

Parts of chapter 2 and chapter 4 have been published in journals previously [123, 290], and the articles have been used as a base for writing the respective chapters. All program and model files used for the thesis and the source code of `MiniCellSim` are available from the author and on the author's web page under `http://www.tbi.univie.ac.at/~luen/Diss/`.

Zusammenfassung

Mathematische Modelle sind wertvolle Werkzeuge um die komplexen Netzwerke zu verstehen, die biologischen Systemen zu Grunde liegen. Besonders in der Molekularbiologie sind mathematische Modelle von regulatorischen und metabolischen Netzwerken essentiell, um von einer Betrachtung isolierter Komponenten und Interaktionen zu einer systemischen Betrachtungsweise zu kommen. Genregulatorische Systeme eignen sich besonders gut zur Modellierung, da sie experimentell leicht zugänglich und manipulierbar sind. In dieser Arbeit werden verschiedene genregulatorische Netzwerke unter Zuhilfenahme von mathematischen Modellen analysiert. Weiteres wird ein Modell einer *in silico* Zelle vorgestellt und diskutiert.

Zunächst werden zwei zyklische genregulatorische Netzwerke - der klassische Repressilator und ein Repressilator mit zusätzlicher Autoaktivierung - im Detail mit analytischen Methoden untersucht. Beide Systeme können sowohl verschiedene Anzahlen von stationären Zustaenden, als auch Grenzzyklen mit periodischen Oszillationen zeigen. Der Repressilator mit Autoaktivierung kann weiters stabile heterokline Zyklen aufweisen, was zu Oszillationen mit anwachsender Periode führt. Parameterabhängigkeiten von und Kriterien für Stabilitiät und Existenz von stationäeren Zuständen und Grenz- sowie heteroklinen Zyklen werden im Detail für Systeme mit beliebiger Anzahl von Genen abgeleitet. Um den Einfluß zufällig schwankender Molekülzahlen auf die Dynamik der beiden Systeme zu untersuchen, werden stochastische Modelle erstellt und die beiden oszillierenden Systeme verglichen. Dabei zeigt sich, daß die zusätzliche Autoaktivierung zu einheitlicheren Oszillationen mit längeren Autokorrelationszeiten als beim klassischen Repressilator führt.

Weiteres werden mögliche Auswirkungen von Genduplikationen auf ein einfaches genregulatorisches Netzwerk untersucht. Dazu wird zunächst ein kleines Netzwerk von GATA Transkriptionsfaktoren, das eine zentrale Rolle in der Regulation des Stickstoffmetabolismus in Hefe spielt, modelliert und das Modell mit experimentellen Daten verglichen, um Parameterregionen einschränken zu können. Außerdem werden potentielle Topologien genregulatorischer Netzwerke von GATA Transkriptionsfaktoren in verwandten Fungi mittels sequenzbasierender Methoden gesucht und verglichen. Ein Modell eines einfachen, autoaktivierenden GATA Transkriptionsfaktors wird verwendet, um die Auswirkungen

von Genduplikation zu untersuchen. Dabei stellte sich heraus, daß die Autoaktivierung zu einem starken Gendosiseffekt führen kann. Möglicher Mutationen, die diesen Effekt abschwächen können, führen zu häufig gefunden genregulatorischen Motiven - zum Beispiel regulatorischen Kaskaden, oder Feed-Forward Schleifen. Eine Mutation - der Verlust der transkriptionellen Aktivierung in einem der paralogen Gene - kann sogar ein stabil oszillierendes System zur Folge haben.

Im letzten Teil der Arbeit wird `MiniCellSim` vorgestellt, ein Modell einer selbständigen *in silico* Zelle. Es erlaubt ein dynamisches System, das eine Protozelle mit einem genregulatorischen Netzwerk, einem einfachen Metabolismus und einer Zellmembran beschreibt, aus einer Sequenz abzuleiten. Nachdem alle Parameter, die zur Berechnung des dynamischen Systems benötigt werden, ohne zusätzliche Eingabe nur aus der Sequenzinformation abgeleitet werden, kann das Modell für Studien zur Evolution von genregulatorischen Netzwerken verwendet werden.

Teile der Kapitel 2 und 4 wurden als Artikel veröffentlicht [123, 290], und diese Artikel als Grundlage für die jeweiligen Kapitel herangezogen. Alle Programm- und Modelldateien sowie der Quellcode von `MiniCellSim` sind vom Autor direkt auf Anfrage oder auf der Webpage des Autors unter `http://www.tbi.univie.ac.at/~luen/Diss/` erhältlich.

Contents

List of Abbreviations

AMP   Adenosine Monophosphate

AreA/B   GATA type transcription factors involved in nitrogen reulation in *Aspergilli* [439]

ARN   Artificial Regulatory Network

bp   base pair

cAMP   cyclic Adenosine Monophosphate

CAP   Catabolite Activator Protein

CellML   Cellular Markup Language

cI   Transcription factor in bacteriophage λ able to act both as a repressor and activator and involved in maintaining lysogeny [188]

CME   Chemical Master Equation

DAL80   Repressing GATA-type transcription factor involved in yeast NCR [84]

DNA   Deoxyribonucleic Acid

DOR   Dense Overlapping Regulon

EBA   Energy Balance Analysis

FBA   Flux Balance Analysis

FFL   Feed Forward Loop

FMN   Flavin Mononucleotide

GAL   The Leloir or galactose utilisation pathway in yeast

GAT1   Activatory GATA-type transcription factor involved in yeast NCR [84]

GATA   a family of transcription factors commonly binding DNA motifs containing the sequence `GATA`

GLN3   Activatory GATA-type transcription factor involved in yeast NCR [84]

Gln   Glutamine, or some other readily available nitrogen source

GZF3   Repressing GATA-type transcription factor involved in yeast NCR [84]

IHF   Integration Host Factor protein

$\mathbf{J}(\mathbf{p}, \mathbf{x}(t))$ Jacobian matrix of an ODE system with parameters $\mathbf{p}$ and time dependent variables $\mathbf{x}(t)$

λ	Eigenvalues
LacI	lac repressor protein of *E. coli*
MAPK	Mitogen-Activated Protein Kinase
MFE	Minimal Free Energy
miRNA	micro RNA
MoMA	Method of Minimization of metabolic Adjustment
MPF	Maturation Promoting Factor
mRNA	messenger RNA
N	stoichiometric matrix
Nanog	Transcription factor involved in regulation of embryonic stem cell survival and pluripotency in mice and humans [78]
NCR	Nitrogen Catabolite Repression
O	Origin of a coordinate system.
OCT4	OCtamer-binding Transcription factor 4: Transcription factor involved in regulation of embryonic stem cell pluripotency in mice and humans [78]
ODE	Ordinary Differential Equation
ORF	Open Reading frame of a gene - the protein coding part of the sequence.
PDE	Partial Differential Equation
RBN	Random Boolean Networks
RepAuto	A repressilator with autoactivation. A cyclic gene regulatory network in which each gene product activates its own expression and represses expression of its successor in the cycle. Repression is modelled to be tight.
RepLeaky	The classical repressilator. A cyclic gene regulatory network in which each gene product represses expression of its successor in the cycle. Repressor binding is cooperative and repression leaky.
RNA	Ribonucleic acid
RNAP	RNA polymerase
ROOM	Regulatory On/Off Minimization
$\mathcal{S}(\mathbf{x})$	Support of point $\mathbf{x}$. The list of the indices of all nonzero coordinates.
SBML	Systems Biology Markup Language
SIM	Single Input Motif

SOX2	SRY-box 2: HMG-box containing transcription factor involved in regulation of embryonic stem cell pluripotency in mice and humans [78]
TAD	TransActivatory Domain of an transcriptional activator
TF	Transcription Factor
TOR	Target of Rapamycin. Serine/threonine kinases that are inhibited by Rapamycin.
tRNA	transfer RNA
trp operon	A group of genes in *E. coli* encoding proteins involved in tryptophan metabolism transcribed on one mRNA
u	unit Eigenvector
Ure2p	Prion-like yeast protein binding to and sequestering Gln3p in the nucleus at high levels of nitrogen availability [84, 263]
URR	Upstream Regulating Region - DNA region located 5' of the promotor region of genes binding transcription factors.
WGD	Whole Genome Duplication events.
XML	eXtended Markup Language
T_h	Lymphocytes expressing the CD4 antigen and directing and activating other immune cells

List of Figures

Figures

List of Tables

1 Introduction

1.1 Modelling of Biological Networks

Molecular biological research for a long time has been focused on the mapping of functions to single components and on characterising the interactions between them. While this approach has proved very successful, its explanatory and predictive power is limited by the complexity of biological systems. Most biological phenomena cannot be attributed simply to single molecules or genes, but instead they arise from the interplay and interaction of many agents. Such complex systems cannot just be understood by looking at isolated elements, but rather need consideration of all parts and interactions together. A quite original view, and critique of the classical reductionist approach employed in biology is given by Lazebnik (2002) [238].

In the last two decades, research in molecular biology has shifted away from looking at components in isolation to studying whole systems [63, 225]. Technological advances in molecular biological research, especially the advent of high throughput methods, allow the measurement of a multitude of components simultaneously and can determine the status of a whole biological system at defined time points.

Making sense of such a vast amount of data is by far not a trivial problem. Even the behaviour of a comparably small and well characterised system, such as the lysogeny-lysis switch of the bacteriophage λ, encompassing only six genes, can be far too complex to be intuitively understood or predicted by just looking at the components [13, 276]. For the prokaryote *E. coli* on the other hand, Regulon DB[1] [135] lists 4622 genes regulated by more than 2700 interactions (Release 7.2, May 2011), and a recent metabolic reconstruction contains 1387 metabolic reactions [117]. To understand such systems mathematical models need to be constructed and to be tested against existing knowledge. Once validated such models can be used to find novel behaviours, analysed to identify essential sub-components, optimise or alter a system, and much more.

While mathematical modelling of biological systems and processes has a long history, until the widespread use of computers it was mainly restricted to the study

---

[1] `http://regulondb.ccg.unam.mx/`

small systems in ecology or the kinetics of single enzymes. Since the ground-breaking work of Hodgkin and Huxley [189] on action potential formation in squid axons and of Chance [74] on the kinetics of catalase using analog computers to obtain numerical solutions, the number and complexity of mathematical models in biology increased vastly.

Creating a model is quite a complex process, and its formulation not only depends on the biological system to be investigated, but also on the final function of the model. This, together with the amount of knowledge available, often determines the detail and granularity necessary (see Fig. 1).
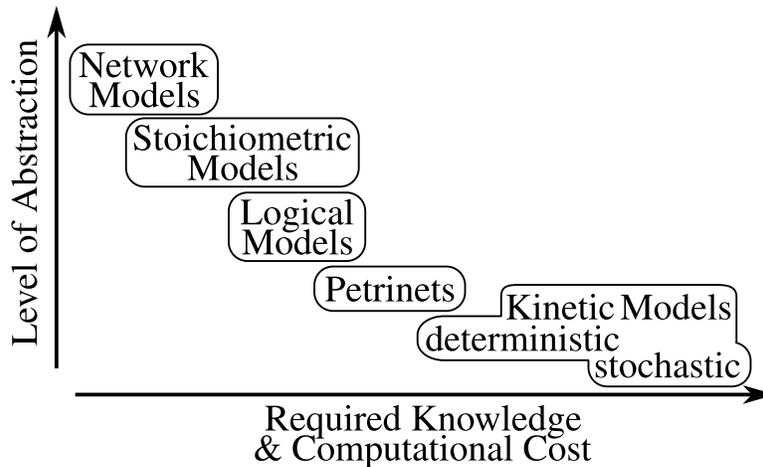
Figure 1: Comparison of detail and computational costs of some common modelling approaches. Network and topology based approaches can often readily be employed to gain general insights, but do not allow for detailed description of complex dynamics and behaviours. Kinetic models on the other hand require much more detailed information on interactions and parameters, but can give mechanistic explanations. Due to the higher computational cost of kinetic models, more abstract representations, such as logical models, are sometimes used quickly screen whether network topologies fit or contradict experimental data.

Biological reaction networks have many different logical layers, from a list of components, reactions and stoichiometries, over network of regulatory interactions, to a detailed quantitative, kinetic description with mathematical expressions for reaction velocities and defined parameter values [63]. Which layers need to be combined depends both on the knowledge available and the intended use of a model.

1.2 Network Models

One of the most basic ways of describing a biological system is as a network of components and their interactions. Interactions can encompass quite different phenomena, from direct physical interaction, over participation in a reaction, to regulations of an activity or state. While this of course cannot fully describe the dynamics of a system, it still offers valuable insights.

One way of formally representing such networks is as graphs, that is lists of nodes, representing the components, connected to each other by edges standing for their interactions. The edges can be undirected, such as for protein-protein interaction networks, or directed, as in genetic regulatory and signalling networks. To indicate the type of interaction, for example activating or inhibiting regulatory interactions, the edges can also be labelled.

The formal description of a network as a graph has permitted the use of standard tools and measures of graph theory to investigate the features of diverse biological networks [3]. Amongst the most basic quantifiable characteristics of a network are the degree or connectivity, the shortest path lengths, and the clustering coefficients. The degree, k, is simply the number of connections a node has, and its distribution $P(k)$, that is the probability of a node having k connections, allows the global classification of a network, independent of its size. One interesting finding which came about by using graph based analysis was that relatively diverse types of biological networks - for example protein-protein interaction, gene regulatory and metabolic networks [19, 29, 207, 333, 342] - share a common feature. In many of these networks, the degree distribution is not just a binomial distribution, as would be expected for random graphs, but instead follows a power law, $P(k) \propto k^{-\gamma}$. This means that while most nodes are only loosely connected, there exist a few highly connected hubs. Such a network architecture is called scale-free, and is also found in other systems, such as the internet, or some social networks, giving rise to the "Small World" phenomenon. For at least a certain range of γ values , such networks possess properties favourable to biological systems, that can lead to smaller average paths lengths and increased robustness against node failures [29].

1.2.1 Stoichiometric Models

One way of describing a biochemical system is as a list of reactions with stoichiometric coefficients for the reactants and products. It can be represented as a directed hypergraph, or, equivalently, as a directed bipartite graph. In a hypergraph, different from simple graphs, each edge can be connected to more than one vertex. By contrast in a bipartite graph there exist two kinds of vertices, for example molecular species and reactions. Each node of a bipartite graph can only be connected to a node of the different type, so molecular species would be bridged by reactions. The information needed to build up such a graph can be concisely stored in the stoichiometric matrix, N. N consists of a row for each of the n metabolites and a column for each of the m reactions of a biochemical system. Its entries are the stoichiometric coefficients ν_{ni} of each metabolite m_n in each reaction r_i.

The stoichiometric matrix contains a wealth of information, especially for the analysis of metabolic systems. By combining knowledge of single reactions and their catalysing enzymes with genomic data, whole genome metabolic reconstructions for a plethora of organisms have been created, from various prokaryotes over yeast to humans [101, 106, 181, 303, 332]. Adding gene expression and proteomics data to this further allows the creation of cell specific reconstructions, for example for human liver cells, macrophages and erythrocytes [53, 54, 146].

In general the stoichiometric matrix can be analysed in two ways, structurally on its own, or, together with certain constrains and optimisation criteria, to find particular solutions to flux distributions. General structural analysis can show futile cycles, connectivity and robustness of the network. The left null spaces of N give conservation relations between metabolites and moieties pools, while the base vectors of the right null space constitute the base for all steady state pathways of the system [115]. The linearly independent column and row vectors of N elucidate potential dynamic behaviours of the system. The row space contains all possible vectors of reaction rates of the system, that do not lead to steady states. The column space on the other hand contains all possible concerted concentration changes of the system. This information can be used to derive – via the mass balances – the thermodynamic driving forces [317].

Structural analysis of N can also be used to decompose a reaction network into

alternative sub-pathways. These can either be balanced in themselves or under certain constraints on import and export fluxes. One type of such sub-pathways, the Elementary Flux Modes, and a subset of them called Extreme Pathways, can be used to find cycles and optimal paths, or to analyse growth capabilities of strains [206, 393]. The number of different sub-pathways leading to certain metabolites give insight to robustness to reaction deletion, and flexibility of a reaction network. They can also be used to find minimal cut-sets, the minimal amount of reactions, or enzymes, that need to be perturbed, so that a certain objective - say growth or a specific metabolite conversion - can no longer occur in steady state [226]. An excellent review of different structural methods used for pathway analysis is given in Papin *et al.* 2004 [206].

Another approach to exploring a reaction networks capabilities using purely stoichiometric information, is combining stoichimetric models with additional biophysical and experimental constraints, such as upper and lower bounds to fluxes and thermodynamic and energetic criteria. Flux Balance Analysis (FBA) combines this approach with an optimization criteria to retrieve particular steady state flux solutions [317, 425] and it has been widely used to analyse the metabolism of various prokaryotic and eukaryotic organisms [252, 340]. Various variants of FBA have been derived, such as MoMA (Method of Minimization of metabolic Adjustment) [374] and ROOM (Regulatory On/Off Minimization) [380] to analyse the impact of gene knockouts on a metabolic network, and algorithms to couple stoichiometric models to transcriptional regulation (rFBA) [86]. Energy Balance Analysis (EBA) is a variant of FBA that uses detailed energetic constraints [32], to allow only for thermodynamically feasible solutions. FBA can be coupled to other modelling frameworks as shown in integrated FBA (iFBA). There a stoichimetric model of metabolism is coupled with Boolean and ordinary differential equation (ODE) descriptions of gene regulation and signalling [87].

1.2.2 Logical and Boolean Models

Reactions are only one form of interactions between biological entities, other important aspects are regulatory interactions. Often regulatory interactions are deduced fromm different studies, via a variety of techniques, so that their quantification, such as for stoichiometric coefficients, is impossible. This means that interactions can often only be classified as being either activatory or inhibitory

under certain conditions, and they are unable to be further characterised. For gene regulatory and signalling networks, especially, interactions are often only derived genetically or by perturbation experiments without detailed description of exact mechanisms.

Again a regulatory network can be represented as a graph with the agents, such as genes and transcription factors, as nodes, and regulatory interactions as labelled edges. Analysis of the connectivity of the graph can give interesting insights. Cycles can hint at feedback between genes and strongly connected subgraphs can indicate co-regulated groups of genes or proteins.

Qualitative regulatory interactions can be used to analyse the dynamical behaviour of a system by applying a discrete, logical framework. Logical frameworks have the advantage that they need much less prior information than quantitative dynamical descriptions such as differential equations or stochastic approaches. They also generally require less computational effort and many forms can be treated analytically to predict the global behaviour for all possible initial conditions of a system. This allows us to check whether experimental data can be explained by a model, and also to infer potential network architectures compatible with experimental results [66, 227, 236, 270].

In logical representations the interacting agents of a network, for example genes, proteins, or cells, are represented by logical variables, which only exhibit a limited range of distinct values. In the extreme case, the Boolean description, only two levels, for example 0 and 1, or *active* and *inactive*, are considered. The values of these logical variables can be conceived to be values of the concentration or activity of the associated biological entity relative to certain threshold values. This projection of continuous values of a protein activity or gene expression level can be justified by the observation that the activation of some proteins and genes depends on an effector or transcription factor concentration in a highly non-linear fashion. Often it changes from a basal level to saturation over a relatively small range of effector concentrations, giving rise to a so called sigmoidal response curve [426]. Logical frameworks can be seen as the limiting case of step-like activation functions [406, 409].

The temporal logical model consists of a system of logical equations allowing to determine each variable's value depending on previous states. All values of the logical variables at a certain step, or time point, together give the overall state

of the biological system, and the sequence of states correlates to its temporal behaviour. As logical and especially Boolean models can be efficiently analysed, it is possible to derive the stationary states, and recurring loops of states of a system over a big range of initial states [227, 409].

In the simplest form each state depends only on the previous one, leading to synchronous updating, making the notion of an explicit time variable unnecessary. This modelling approach has been criticised to create unrealistic behaviour as it assumes the same time scale for all responses, thus enabling the simultaneous commutation of variables. To avoid this, a concept of precedence or time delays for variable changes was proposed, in which each change of a variable takes a certain amount of time, and changes with a shorter delay are performed first [405, 406]. This way some changes can be cancelled out creating a much richer variety of behaviours, potentially even deterministic chaos. In multivalued logical models logical parameters can be included, to further quantify the effect of interactions on an element [406, 408].

Boolean models have been successfully used for analysing various different biological networks, for example the gene regulatory network underlying segment polarity in *D. melanogaster* embryos [5, 77], the networks controlling the cell cycle [21, 116, 248], and various signalling networks [227, 351]. Furthermore, some formalisms to account for the stochastic nature of biological systems have been created using a logical framework. One way to achieve stochasticity is by randomly flipping states, for example to look at the robustness of large scale-free Boolean networks [6]. Another approach is to introduce randomly varying time delays [408]. This has been used to understand the noise robustness of segment polarity [77] and the yeast cell cycle [58].

As Boolean models require much less detailed knowledge about a system than other forms of dynamical modelling, they can be used to first explore potential network topologies and parameter values, and later be transformed into more mechanistic, continuous time models [438].

1.3 Differential Equations

Ordinary Differential Equations (ODEs) are another commonly used approach to describe the dynamics of biological networks. For this approach, the system - or

at least each modelled subsystem - is assumed to be spatially homogeneous. This is also known as the *well-stirred approximation*. If spatial concentration gradients are influencing a systems behaviour, Partial Differential Equations (PDEs) can be applied to account for diffusion and spatial heterogeneity. Another important point for the validity of an ODE approach is to determine, whether the number of molecules of each reactant are sufficiently large such that they are not to be influenced by the stochastic nature of chemical reactions. Estimations for molecule numbers below which stochastic fluctuations need to be considered commonly range from 100s to 1000s of molecules per cell [147].

A chemical reaction generally can be seen as a transformation of one set of substances, the reactants, to another set called products. Reactants combine with a fixed ratio to form products, indicating the numbers of molecules of each type of chemical species consumed or produced with each reaction event. The number of molecules of species i consumed in reaction j is called the stoichiometric coefficient, ν_{ij}. Stoichiometric coefficients are negative for reactants and positive for products, the sign indicating consumption or production, respectively. Equation (1) depicts a simple reaction of A and B combining to form the product P.

$$\text{aA} + \text{bB} \longrightarrow \text{pP} \tag{1}$$

The stoichiometric coefficients for this reaction are $\nu_A = -a$, $\nu_B = -b$ and $\nu_P = p$. If this is the only reaction in the system influencing the concentration of A,B, and P, the temporal evolution of the concentrations of these substances can be described by the following system of ODEs in which v represents the reaction rate or reaction velocity:

$$\frac{d[A]}{dt} = -a \cdot v \tag{2}$$

$$\frac{d[B]}{dt} = -b \cdot v \tag{3}$$

$$\frac{d[P]}{dt} = p \cdot v \tag{4}$$

In general, the dynamics of the concentrations of n chemical species taking part in m reactions can be described as a system of ODEs using the stoichiometric matrix

$\mathbf{N}$ and the vector of reaction velocities $\mathbf{v}$. With $\mathbf{x}$ as the vector of concentrations and $\mathbf{p}$ as the vector of parameters, the ODE system looks as follows:

$$\frac{d\mathbf{x}}{dt} = \mathbf{N} \cdot \mathbf{v}(\mathbf{x}, \mathbf{p}) = \mathbf{f}(\mathbf{x}, \mathbf{p}) \tag{5}$$

Where $\mathbf{x}$ is an n dimensional vector, and $\mathbf{v}$ and $\mathbf{f}$ are m dimensional vectors. $\mathbf{N}$ is an $n \times m$ matrix. The real valued function $\mathbf{f}$ is also called the right hand side of the ODE system. As this kind of description is deterministic, the trajectory of such an ODE system is fully defined by giving an initial vector of concentrations $\mathbf{x}_{t=0}$. The trajectories of such a system can never cross, as each point must be the start of a unique solution.

Normally the reaction rates in biological systems involve highly non-linear terms. In consequence the resulting non-linear differential system often can not be solved analytically, and instead has to be integrated over time with the help of numerical methods. While it is often not possible to explore the global behaviour of such a system, fast integration algorithms can be used to analyse dependencies over certain ranges of parameters and initial conditions at least by sampling.

There are various ways to get around the problems of ill-defined parameters and initial conditions. The most straight forward way is to go over a range of parameter values and initial conditions, integrate the system for each combination, and analyse the resulting trajectories or the solution space. Another possibility, if experimental time-series data, or other quantitative results are available, is to estimate fitting model parameters using optimisation methods [79, 278, 285].

1.3.1 Reaction Kinetics

One problem with kinetic approaches in comparison to the before mentioned methods, is the greater amount of information needed in comparison with Boolean and network structure based approaches. As the dynamics of the systems depend on the reaction rates, the ODE description needs detailed information on the mathematical form of the rate laws and the values of parameters involved. Such detailed information is rarely available, and even if it is available, it can still be difficult to translate it into a computationally usable form. The mathematical form of expressions describing the rate of reactions is of great importance

for the behaviour of a dynamic model. If the mechanism of a reaction is well characterised, the rate law sometimes can be derived from first principles.

One of the most general forms for deriving a rate law is mass action kinetics. While it is in principle only applicable to elementary reactions, it still is widely used in biological modelling [198, 362]. In mass action kinetics, the reaction rates are assumed to be directly proportional to the product of the concentrations to the power of their stoichiometric coefficients. The proportionality constant, k, is called the rate constant. For the general reaction in equation (1), the reaction velocity $v([A], [B])$ can be written as follows:

$$v([A], [B]) = k[A]^a[B]^b \tag{6}$$

The exponents in this expression are called the partial orders or molecularities of the reaction, while their sum is the overall order. While molecularities often are identical to stoichiometric coefficients, they also depend on the reaction mechanism, that is the number of molecules involved in the rate limiting elementary reaction. Orders greater than two are quite rare in classical reaction kinetics, as normally collisions of more than two molecules are statistically unlikely. One way of obtaining higher exponents are fast intermediate reaction steps at quasi-steady state. The oxidation of NO with an overall reaction of $2\,NO + O_2 \longrightarrow 2\,NO_2$, for example, has an order of 3 due to the fast dimerisation of NO to N_2O_2 followed by a slower reaction with O_2 [16]. Especially in biological systems higher and non-integer exponents, as well as time dependent rate constants, can also occur due to dimensionally restricted and anisotropic diffusion, such as along the cyto-skeleton or DNA, and molecular crowding [230, 356, 361].

Mass action kinetics are widely used in chemical kinetics, but they can lead to a high number of intermediate steps and parameters in biological systems, many of which might not be experimentally determinable. For the complex mechanisms found in enzyme catalysed reactions, simplified rate laws can be derived using quasi-steady state or rapid equilibrium approximations. A number of common rate laws have been derived this way are described in reference books [85, 373]. There also exist various methods supporting the derivation of rate laws from mechanisms, such as the graph-based one by King and Altman [80, 224].

As the exact mechanistic details of biological reactions are often unknown, generic

and approximate rate laws are widely employed in mathematical modelling. One of the most commonly used one is the irreversible Michaelis-Menten equation. For the reaction of a substrate S to a product P, catalysed by the enzyme E, the following rate law can be derived using the quasi-steady state approximation:

$$\text{E} + \text{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{ES} \xrightarrow{k_2} \text{E} + \text{P} \tag{7}$$

$$v([\text{S}]) = v_{max} \frac{[\text{S}]}{K_M + [\text{S}]}$$

$$\text{with}: v_{max} = k_2 \cdot [\text{E}_0],\ K_M = \frac{k_{-1} + k_2}{k_1}$$

In this mechanism, v_{max} is the maximal velocity for a given overall concentration of enzyme, $[\text{E}_0]$. The Michaelis constant K_M gives the concentration of substrate, at which the rate is half the maximal velocity v_{max}.

While the Michaelis-Menten equation was derived for a defined, simple enzymatic mechanism under the assumption of a quasi-steady state of the enzyme-substrate complex and in absence of products, it is quite often employed in models where this assumption does not necessarily hold [45, 218, 349].

Enzymatic and other biological reactions show a variety of important properties, that can not directly be found in simple reactions obeying mass action kinetics. One of these properties is saturation at high substrate levels due to the limited amount of enzyme available. This also means that the apparent molecularity of the reaction changes with the substrate concentrations. The simple irreversible Michaelis-Menten rate law, for example, changes from an apparent order of one at low substrate concentrations, to an apparent order of zero at saturating substrate levels.

Another important property of biological reactions, is that their rates can be modulated by molecules other than products or substrates. These molecules can either compete directly for binding with the reactants, or alter the activity of an enzyme by binding to different sites, a process called allosteric modulation. Often these modulations display cooperativity, that is, the effect of the ligands depends on their concentrations in a non-linear fashion, where the whole is more (or less) than the sum of its parts. One explanation for cooperative behaviour

are multiple binding sites for a ligand, which influence each other.

The simplest description of a cooperative binding process is the Hill equation, first derived for cooperative binding of oxygen to haemoglobin [183]. It describes the fractional occupancy of an protein bound to a ligand L.

$$\bar{y} = \frac{[\text{L}]^h}{K_H + [\text{L}]^h} \tag{8}$$

In the Hill equation K_H is an apparent dissociation constant, and h, the Hill coefficient, indicates the degree of cooperativity. The exponent h is not necessarily an integer, and in general is different from the number of binding sites n, although n often constitutes an upper bound for h. This is exemplified by Hill's original study, in which the Hill coefficient of oxygen binding to haemoglobin ranged from $h = 1.6$ to $h = 3.2$, while the protein possesses four binding sites for O_2.

One measure of the influence of an effector, J, on the activity of an enzyme is the response coefficient, R_J. For all other conditions being equal, R_J has been defined as the ratio of the concentration of effector leading to 90% of the maximal activity to its concentration producing 10%, $R_J = [\text{J}_{0.9}]/[\text{J}_{0.1}]$ [156]. The irreversible Michaelis-Menten rate law (eq. (7)) has a response coefficient to the substrate S, R_S, of 81, that is, the substrate concentration needs to change by a factor of 81 to increase the reaction rate from 10% to 90% of its maximal value. A rate law following the form of the irreversible Hill equation (eq. (8)) on the other hand, would have response coefficients, R_L, depending on the Hill coefficient h. For a Hill coefficient h greater than 1, R_L becomes smaller than 81, in case of $0 < h < 1$ it becomes bigger. The response coefficient for a Hill coefficient, h, of 2, for example equals 9, indicating that the relative change of effector concentration needed to switch from 10% to 90% of activity is 9 times smaller compared to the Michaelis Menten equation. Goldbeter and Koshland [156] coined the term *ultra-sensitivity* for processes exhibiting response coefficients smaller than 81, that is, being more sensitive to an effector concentration than the irreversible Michaelis Menten equation.

Ultra-sensitivity to a stimulus or effector can arise in enzyme cascades following non-cooperative Michaelis-Menten or mass action kinetics. Two simple mechanisms described are multistep and zero-order ultrasensitivity [156, 157]. Both

mechanisms assume reversible covalent modifications to one enzyme by others, such as found in protein kinase cascades. They can explain an ultrasensitive change of the ratio of the modified to the unmodified form of the enzyme in dependence of a stimulus influencing the rates of the modifying enzymes.

Multistep and zero-order ultrasensitivity can only occur under certain conditions in protein modification cascades. For multistep ultrasensitivity, the effect of the stimulus on the rates of modifying enzymes needs to fulfil certain criteria, while in the case of zero-order ultrasensitivity, the total concentration of the enzyme to be modified needs to be high enough for the modifying enzymes to work in the saturated or zero order regime [156, 157]. These mechanisms have been argued to be involved in the high apparent Hill factors of up to 4.9 reached in mitogen activated kinase cascades, that could hardly be explained by cooperative binding [198].

For use in modelling a generic reversible Hill equation has been derived amongst other, more complex frameworks for cooperative behaviour [194]. Another approach, the Adair-Klotz model, assumes different sequential apparent binding constants for the first, second, and following ligands [2, 228, 229]. In the Monod-Wyman-Changeux model, the bound protein can exist in several conformational states, which coexist in equilibrium, and have different affinities for the ligand [286].

Many generic rate laws have been created to allow both inclusion of allosteric regulation and cooperativity, as well as thermodynamic parameters [85, 105, 194, 241, 249, 250]. The latter point is especially important, as equilibrium constants and free energy data are often more readily available than kinetic parameters. As the net-rate of every reversible enzymatic reaction has to approach zero at equilibrium, there exists at least one relation between the equilibrium constant and the kinetic parameters. These so called Haldane relationships can be used to restrict the amount of free kinetic parameters [85].

Another possibility is the use of approximate rate laws [177], such as generalised mass action and lin-log kinetics [354, 428]. These approaches have some desirable properties, such as a low number and easy estimation of parameters, as well as direct incorporation of characteristics derived from perturbation experiments [177, 431]. On the other hand, they are generally not valid over wide ranges of concentrations, but only close to defined system states.

1.3.2   Stationary Points and Limit Cycles

An important characteristic of a dynamical system is its long-term behaviour. The question whether a system tends to a defined equilibrium or steady state, exhibits periodic behaviour, or whether some solutions grow without bounds is important when comparing it to its biological counterparts. Multistability and oscillations in particular have been found to play important roles in living cells, from signalling [55, 155, 182, 296], over cell cycle control [90, 306, 375], to developmental processes and metabolism [20, 94, 186, 293].

One method of analysing the long term behaviour, of a dynamical system is to look at special points, or regions, in concentration space. These can for example be stationary points, or closed orbits in phase space.

At a stationary point or equilibrium, $\mathbf{x}_s$, the right hand side of the ODE system, $\mathbf{f}(\mathbf{x}, \mathbf{p})$, equals 0.

$$\dot{\mathbf{x}}_\mathbf{s} = 0 = \mathbf{f}(\mathbf{x}_s, \mathbf{p}) \tag{9}$$

A stationary point can be asymptotically stable, or unstable, depending on whether, after a small perturbation, the system tends back to the stationary point or diverges from it. More formally, a fixed point is described as asymptotically stable, if there exists a small region with diameter $\epsilon > 0$ around $\mathbf{x}_s$ such that all points in that region (that is with $|\mathbf{x}(t=0) - \mathbf{x}_s| < \epsilon$) tend towards $\mathbf{x}_s$ and $\lim_{t \to \infty} |\mathbf{x}(t) - \mathbf{x}_s| = 0$.

An important characteristic of the system at a stationary point is the Jacobian matrix, $\mathbf{J}(\mathbf{x_s})$, an $n \times n$ matrix consisting of the partial differentials of $\mathbf{f}(\mathbf{x_s})$ to the various concentrations, with each entry defined by:

$$J_{ij} = \frac{\delta f_i}{\delta x_j} \tag{10}$$

The right hand side of an ODE system can be linearised around a point $\mathbf{x}_0$ by performing a Taylor series expansion using the Jacobian matrix and neglecting

higher order terms. With $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}_0$ follows

$$\dot{\mathbf{x}} = \dot{\mathbf{x}}_\mathbf{0} + \mathbf{J}(\mathbf{x}_0) \cdot \Delta \mathbf{x} + \mathcal{O}(\Delta \mathbf{x}^2) \qquad (11)$$

As in a stationary point $\mathbf{x}_s$ the $\dot{\mathbf{x}}_\mathbf{s} = 0$, the Jacobian matrix $\mathbf{J}(\mathbf{x}_s)$ defines the stability behaviour around such a point. The time development of the deviation from $\mathbf{x}_s$ can be approximated by the following system of linear differential equations:

$$\dot{\Delta \mathbf{x}} = \mathbf{J}(\mathbf{x}_s) \cdot \Delta \mathbf{x} \qquad (12)$$

This linear differential equation system can be readily solved, showing that the eigenvalues, λ, and unit eigenvectors, $\mathbf{u}$, reveal the stability behaviour around the stationary point. For a system with n eigenvalues λ and eigenvectors $\mathbf{u}$, solutions for the approximation around $\mathbf{x}_s$ depend on whether the eigenvalues are real or complex. For real eigenvalues a valid solutions looks like this:

$$\Delta \mathbf{x}(t) = \mathbf{u}_i \cdot \exp^{\lambda_i \cdot t} \qquad (13)$$

For complex eigenvalues $\lambda_i = \alpha + i\beta = \bar{\lambda}_{i+1} = \alpha - i\beta$, with the eigenvectors $\mathbf{u}_i = \bar{\mathbf{u}}_{i+1}$, the following solution is possible [398]:

$$\Delta \mathbf{x}(t) = \exp^{\alpha t} \left(\mathbf{u}_i \cdot (\cos \beta t + i \sin \beta t) + \bar{\mathbf{u}}_i \cdot (\cos \beta t - i \sin \beta t) \right) \qquad (14)$$

In the linearised system the real part of each eigenvalue determines whether the deviation $\Delta \mathbf{x}(t)$ grows, in case of a positive real part $\mathbb{R}(\lambda_{max}) > 0$, or decays, for negative real parts $\mathbb{R}(\lambda_{max}) < 0$, over time, and the eigenvectors give the principle directions along which the trajectories leave or approach. Complex eigenvalues indicate a spiralling mode of the trajectories, or oscillatory behaviour. In case that the real part equals 0, $\mathbb{R}(\lambda_{max}) = 0$, the linear system has a center and the imaginary part can give a limit frequency (see Fig. 2) [398].

This approximation of the systems behaviour gives some simple dependencies of the stability of a stationary point on the eigenvalues of the Jacobian matrix. If the eigenvalue with the largest real part, λ_{max}, is positive, $\mathbb{R}(\lambda_{max}) > 0$, the
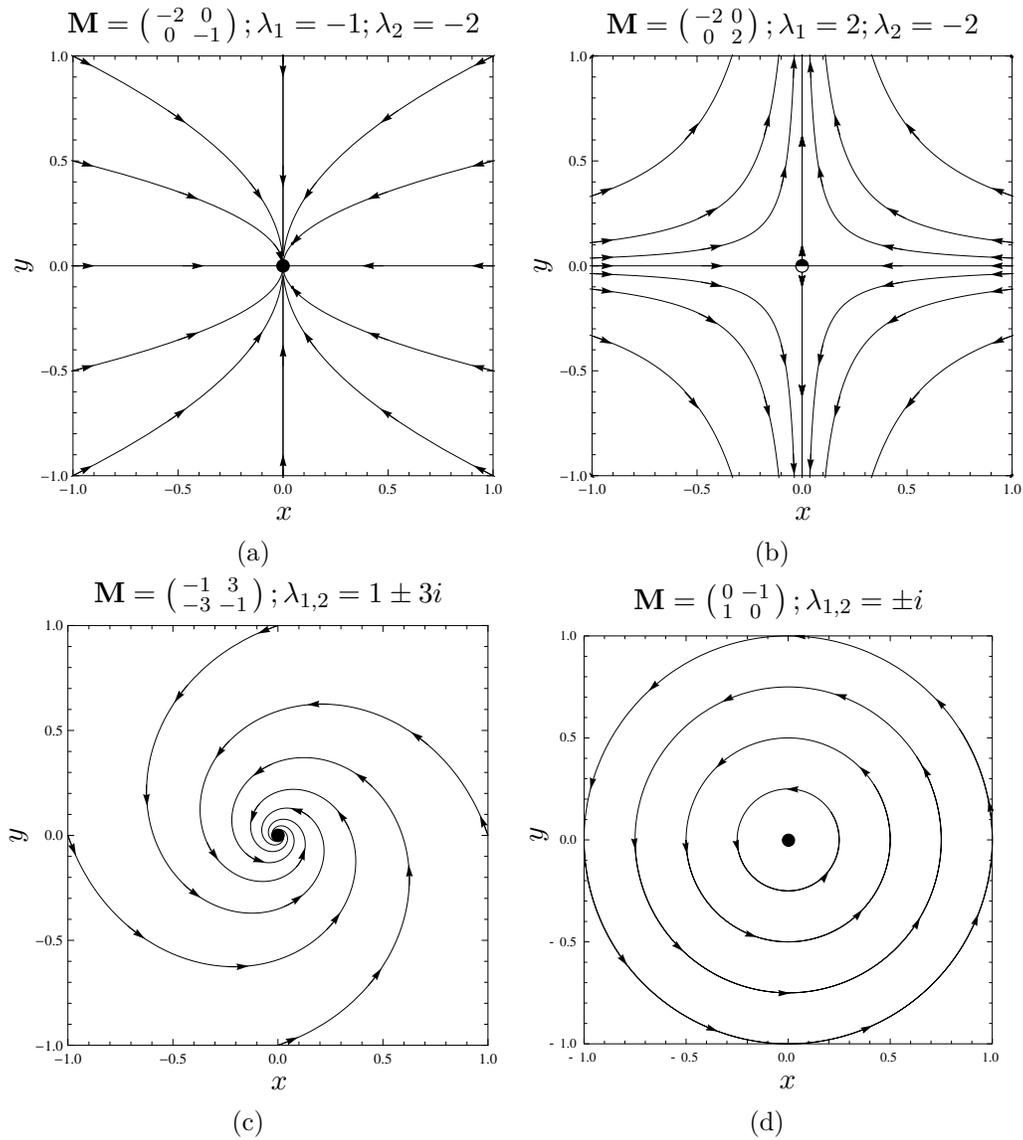
Figure 2: Trajectories of two dimensional linear systems of the form $\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$. The values of the matrix $\mathbf{M}$ and its eigenvalues λ_1 and λ_2 are given above each figure. The origin is a unique stationary point for all systems shown. In (a) it is a stable node, in (b) a saddle. (c) has complex conjugated eigenvalues and the stable node is approached via a spiral. In (d) the eigenvalues are purely imaginary and the center is surrounded by closed periodic orbits.

stationary point is unstable. On the other hand, if $\lambda_{max} < 0$, the point is stable. For cases with $\mathbb{R}(\lambda_{max}) = 0$, higher order terms may need to be considered in the approximation. Unstable stationary points with both positive and negative eigenvalues are also called saddle points.

Another important feature of dynamical systems are closed orbits in phase space.

In some cases closed orbits can be periodic, which means that each of the points on the closed orbit at time t, $\mathbf{x}_o(t)$, recurs after a period of τ, $\mathbf{x}_o(t) = \mathbf{x}_o(t + \tau)$. Limit cycles are a special kind of periodic orbits on a two dimensional manifold. They have the additional criteria, that trajectories starting close to them in phase space either are attracted to them or depart from them. Similar to stationary points, limit cycles are classified as being either stable, or attracting, and unstable.

Stable, or attracting, limit cycles are those periodic orbits for which a neighbourhood exists, in which all trajectories approach the limit cycle in the limit of $t \to \infty$. With unstable limit cycles, on the other hand, neighbouring trajectories withdraw from the orbit - that is they approach the orbit with the limit of $t \to -\infty$. Stable limit cycles are examples of self sustaining oscillations, in which a system returns to periodic behaviour after small perturbations, and have been investigated in models of glycolytic oscillations [94, 346], circadian rhythms [244, 245], as well as the cell cycle [306, 414].

Stationary points can sometimes be connected by trajectories leaving an unstable point along the directions of the eigenvectors associated with positive eigenvalues and approaching over the ones with negative eigenvalues. When an orbit connects two different equilibria, it is called heteroclinic, while if it joins a saddle point to itself, it is called homoclinic. Multiple heteroclinic orbits can create loops from saddle points back onto themselves, creating a heteroclinic cycle. While the cycle itself cannot be followed by a trajectory, as the trajectory would stop as it approaches each saddle point, trajectories can be attracted to or repelled away from the heteroclinic cycle similar to a limit cycle. However, different from a limit cycle, a stable heteroclinic cycle creates oscillations with increasing periods, as the trajectories get closer to the saddle points and stay longer and longer in their neighbourhood [191, 274, 326].

An important concept in connection with stationary points and special orbits, such as limit cycles, are attractors and basins of attraction. An attractor is a closed set of points in concentration space, for which all trajectories starting from a point in the set remain inside. An attractor's basin of attraction is the open set of points in concentration space, from which trajectories tend to the attractor. Stable equilibria and saddles, as well as attracting limit and heteroclinic cycles and orbits are examples for attractors. Normally it is very hard to impossible to

derive the basins of attraction exactly for the large non-linear systems commonly used in biology. Nevertheless, by scanning over different sets of initial conditions, often at least an overview of some interesting regions can be gained.

1.3.3 Multistability, Oscillations and Bifurcations

How the qualitative behaviour of a dynamical system depends on parameter values and intial conditions is an important question to address. For example a system can change the number and stability of its stationary points, or suddenly switch from approaching a single steady state to oscillations, in response to changes in its parameter values. Such changes in the qualitative behaviour of a system are called bifurcations, and are often encountered in biological systems. A common example is switching behaviour in which, depending on an external signal or a change in environment, a cell suddenly shifts from one state to another, for example in stem cell or B-cell differentiation [42, 78], or in cell cycle checkpoints [399, 416].

Bifurcation analysis can be performed analytically, as demonstrated for the system described in chapter 2. However, this is only possible for simpler systems and often needs further reduction of variables. Another possibility is to use numerical techniques. Two software packages widely applied for analysing biological models using continuation methods are the `AUTO` package [99], which is integrated in several programs such as `XPPAut` [113] and the Systems Biology Workbench (SBW) [353], and `MatCont` [98], a *Matlab* package. In addition to these two software packages, a versatile python library for bifurcation analysis, `PyDStool`[2] [82], is also used in parts of this thesis.

One of the simplest bifurcations is the *saddle-node* bifurcation, which occurs, when a stable and an unstable equilibrium collide and annihilate. Similarly the *trans-critical* bifurcation, in which two equilibria change stability upon collision is a second example of a simple bifurcation event [398].

Saddle node bifurcations can be encountered in systems showing switch-like behaviour. In such situations, two saddle node bifurcations occur at two critical values of a specific parameter, for example a kinase activity or a stimulus

---

[2]Clewley RH, Sherwood WE, LaMar MD, and Guckenheimer, JM (2007), `http://pydstool.sourceforge.net`

strength. If these bifurcations both involve the same unstable steady state, the region between them can feature two stable stationary states and one unstable stationary state, while parameter ranges above and below the critical values possess only a single stable equilibrium. This critical parameter can for example be a mitogenic stimulus [399]. In case of increasing stimulus strength a system stays on one stable branch of equilibria past the lower critical stimulus strength until it reaches the upper critical stimulus. At this point it suddenly switches to another stable branch, and stays on it, even if the stimulus strength falls, as long as it stays above the lower critical value. This history dependent behaviour of the system is also known as hysteresis (see Fig 3).

Switches with hysteresis are found in a variety of biological systems, for example in the regulatory systems underlying the cell cycle [375, 399], apoptosis [20] or stem cell differentiation [78]. If the lower critical parameter value lies beneath the feasible range, for example at a negative signal strength, the switch can be irreversible. In some cases the two saddle node bifurcation points collide and vanish over a range of values of at least two parameters forming a cusp [398] (see Fig 3).

Slightly more complex than saddle-node bifurcations are *pitchfork*, or *double point* bifurcations, in which a single equilibrium splits into three. Two classes of pitchfork bifurcations are distinguished, *sub-* and *supercritical*. In *supercritical* pitchfork bifurcations, a stable equilibrium splits into two stable and one unstable stationary points, while conversely in the *subcritical* case an unstable stationary point gives rise to two unstable and one stable equilibria [398].

An *Andronov-Hopf* or for short *Hopf* bifurcation is defined as a change of the stability of an equilibrium under appearance or disappearance of a limit cycle. As with the pitchfork bifurcation, Hopf bifurcations are classified as *sub-* or *supercritical* [398].

In a *supercritical* Hopf-bifurcation, a stable equilibrium becomes unstable, giving rise to a stable limit cycle with zero amplitude and a finite frequency. This manifests itself in oscillations with gradually increasing amplitude as the limit cycle start off from the bifurcation point.

In the *subcritical* case, an unstable equilibrium changes stability to become stable, creating an unstable limit cycle. Approaching the bifurcation point from the
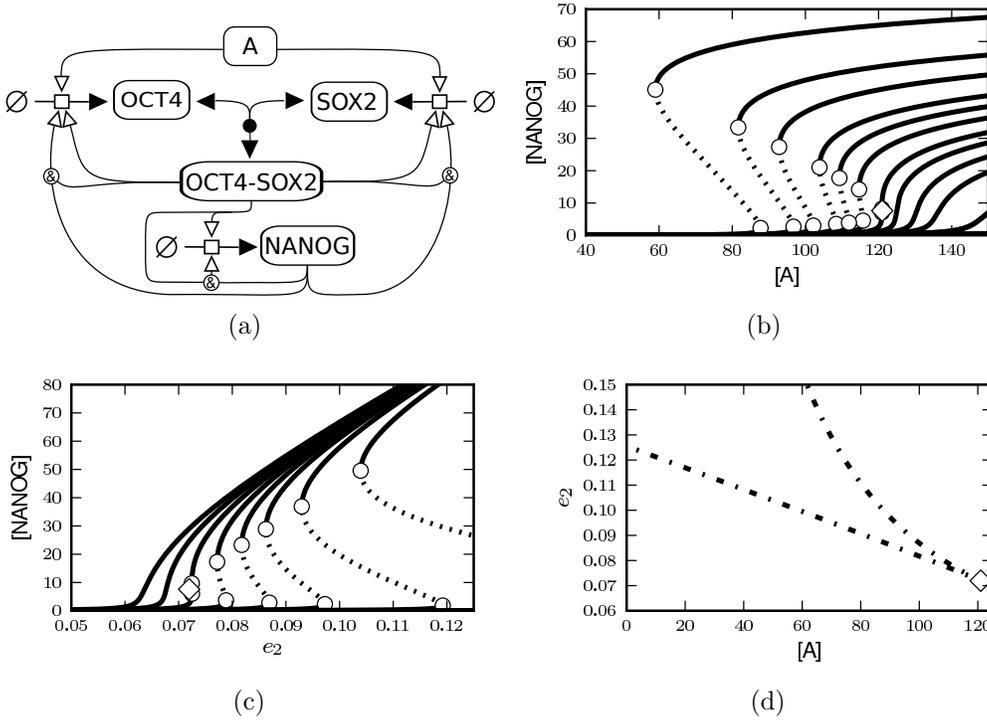
Figure 3: Bistable switch with hysteresis and cusp in the model of the OCT4-SOX2-NANOG gene regulatory network described in Chickarmane *et al.* 2006 [78]. (a) gives an overview of the network. Open arrows stand for transcriptional activation, full arrows for reactions. In this system, A is a stimulating transcription factor or signal that activates expression of both OCT4 and SOX2. OCT4 and SOX2 bind to form the transcriptional activator OCT4-SOX2, which in turn activates transcription of the genes encoding OCT4, SOX2, and NANOG. NANOG can only bind and activate transcription of all three genes in combination with the OCT4-SOX2 complex. The positive feedback by NANOG on itself is directly proportional to the parameter e_2. (b) shows the steady state concentration of NANOG at different concentrations of A. Each curve is calculated for a constant value of the parameter e_2 (from left to right: 0.1, 0.09, 0.085, 0.08, 0.075, 0.075, 0.072, 0.07, 0.075, 0.065, 0.06 and 0.05). Solid lines indicate stable steady states, dotted lines unstable states. Circles indicate saddle-node bifurcation points, the diamond a cusp point. The system is bistable and shows hysteresis in regions between two saddle node bifurcations. (c) shows the steady state concentration of NANOG for varying values of e_2 and constant $[A]$ (from left to right: 10, 25, 50, 75, 90, 100, 110, 120, 130 and 140). (d) shows the saddle-node bifurcation curve in the $[A]$-e_2 plane and the annihilation of the bistable region after the cusp point. In the model, decreasing feedback of NANOG on its own expression leads to loss of bistability.

stable side, a subcritical bifurcation exhibits quite different behaviour from the supercritical one. With the disappearance of the unstable limit cycle and change of stability of the equilibrium, the system tends to a new attractor. In case this

is a stable limit cycle, oscillations can suddenly start with the amplitude of the surrounding cycle [203, 415]. This type of behaviour occurs in various models of neuronal excitation such as the Hodgkin-Huxley model [189, 203]. Further, subcritical Hopf bifurcations have been indicated to occur in cell cycle related oscillations [305, 306, 414].

The change of stability in a Hopf-bifurcation is accompanied by a pair of complex conjugated eigenvalues of the Jacobian matrix crossing the imaginary axis, $\lambda_i = \bar{\lambda}_{i+1} = \beta$i. The imaginary part, β, of this pair gives the frequency, $\beta/(2\pi)$, of the zero-amplitude oscillations in the bifurcation point [203, 398].

In addition to changes of simple stationary points, there also exist cases in which bigger invariant sets, such as limit cycles, collide. These are also known as global bifurcations, as they encompass bigger ranges of phase space. In analogy to the saddle-node bifurcation, the collision and annihilation of two limit cycles of opposed stability is called a saddle-node bifurcation of cycles [203, 398]. Other bifurcations involving limit cycles are the infinite-period, or saddle-node on invariant cycle, and homoclinic bifurcation.

In the infinite-period bifurcation the period of a limit cycle increases until a saddle point appears on the limit cycle and it suddenly becomes infinite. The similar homoclinic bifurcation occurs when a limit cycle approaches a saddle point and at the bifurcation point fuses with it to become a homoclinic orbit [203, 398].

Another interesting dynamical phenomenon is the appearance of deterministic chaos. Chaotic behaviour has been intensively studied in chemical systems [112], but also has been described in biological systems exhibiting oscillatory behaviour, especially in neurons [272], glycolytic oscillations [299], and calcium oscillations [197]. One commonly found way to chaos is via a series of period doubling bifurcations of an oscillating system until a point of infinite period is reached at which aperiodic oscillations occur [67, 112, 272, 299].

1.3.4 Feedback Mechanisms

Of special interest with respect to multi-stationarity and complex behaviour are feedback circuits, or loops, in which for example a molecule influences its own concentration, or a gene product modulates the rate of their own expression.

These regulatory circuits can either be direct, or they can encompass various intermediate interactions. Feedback circuits are especially well researched in gene regulatory networks. Depending on the number and signs of interactions in such a circuit, the overall effect of each gene product on its own expression can be enhancing or inhibitory, constituting a positive or negative loop, respectively. A loop containing only positive and/or an even number of negative interactions constitutes a positive loop, while one containing an odd number of negative interactions represents a negative one [407, 408].

Regulatory systems containing feedback loops give rise to some remarkable phenomena. As such they have been the subject of extensive experimental and theoretical investigation.

One important role ascribed to positive feedback is increased sensitivity in responses to stimuli or signals – all-or-none, or switch-like behaviour. While such behaviour can be achieved by other means, positive feedback has been found to increase the sharpness of response, for example in *Xenopus laevis* oocyte maturation [119]. *Xenopus laevis* oocyte maturation is induced by progesterone, via activation of a mitogen-activated protein kinase (MAPK) cascade. While activation of the MAPK cascade shows ultrasensitivity to progesterone on its own, it is further increased by protein expression dependent positive feedback [156, 198].

Positive feedback loops have been identified as a necessary condition for the existence of multiple steady states [12]. In the common case of two stable steady states, bistability, transitions between the two states can be triggered by changes in the systems input parameters. Bistable switches are assumed to underlie various essential processes, for example in apoptosis, cell cycle control and some forms of signal transduction [20, 41, 416, 448].

Gene regulatory switches were amongst the first to be identified to contain positive feedback loops, such as the lysogeny/lysis switch in λ phage [335]. In mammalian embryonic stem cells a genetic switch consisting of three mutually activating transcription factors (TFs) is involved in differentiation [78]. Another cell differentiation process, blood cell differentiation triggered by erythropoietin, involves switching the transcription factor GATA1 from a low to a high expression level via two positive feedback circuits [314]. In an engineered system a single positive feedback loop was sufficient to construct an inducible toggle switch in mammalian cells [381].

Feedback loops containing an odd number of negative interactions leading to an overall positive feedback have also been found in biological systems, for example in the regulation of the λ phage cI repressor via the cro transcription factor [337, 403]. Recently a double negative feedback loop has been found to play a central role in an epigenetic switch in *B. subtilis* [73]. A similar approach with two mutually repressing transcription factors was employed to produce synthetic toggle switches in *E. coli* [138].

Negative feedback in gene regulation and signal transduction has been indicated to lead to homeostasis and noise reduction [35, 131, 355, 394] and has been connected to faster response times [302]. Further, negative feedback has been implied as a requirement for robust perfect adaptation [449] - the capability of sensing mechanisms to adapt to persistent stimuli. This is, for example, found in bacterial chemotaxis, in which the bacterium needs to sense changing concentrations of small molecules. There it is achieved by chemoreceptors, whose activity is regulated via methylation in a negative feedback loop, thus allowing the adaption of sensitivity to temporal differences in concentration, rather than absolute levels [8, 30].

Early studies in transcriptional regulatory systems indicated, that negative feedback can lead to stable oscillations [159, 162]. Negative feedback seems to be a requirement in all the biological oscillators characterised to date. Analysis of mathematical models has shown, that in addition to negative feedback some sort of a time delay is necessary for sustained oscillations to appear [307]. In metabolic oscillators the time delay can consist in intermediate reaction steps, such as is seen in glycolytic oscillations [40, 260], while in gene regulatory networks it can additionally be achieved by mRNA export, translation, and protein import to the nucleus [307]. Without an explicitly introduced time delay or destabilisation by an additional positive feedback, a negative feedback loop must contain at least three intermediate steps to exhibit stable oscillations [409].

Another possibility to achieve a time delay, and subsequently sustained oscillations, is by coupling negative and positive feedback. In so called hysteresis driven oscillations a bistable switch is coupled with a negative feedback, that drives it from one state to the other [307, 415]. Such hysteresis driven oscillators are thought to underlie for example the periodic cyclic AMP (cAMP) production in *D. disoideum* [269] and MPF activation in *Xenopus laevis* oocyte extracts

[305]. Different from delayed negative feedback oscillators, hysteresis driven oscillators can undergo a subcritical Hopf bifurcation and start oscillating with a large amplitude (see Fig. 4).
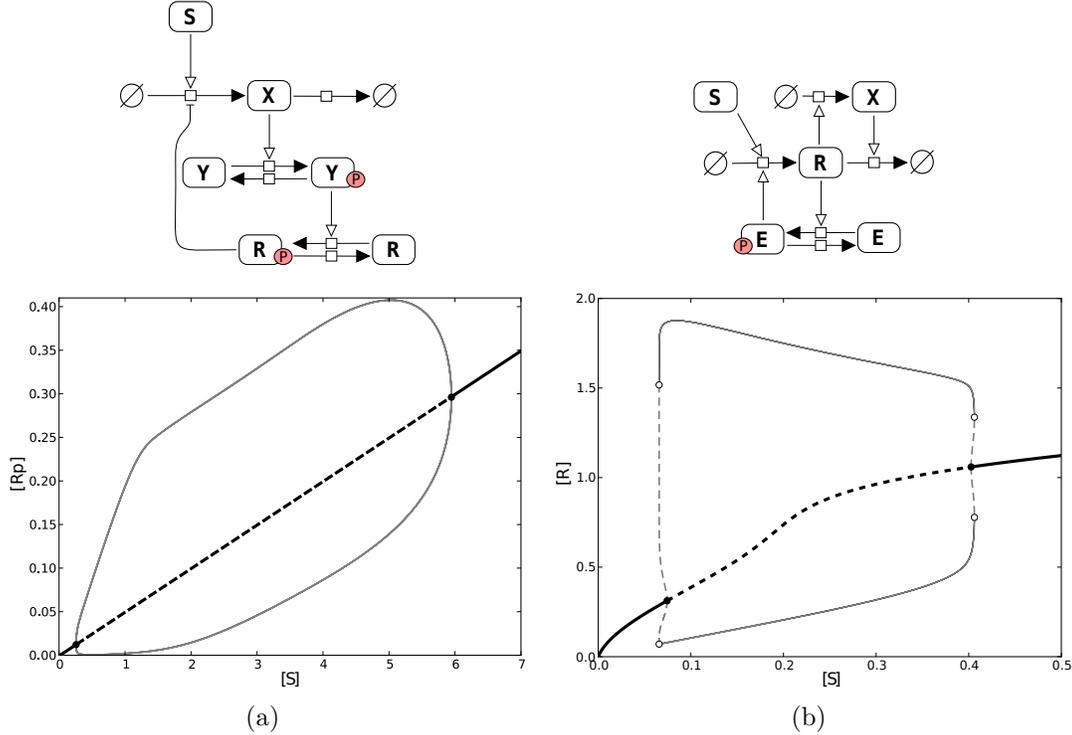
Figure 4: Network topology (top) and bifurcation diagrams (bottom) of a delayed negative feedback (a) and a hysteresis oscillator (b) (after fig. 2 a) and b) in [415]. In the delayed negative feedback oscillator (a) a signal S leads to activation of a protein R via a cascade. The active form, Rp, inhibits the start of the cascade. In the mixed positve-negative feedback oscillator (b) the signal S directly leads to production of R, which enhances its own syntheses via the positive feedback-cycle R → Ep → R. At the same time it increases the synthesis rate of X, which leads to faster degradation of R. The bifurcation diagrams show both the concentration of Rp and R, respectively, at the stationary points (black line) and their maximal and minimal concentrations during limit cycle oscillations with varying signal strength [S]. Solid lines indicate stable, dashed unstable equilibria or limit cycles. The solid black dots are the Hopf bifurcation points. The delayed negative feedback oscillators undergoes supercritical Hopf bifurcations leading to smoothly increasing oscillation amplitudes, while the hysteresis oscillator oscillations start with a high amplitude at subcritical Hopf bifurcations.

Gene regulatory feedback has been found at the heart of other circadian oscillators in cyanobacteria and many eukaryotes [103]. The circadian rhythms in *D. melanogaster* seem to depend on a negative feedback of the two proteins Period and Timeless on their own expression [152, 244].

In early chordate development, a negative transcriptional feedback of the Hes1 protein on its own expression drives cell autonomous oscillations essential for somite segmentation [186]. The oscillatory expression of the tumour suppressor p53 upon DNA damage seems to depend on a negative feedback via the mdm2 protein [28, 235, 443]. In this case p53 activates transcription of mdm2, which in turn sequesters p53 and targets it for degradation. A similar system exists for the mammalian transcriptional activator NF-κB, which controls, amongst other targets, the expression of IκB. IκB binds NF-κB and sequesters it in the cytoplasm, thereby abolishing its trans-activating capability, leading to a negative feedback [193].

Negative transcriptional feedback has also been used to produce a number of synthetic oscillatory systems [93, 108, 221]. A mixed strategy of positive autoregulation and a negative feedback loop was employed to create a two protein oscillator in *E. coli*, consisting of an activating transcription factor, NRI, and a repressor, LacI [17]. NRI both activates its own expression and that of the lac repressor, LacI, which in turn inhibits NRI transcription, leading to a hysteresis driven oscillator.

1.4 Stochastic Approaches

ODE based approaches have some significant deficiencies when modelling biological processes. First they are assuming variables with continuous values, which does not reflect the discrete nature of molecule numbers. Another problem is that deterministic approaches do not account for the random nature of chemical reactions and biological processes due to thermal noise, random collisions, and diffusion of particles.

Both discrete numbers and stochasticity can sometimes be neglected for high molecule concentrations, such as often found in metabolic systems, but in many biological systems, ranging from gene expression [109, 121, 319] to cell signalling [220] and differentiation [315], the numbers of key molecules per cell can be very low and noise plays an important role. Genes normally only exist in one or two copies per cell, and many transcription factors and proteins involved in signal transduction can be available in low numbers.

In such cases deterministic approaches can sometimes create misleading results,

for example occluding potential oscillatory behaviour of genetic regulatory networks [427], or underestimating the sensitivity to a signal, such as seen in stochastic focussing and resonance [172, 321].

Random fluctuations in chemical systems generally are proportional to the square root of the particle number N, $\sqrt{N}$. As a rule of thumb stochastic variation and discrete numbers need to be considered in systems with molecule numbers below hundreds or thousands of molecules per cell [147].

1.4.1 The Chemical Master Equation

Distinct from deterministic approaches, a stochastic framework does not produce definite solutions for the concentrations at a time point t, but only gives the probability of a certain state at a certain time. For a chemical reaction system of n species undergoing m reactions, these probabilities are determined by the Chemical Master Equation (CME) [147].

$$\frac{\delta P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\delta t} = \sum_{j=1}^{m} [a_j(\mathbf{x} - \nu_j) P(\mathbf{x} - \nu_j, t | \mathbf{x}_0, t_0) - a_j(\mathbf{x}) P(\mathbf{x}, t | \mathbf{x}_0, t_0)] \quad (15)$$

In this $\mathbf{x}$ is an n-dimensional vector with each x_i standing for the number of particles the ith molecular species. $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ stands for the probability that the state $\mathbf{x}$ is reached at time t given the state $\mathbf{x}_0$ at time t_0. ν_j is a vector with the stoichiometric coefficients of each species for reaction j, that is the jth column vector of the stoichiometric matrix $\mathbf{N}$, and $a_j(\mathbf{x})$ is called the propensity function of reaction j.

Exact analytical solutions of the CME are only possible in very simple, linear cases and seldom feasible for systems encountered in biology. One way of obtaining approximate solutions was derived by van Kampen. In the van Kampen size expansion the discrete, stochastic molecule numbers $\mathbf{x}(t)$, are split into a deterministic, macroscopic part, $\Omega \phi(t)$, and random fluctuations proportional to the square root of the system size Ω around it [369, 423]:

$$x(t) = \Omega \phi(t) + \sqrt{\Omega} \xi \quad (16)$$

In this ξ stands for a random variable. When defining the system's size Ω as the volume, $\phi(t)$ is the deterministic solution for the temporal development of the concentrations.

1.4.2 Stochastic Simulation Algorithms

While the CME can be solved numerically, this is computationally very expensive for bigger systems. There exist however a range of efficient algorithms for simulating possible trajectories $\mathbf{X}(t)$ of a system, with the overall distribution of such trajectories satisfying the CME.

Two exact algorithms for obtaining such solutions are the direct and the first reaction methods developed by Daniel Gillespie [147]. They constitute two complementary, stepwise approaches. In the direct method, in each time step first the time to the next reaction is randomly determined, and afterwards the kind of reaction, while in the first reaction method the individual time intervals until each of the m reactions occurs are sampled, and the next reaction to occur is chosen [147, 148]. The direct method is generally faster, but an improved version of the first reaction method, called next reaction method, has been developed and is also frequently employed [145].

For these algorithms each of the m reactions is represented by a reaction channel R_i, with an associated vector of stoichiometries coefficients $\nu_i = \nu_{1i}, ..\nu_{ni}$ for each of the n molecular species in this reaction, and a propensity function $a_j(\mathbf{x})$.

The propensity function is central to the stochastic approach, and $a_j(\mathbf{x}) \cdot dt$ gives the probability that reaction j will occur during the time interval $[t, t+dt)$. Under the well stirred assumption at least for reactions in dilute gases propensity functions for elementary reactions can be explained and derived from kinetic theory [147]. For simple elementary reactions, they are similar to deterministic rate laws, and their parameters can be related to the classical, deterministic kinetic constants used in mass action kinetics.

In the case of a monomolecular reaction, $X \longrightarrow P$, such as radioactive decay or dissociation of a protein complex, the probability of a single molecule to react in the interval dt is $c \cdot dt$ for each molecule. For x molecules the propensity function follows as $a(x) = c \cdot x$, similar to the deterministic rate law.

1. Introduction

In a simple bi-molecular reaction, $X + Y \longrightarrow P$, the propensity function $a(x, y) = c \cdot x \cdot y$, is similar in form to the deterministic reaction velocity, $v(x, y) = k \cdot [X] \cdot [Y]$, only using molecule numbers, x and y, instead of concentrations, $[X], [Y]$. A special case are reactions of two instances of the same molecule, such as the homo-dimerisation $X + X \longrightarrow P$ for example. As each molecule cannot react with itself, and in each pair of reactants the molecules are indistinguishable, the propensity function follows as $a(x) = \frac{c}{2} x \cdot (x - 1)$, and is markedly different from the deterministic rate law $-dX/dt = k \cdot [X]^2$.

While the propensity constants cannot be easily derived from first principles, they can be related to the conventional deterministic rate constants. For the monomolecular reaction the value of c is independent of the system's volume V and equal to the deterministic rate constant k. The propensity constant for bi-molecular reactions of two distinct substrates on the other hand, depends on the system size and follows as k/V with potential additional terms for unit conversion if k is given, for example, in moles. For the homo-dimerisation reaction mentioned above, an additional factor of 2 has to be considered and the propensity constant c follows as $2k/V$.

Most stochastic simulation methods do not try to predict a potential trajectory $\mathbf{x}(t)$ by solving $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$, but a different probability function $p(\tau, j | \mathbf{x}, t)$, with $p(\tau, j | \mathbf{x}, t) \cdot d\tau$ representing the probability that the next reaction will be reaction R_j and occur in the interval $[t + \tau, t + \tau + d\tau)$. Defining the sum of all propensity functions at a given time point as $a_0(\mathbf{x}) = \sum_{i=1}^{m} a_i(\mathbf{x})$ this probability function can be derived as [147, 150]:

$$p(\tau, j | \mathbf{x}, t) = a_j(\mathbf{x}) \cdot e^{-a_0(\mathbf{x}) \cdot \tau} \tag{17}$$

Based on equation (17), the direct method uses two random variables, r_1 and r_2, to derive the time τ to the next reaction event and the reaction channel j. The numbers are drawn from the interval $(0, 1]$. The time, τ, to the next reaction event follows a Poisson distribution and is determined as:

$$\tau = \frac{1}{a_0(\mathbf{x})} \cdot \ln\left(\frac{1}{r_1}\right) \tag{18}$$

28

Then the type of reaction is determined as the smallest integer j satisfying:

$$\frac{\sum_{i=1}^{j} a_i(\mathbf{x})}{a_0(\mathbf{x})} \geq r_2 \tag{19}$$

The simulation time is then increased by τ, the molecule numbers are changed according to the stoichiometric coefficients of reaction R_j, and the next time step can be calculated with the updated propensities.

In addition to the exact methods, there exist some approximate ways of generating stochastic trajectories. One of them, τ-leaping [149], jumps over time intervals of length τ, under the assumption that the propensity functions of the system do not change significantly during this time period. The τ-leaping approximation is taken a step further in the Chemical Langevin Equation [149, 369], which describes the system's evolution as a system of continuous stochastic differential equations with a deterministic part equivalent to deterministic reaction kinetics and an added noise term proportional to the square root of the propensities ($W(t)$ represents a Wiener process equivalent to white noise [369]):

$$\frac{d\mathbf{X}(t)}{dt} = \sum_{j=1}^{m} \nu_j a_j(\mathbf{X}(t))dt - \sum_{j=1}^{m} \nu_j \sqrt{a_j(\mathbf{X}(t))}dW(t) \tag{20}$$

However, the level of detail of discrete stochastic modelling comes at a cost. First a lot more information about a system is needed than for the approaches introduced above. As all reactions need to be modelled as elementary reactions, detailed mechanisms and more rate constants, or propensities, are necessary. The overall computational cost is in general considerably higher, too. Even with efficient algorithms for solving or approximating the chemical master equation, many of them scale with the number of reaction events per time. Furthermore, and more general a problem, each solution over time represents only one possibility, so a sample of solutions needs to be computed and analysed (see Fig. 5).

In many cases a system has to be interpreted using different modelling frameworks. For example in genetic regulatory networks, a continuous deterministic framework might be useful for bifurcation analysis of the qualitative behaviour, while a probabilistic discrete approach could be used to explore robustness and
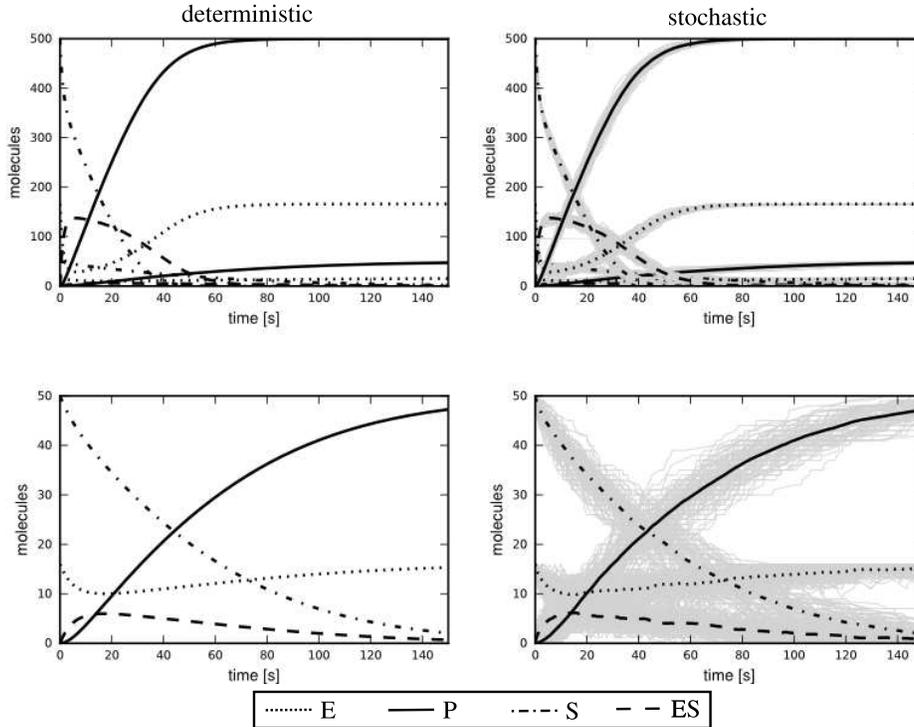
Figure 5: Deterministic (left) and stochastic (right) time-course simulations for a simple enzymatic reaction of a substrate S to a product P catalysed by the enzyme E : $S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_3}{\longrightarrow} P$. In the stochastic time-courses the black lines indicate the mean of a 100 simulations, while the grey lines show the individual trajectories. The initial conditions were $S = 500$, $E = 166$ and $ES = P = 0$ for the upper panels, and $S = 50$, $E = 17$ and $ES = P = 0$ for the lower panels. The rate and propensity constants, respectively, were $k_1 = 0.00166; k_2, 0.0001; k_3 = 0.1$. Time courses were calculated using the tool `SloppyCell`[3] [294], which employs the direct method for stochastic simulations.

behaviour at low concentrations of some species [108, 316, 388].

A general problem of both the deterministic and the stochastic approaches is the validity of classical chemical kinetics in the environments found in organisms. Most rate laws assume a well stirred, isotropic solution, in which particles are allowed to diffuse freely - conditions approximately true in gas phase or diluted solutions, but not inside cells or living organisms. Intracellular environments are filled with macromolecules, and fibrous and membranous structures that lead to a reduction of the available solute volume and diffusion of molecules with a strong, nonlinear size dependency [283, 284]. The occupation of up 40% of the cellular fluid volume by macromolecular structures leads to effects subsumed under the term molecular crowding [171]. Molecular crowding has effects on both reaction thermodynamics and rates, especially on the macromolecules, such as proteins [171, 361]. Molecular crowding can lead to time dependent rate constants and fractal-like kinetics with rate laws differing substantially from classical mass action kinetics [164, 230, 361]. Under these circumstances it is sometimes a little bit surprising how well some models based on ideal reaction kinetics can predict even complex behaviours.

1.5 Gene Regulatory Networks

Genes display a wide variety of different expression patterns ranging from simple constitutive and virtually unregulated cases, over homeostasis and switch-like behaviour, to complex oscillations with stable periods over a wide range of conditions. Many of these observed patterns stem from the capability of gene products, protein or RNA, to modulate gene expression by activating or inhibiting their transcription.

1.5.1 Layers of Control in Gene Expression

The expression of a gene can be regulated at different stages. To create a functional protein, a gene first needs to be transcribed into mRNA by a polymerase, which then is then translated into a protein by ribosomes. In prokaryotes mRNA transcription and translation are not spatially separated, while in eukaryotic cells

---

[3] http://sloppycell.sourceforge.net/

transcription occurs in the nucleus and the transcribed RNA needs to be modified commonly, sometimes spliced, and exported to the cytoplasm to be translated. This creates an additional layer of regulation in gene expression.

Mechanisms to control expression exist at all stages of this process. It is convenient to separate them into processes that regulate the first process, transcription, alone, and those who influence all later stages.

1.5.2 Transcriptional Regulation

Gene transcription in prokaryotes can be divided into several different stages: polymerase binding, initiation and promoter clearance, elongation and termination. First the RNA polymerase (RNAP) binds a special region upstream of the coding region called the promoter. This region contains the core promoter elements which are located between 10 and 35 base pairs (bp) away from the transcription start site. Recognition of the core promoter elements requires RNAP to associate with proteins called σ-factors. After binding the RNAP separates the two DNA strands and initiates synthesis of the RNA molecule, and shortly after leaves the promoter region [62]. If the RNAP fails to dissociate from the promoter, an incomplete transcript is produced. The RNA molecule is then elongated until termination, at which the RNA dissociates from the RNAP. This can either happen by binding of a special protein, the ρ factor, or by a special hairpin structure formed by the RNA molecule.

One common way of modulating transcription, are sequence specific DNA binding proteins, or protein complexes, called transcription factors. The DNA motifs recognised and bound by transcription factors are called cis-regulatory elements or transcription factor binding sequences. The term cis stems from a genetic classification. *Cis*-regulatory elements have to be encoded on the same DNA molecule as the regulated gene, while *trans*-regulatory factors, such as transcription or specific σ factors, can be encoded on a separate one. Depending on whether they increase or decrease the transcription rate, transcription factors are classified as transcriptional activators or repressors, respectively. Some transcription factors, called dual regulators, can act as both activators and repressors, often depending on the location of their binding sequences [324]. The transcription factor Cra, for example, functions as an repressor, when binding downstream of the promoter region, and as an activator when bound upstream

[352].

Various ways for a transcription factor to mediate repression have been described (reviewed in [62, 422]). First a transcription factor can sterically hinder σ-factor or RNAP recruitment, by binding close to or on the core promoter region, or block elongation by binding downstream of the transcription start site [240, 337, 352]. A common mechanism lies in binding up and downstream of the promoter region and creating a DNA loop that prevents RNAP binding or clearance, such as in the case of the *E. coli* lacI repressor [265]. A similar sterical inhibition mechanism is also found in eukaryotes, for example as with the GalR repressor of *S. cerevisiae* [291]. A different form of repression is modulating the action of an activating transcription factor, as for example the Cyt repressor, CytR, that interferes with activation by the catabolite activator protein (CAP) [291, 421].

Conversely, a common mechanism for activator action is the enhancement of the recruitment of σ-factors or RNAP. Transcription factors can bind upstream of the core promoter and interact directly with the α-subunit of the RNAP, for example the Fis activator at the proP P2 promoter[277], a mechanism called class I activation [62]. Alternatively, they can bind directly next to or on the core promoter and interact with σ-factors, as in the case of λ repressor auto-activation at the cI promoter [188, 337]. Another way of increasing RNAP or σ-factor binding, is alteration of the DNA structure at the promoter region, for example by twisting or bending it, as seen with members of the MerR family of transcription factors, that align sub-optimally spaced core promoter elements [180, 297]. Some transcription factors do not alter the binding of RNAP, but rather activate bound RNAP-holoenzymes, as seen with the NtrC activator [337]. Finally, a transcription factor can also relieve repression, by binding adjacent to a repressor and modulating its activity, a mechanism called antirepression. This has been described for the ComK activator in *B. subtilis*, which modulates repression of its own gene, comK, by RoK and CodK [387].

Sometimes transcription factors alter their activity by binding small molecules, called cofactors, allowing cells to quickly respond to changes in their environment. The famous lac operon shows this kind of behaviour for both the lac repressor, LacI, and CAP. Binding of an inducer, such as allolactose relieves repression by lacI, while transcriptional activation by CAP requires binding of cAMP [337]. Other cases of regulation of gene expression by multiple activators, repressors

and cofactors are the metE and metH genes, regulated by MetJ and MetR [434].

While many activators function independent of each other, in a few cases different activating transcription factors need to work together, that is cooperatively. One possible mechanism for this is cooperative binding, which has been described for the melAB promoter and the transcription factors MelR and CAP [432]. In other promoters, an otherwise sufficient activator might not work by itself, but need to be repositioned by another secondary activator. Activation of the malK promoter by the MalT activator, for example, requires binding of CAP to shift MalT from high to low affinity binding sites [347]. Another, similar, mechanism for secondary activators has been implicated for integration host factor protein (IHF). IHF is assumed to bend DNA so that the transcription factor NarL binding 200 bp upstream of the promoter region can interact with RNAP [364].

Gene transcription in eukaryotes has been found to be governed by more complex processes involving many different factors [195, 246]. Most protein coding mRNAs are transcribed by RNA polymerase II, but require different additional protein complexes, called general transcription factors such as TFIID, for core promoter recognition [246, 300]. Additionally coactivators, such as TFIIA [452], different mediator complexes [246], and chromatin remodelling complexes, such as the Swi/Snf [325], can be required [195]. As covalent modifications of histones also can influence transcription from some promoters, histone modifying complexes can modulate transcriptional activity. For example, the histone acetylase complex SAGA functions as a coactivator [396], while some histone deacetylases such as Sin3/RPD3 can function as corepressors [210].

As in prokaryotes, sequence specific transcription factors play an important role in eukaryotic gene regulation. Many different mechanisms have been found for eukaryotic transcription factors. Some activators, such as the yeast Gal4p, interact directly with various coactivators, and the SAGA complex [43]. The yeast Ume6p transcription factor, on the other hand, represses transcription by recruiting histone deacetylases [210]. Similar to prokaroytic transcription factors, some eukaryotic transcription factors bind close to the core promoter to proximal promoter elements. Others were found to influence the regulation by binding sequences situated far away from the promoter [48]. These so called enhancers are generally cis-acting DNA sequences, that can lie many kilobases upstream, or downstream of the promoter region of the gene they regulate. In some instances

enhancer elements can even directly influence transcription of genes located on different chromosomes [255].

Two important principal ways of transcriptional regulation by enhancer elements have been suggested. In the classical gradual or rheostatic model, transcription factors directly enhance the transcription rate, while in the binary or probability model, binding of transcription factors increases the probability that a gene becomes transcriptionally active, or maintains transcriptional activity [121, 201]. The gradual model predicts, that in dependence on a single transcription factor, expression levels should increase for all cells in a population, while according to the binary model only the number of cells expressing a gene, and not the expression levels should vary. Over cell populations, these two models may be hard to distinguish, as both can lead to gradual increases in averaged expression levels. In single cell studies, the differences between these two models become more obvious, and many studies have shown the predicted all-or-nothing response [120, 214]. Graded responses to single transcription factors have been observed in only a few single cell studies [231].

A widespread phenomenon in transcriptional regulation is cooperativity. This means a more than linear change of transcription in response to a linear change of transcription factor concentration, giving rise to a sigmoidal rather than a hyperbolic response curve. One explanation for this behaviour are cooperative effects in transcription factor binding. Some transcription factor bind as di- or multimers, which can lead to an increased concentration dependence, or interact after binding to the DNA [1, 335]. Another possibility are synergistic effects, which can occur without direct interactions between bound transcription factors. One explanation for this, is that bound transcription factors can cooperate in the recruitment of the transcriptional machinery. This mechanism is widely found in eukaryotes [426], for example in some cases of activation by the Gal4p transcription factor [69], but also in prokaryotic systems, as seen for the *E. coli* transcription factor CRP [65, 209].

1.5.3   Mathematical Models for Transcriptional Regulation

In the literature there exist numerous ways of deriving mathematical expressions for regulation of gene transcription or expression. One of the main problems is to conceive a model of the transcriptional activity of a gene in dependence of

the concentrations, or activities, of its regulating transcription factors and small cofactors.

Various approaches based on logical functions have been suggested [216, 402, 405]. In the generalised threshold models, the activities of promoters are approximated as step functions switching from one value to another at threshold levels of transcription factors [409]. These functions can then be used in a differential equation form [334], or in Boolean or logical [403, 404] models. The step functions can be seen as Hill functions (equ. (8)) at the limit of high cooperativity.

The step function approach can be extended to graded piecewise linear functions [279, 327, 338], in which the transcriptional activity of a promoter is determined by logoid regulation functions. The logoid functions depend linearly on the transcription factor concentration between two threshold values and are constant below and above these [96, 331].

Another approach is to model the probabilities that the regulatory regions of a gene are occupied by transcription factors. One of the first attempts was an ODE approach proposed by Goodwin [159]. It describes a negative feedback of a metabolite, created by the gene product, R, on the transcription of its own mRNA, M. Both mRNA, protein, and metabolite levels are modelled explicitly with production and decay terms. In the version adapted by Griffith [162], the protein directly functions as the repressor. Repressor binding is assumed to depend on its concentration, [R], in form of a Hill function (equ. (8)). The positive Hill exponent n allows for cooperative effects of repression.

$$\frac{d[\text{M}]}{dt} = a_0 \frac{K}{K + [\text{R}]^n} - d_M[\text{M}] \tag{21}$$

$$\frac{d[\text{R}]}{dt} = k_{tl}[\text{M}] - d_p[\text{R}] \tag{22}$$

In this d_M and d_p are first order decay constants for mRNA and proteins, respectively, k_{tl} is a first order translation constant, and K can be seen as a macroscopic dissociation constant for repressor binding. a_0 is the transcription rate in the unrepressed case.

In case of a transcriptional activator, A, a similar Hill like function has been

suggested [163]:

$$\frac{d[\text{M}]}{dt} = a_{max}\frac{[\text{A}]^n}{K + [\text{A}]^n} - d_M[\text{M}] \tag{23}$$

In this case, a_{max}, stands for the rate of the fully activated transcription. A mechanistic assumption for this kind of model, is that the Hill function represents the fraction of promoter bound by the transcription factors [162]. A similar result was obtained and tested against experimental results later for the induction of the lac operon [444], and coupled with delay differential equations for the trp operon [49]. This approach of using Hill functions has been widely used for both prokaryotic [108, 138, 221] and eukaryotic transcription [192, 314, 381, 399]. A study on a wide range of prokaryotic promoters, showed that Hill functions, or similar formed rational functions, in combination with a basal transcriptional level, gave adequate fits to experimental data [213].

A more mechanistic description of the binding of transcription factors to the regulatory regions of a gene requires detailed knowledge of the binding sites and thermodynamics. Shea and Ackers [376] proposed a model for transcriptional regulation during the switch of the λ-phage from lysogenic to lytic mode. Using equilibrium statistical thermodynamics as an approximation and detailed energetic data, they derived expressions for the probabilities of an operator to exist in a specific microscopic state bound to the different transcription factors and RNAP [1, 376]. The equilibrium probabilities of the states containing RNAP can be combined with a rate constant for the isomerisation of the closed complex into the open, productive form, thereby relating fractional occupancies of RNAP with a rate of transcription initiation. A key assumption for this approach is, that the complex assembly at genes' promoters is fast compared to transcription, and that transcription factor binding is close to the thermodynamic equilibrium.

The framework of Shea and Ackers has been widely used and validated in prokaryotic systems, and various expressions have been derived for different regulatory architectures, including complex mechanisms such as DNA looping and cooperativity [46, 140]. It has also been adapted to eukaryotic systems, for example to explain expression patterns in *D. melanogaster* embryonal development [205, 372, 453], and synthetic and genomic promoters in yeast [143]. Various adaptions have been incorporated into the original model to account for the

more complex mechanisms found in eukaryotes, such as synergistic activation, and quenching, as well as short-range effects of transcriptional repressors [176].

As this approach allows for the prediction of the probability of initiation of transcription, it can directly be employed in stochastic frameworks. In this context, the rate constant for initiation has to be transformed into a propensity constant for initiation [275]. The approach of using a partition function to obtain the probability of transcription initiation was for example used to model the λ phage lysis-lysogeny switch [13].

An even more detailed approach used in stochastic modelling, is to explicitly include all binding reactions of transcription factors to the promoter regions. Apart from a detailed model of the promoter architecture, this approach also needs kinetic rate constants, which are much harder to estimate or measure than thermodynamic binding constants. Still, it can be a way for a fast approximation of stochastic behaviour and to fathom robustness of a network architecture [108, 413].

1.5.4 Postranscriptional Regulation

The expression of genes is not only regulated at the level of transcription, but can be influenced at nearly all the subsequent stages on the way to the final gene product.

In prokaryotes transcriptional attenuation, regulation by transcription termination or anti-termination, plays an important role of the expression of various tRNA genes and biosynthetic operons, such as the *trp* operon of *E. coli* [447, 455]. Commonly in transcriptional attenuation, the nascent RNA can form two alternative secondary structures, one of which leads to termination of transcription and premature release of the transcript. In the case of *trp* operon, the formation of the terminator loop depends on the translation of a small leader peptide and the availability of tryptophan charged tRNA. If tryptophan levels are low, translation of the leader peptide stalls, the antitermination rather than the termination loop forms, and transcription of the full length mRNA is presumed.

Translational attenuation is an analogue process acting on the fully transcribed mRNA. In this form of regulation again two alternative RNA structures can be formed, one occluding the main ribosome binding site. Some ribo-switches have

been found to control gene expression via this mechanism, for example the RFN elements in various bacterial genes encoding for riboflavin synthesis. At high concentrations of flavin mononucleotide (FMN), a coenzyme synthesised from riboflavin, the RFN element binds FMN and the mRNA folds in a way rendering the ribosome binding site inaccessible, thus inhibiting expression of the gene. If FMN occurs at low levels, an alternative mRNA structure forms, ribosomes can bind, and the gene is expressed [308, 429].

A family of small, non-coding RNAs, called micro RNAs (miRNAs), also has been found to be of great importance in eukaryotic gene regulation. They can form helices with partially complementary regions in mRNAs, and are known to lead to post transcriptional gene silencing via at least two different mechanisms. Recruiting a protein complex to the bound mRNA, they can either lead to mRNA cleavage or to repression of translation [175].

In addition to these few exemplary mechanisms, many other have been described. Some act on mRNA stability, such as in the case of the mRNA of the areA gene in *A. nidulans* [328]. Also in eukaryotes the processing of RNAs in the nucleus, such as RNA splicing and polyadenylation, as well as transport to and in the cytoplasm, can be a target for expression control [4, 22].

1.5.5 Network Motifs

In natural genetic regulatory networks certain interaction patterns between transcription factors and their target genes have been found to occur with a much higher frequency than would be expected in random networks. These interaction patterns, or network motifs, have been first described in *E. coli* [282, 377] and in *S. cerevisiae* [243]. Similar motifs have been found in various higher eukaryotes [57, 309], leading to the assumption, that they are elementary units of genetic control, and since then, they have been intensely studied both computationally and experimentally.

The simplest and, in prokaryotes, most common, motif is autoregulation, in which a transcription factor directly binds to the regulatory regions of its own gene. While in prokaryotes more than half [404] of all genes encoding transcription factors are autoregulated, autoregulation seems to be significantly less common in yeast (~10% of genes encoding for transcription factors) [243]. In prokaryotes

the main part of transcriptional autoregulatory feedback loops are negative [404]. Direct negative autoregulation has been predicted [355, 433] and found to lead to faster response times in gene expression [302], and to stabilise transcription factor levels and reduce their noise [35]. Direct positive autoregulation, on the other hand, has been shown to lead to a delayed response. Further, it can shift a sigmoid to a graded response, or conversely, create a switch with hysteresis [262]. Another explanation for the prevalence of negative feedback in prokaryotic systems lies in the demand theory of gene regulation by Savageau [355, 357, 358]. Amongst others it predicts that genes, that are not needed to be expressed at high levels under normal environmental conditions, are selected to be under negative control. Purely transcriptional feedback loops with more than two or three components were only found in yeast and in other eukaryotes, but not in *E. coli* [243].

Another common motif found in both pro- and eukaryotic cells are feed forward loops (FFL) [266]. In a transcriptional FFL, two transcription factors both control a common target gene and one of the transcription factors, the master regulator, also regulates the expression of the other, secondary one. As each of the three direct regulatory interactions involved can be activating or repressing, there exist eight different forms of FFLs. If the direct regulation of the target gene by the master regulator is of the same sign as the indirect one via the secondary regulator, the FFL is called coherent, if the direct and indirect regulation have opposing effects, incoherent. In both *E. coli* and *S. cerevisiae* mainly one form each of coherent and incoherent FFLs have been found [266]. The by far most frequent form of coherent FFL is the all activating one, while the most common incoherent FFL consists of one transcription factor activating transcription of both the target gene and the secondary transcription factor, which in turn negatively controls expression of the target gene.

Coherent FFL have been argued to function as a sign sensitive - depending on whether they are activating or repressing - delays in transcriptional response to a stimulus increasing the activity of the master regulator [266]. As such they can also be seen as filters to only respond to persistent, and not to short time, stimulation. Furthermore, the most common type of coherent FFL - the all activating one - shows an increased apparent cooperativity of target gene expression for low levels of stimulation when compared to a simple cascade of transcription factors.

The time delay function of a coherent FFL has been experimentally verified in *E. coli* using the araBAD operon of the l-arabinose utilisation system [268]. Expression of the araBAD operon is activated by both the CRP and the araC transcription factors. Further, the transcription of araC is positively regulated by CRP. Compared to operons just regulated by CRP or araC alone, transcriptional response from the araBAD operon is notably delayed, leading to a delayed onset of expression. Similarly, a coherent FFL has also be shown responsible for a prolonged response in the expression of *E. coli* flagella proteins [211].

Incoherent FFL have been found to increase response time to stimuli compared to simple direct regulation or linear chains of transcription factors in computational studies [266, 377]. *In vivo* this has been verified for the crp-galS-galETK system in *E. coli*. In this system the galETK operon is activated by CRP and repressed by GalS. CRP as the master regulator additionally activates expression of GalS. CRP and GalE are upregulated under glucose depletion, but the speed of the relative increase of expression of GalE depends on the repression by GalS [267].

Another potential function of the incoherent FFL is creation of only a short pulse of target gene expression for a defined time after activation [266]. A synthetic incoherent FFL using the LuxR transcription factor as an activating master regulator for both the λ-repressor cI and GFP as a reporter gene, which was in turn also repressed by cI, gave defined pulses of GFP expression for continuous activation of LuxR by an inducer [31].

Other commonly found motifs are single input (SIM) and multi input motifs or dense overlapping regulons (DORs) . In SIMs one transcription factor controls a set of different genes or operons, and, in *E. coli* often its own expression [377]. The function of SIMs could be temporally coordinated expression of groups of genes, as seen in the expression of genes for arginine synthesis, which are controlled by the ArgR repressor. At low levels of arginine, these genes are expressed in an ordered temporal sequence [451].

In DORs a set of transcription factors control a set of genes together. One function of DORs could be the coordination of the expression of different gene combinations for various growth and environmental conditions, such as for carbon source utilisation, aerobic and anaerobic growth, and different kinds of stress, in which multiple signals lead to different responses [243, 377].

Another motif found in yeast, are regulatory chains consisting of three or more transcription factors. One example for such a chain are transcription factors involved in the yeast cell cycle. In this case transcription factors in one stage of the cell cycle which are controlling the expression of transcription factors needed for the transition into the next stage form a linear chain [243].

1.6 Model Building

Building a mathematical representation of a biological system seldom is a straightforward, linear process. Even in well described systems many parts are only ill defined, and some completely unknown. To generate accurate predictions, the model needs to be validated against experimental results and often adjusted and refined in an iterative process (see Fig. 6).
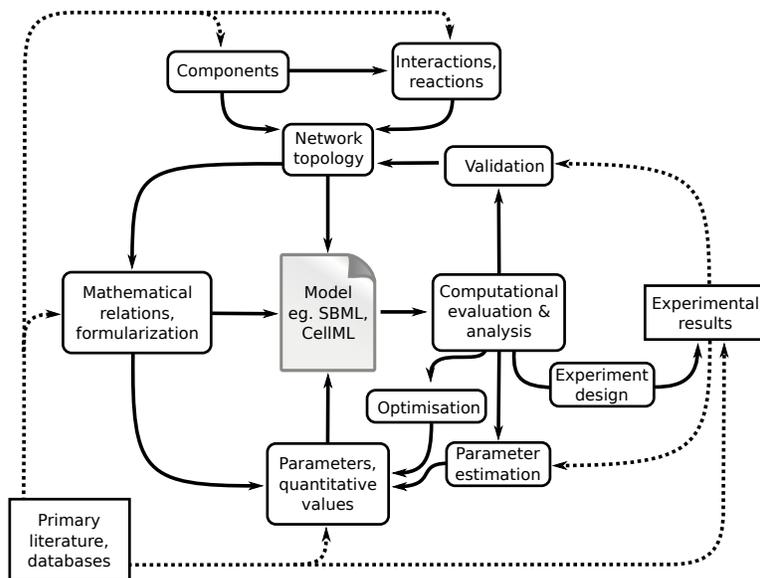
Figure 6: Scheme of an iterative model creation process. External inputs come from literature, databases or experiments. Validation against, and parameter estimation to experimental data, give direct feedback on the model, while the predictions can also be used to guide experimental design. (taken from [110]and modified)

Prior to the construction of a detailed kinetic model, identification and listing of key interacting species, and other observables of the system in question is necessary. In addition, their known regulatory interactions and chemical reactions have to be gathered from the scientific literature and databases.

The kind of databases to scan depends very much on the system to be modelled.

An overview of resources for modelling is given in [298] and [435]. For biological pathways and reactions, KEGG [212], Reactome [273], PANTHER [280] and the Meta- and BioCyc [71] databases are especially useful, as they allow data export in computer readable formats.

Another possibility is to start with existing models accessible through databases and repositories such as BioModels Database[4] [247], JWS Online [389], the CellML model repository[5] [254], the Database of Quantitative Cellular Signalling [383], or ModelDB [184]. The models stored in these databases not only give an idea of the species and their interactions, they can also be excellent starting points for new modelling efforts.

To create a quantitative model, values for the various constants and parameters used in mathematical relations have to derived. While there exists a vast number of experimentally derived values in the scientific literature, it is often hard to find the relevant ones in the multitude of publications.

For enzyme-catalysed reactions, there exist various databases helping to identify the appropriate values. Two databases providing kinetic parameters are BRENDA[6] [75] and SABIO-RK[7] [437]. Both offer a wide range of parameters and reactions extracted from primary literature, with powerful search options. SABIO-RK additionally offers the mechanism assumed in the original source and the ability to export reactions in SBML format. Parameter specific text-mining for searching the primary literature is offered by KMedDB [170]. It allows the searching of PubMed abstracts for various kinetic parameters in combination with compound, organism, or enzyme reaction identifiers. Further information on the thermodynamics of biological reactions is available at the TECRDB [154]. Another helpful source of general interaction parameters is given by the Kinetic Data of Bio-Molecular Interaction Database (KDBI) [233].

Most of the parameters derived from the literature can nevertheless only be taken as guideline values for modelling. If measured time-series or steady state data exist for the system to be described, several algorithms have been implemented for parameter estimation and refinement [reviewed in 88, 278, 285]. Similar methods can be used to optimise a system, for example for engineering metabolic

---

[4] `http://www.ebi.ac.uk/biomodels`
[5] `http://models.cellml.org/cellml`
[6] `http://www.brenda-enzymes.org/`
[7] `http://sabio.villa-bosch.de/`

systems [23].

Another interesting, complementary approach is inverse bifurcation analysis [257]. It can be used to find parameter values exhibiting certain qualitative behaviour. For example it can help in finding regimes that display certain kinds of switching behaviour, or creating more robust oscillators with a given topology.

1.6.1   SBML

Biochemical or genetic regulatory networks can be represented in various forms and formats. Traditionally, models were directly encoded in a programming language, often Fortran or C, which considerably hinders using different tools.

As dedicated tools, targeted at different tasks, often use distinct entry formats, their use requires converting or reimplementation of the original models into compatible formats. Manual conversion of a model from one format to another can be a very tedious and error prone task, especially for bigger systems. To facilitate exchange and reuse of models, the Systems Biology Markup Language (SBML) [8] [200] and the Cellular Markup Language (CellML)[9] [253], have been created.

Both of these languages are XML based and use MathML 2.0 [18] to represent mathematical expressions. CellML is geared more towards physiological, multiscale modelling and, while it allows a greater amount of modularity, it lacks some of the semantics integrated in SBML, such as the distinction between reacting entities, reaction compartments, and parameters. SBML also offers a much broader support by third party tools, which made it the format of choice for this thesis.

SBML is being developed as a community effort, and most of the tools supporting it are freely available. A key to the rapid adoption of SBML by developers is the availability of a general API library, libSBML [56], with bindings to scripting languages such as Perl and Python.

The main tools for model creation and manipulation used for this thesis were CellDesigner [133] and Antimony[10] [386]. CellDesigner is a graphical editor for

---

[8] http://sbml.org
[9] http://www.cellml.org/
[10] http://antimony.sourceforge.net/

biological networks, and was used to create new SBML models, while the model definition language Antimony was used to quickly create modified versions.

1.7 Motivation and Organisation of this Work

While experimental research forms the foundation of biological research, mathematical abstractions and models have become essential to understand the complex systems underlying observed phenomena. Particularly in molecular biology mathematical models of reaction and regulatory networks help to extend knowledge of single interactions and entities to a systems-level. Gene regulatory networks are especially well suited targets for modelling as they are experimentally accessible and easy to manipulate. Furthermore, gene regulatory networks can be synthetically created *in vivo* and used to test predictions or implement systems with specific behaviours [390].

In the second chapter of this thesis the qualitative behaviour of two types of auto-regulatory gene networks is inspected in detail using analytical bifurcation analysis. While the use of such analytical methods in general is limited to smaller systems, they can assist in the elucidation of the qualitative behaviour over large ranges of parameter space. To examine the behaviour under random fluctuations, stochastic versions of the systems are created. The analytical results are employed to obtain parameter values leading to sustained oscillations in the stochastic versions.

The third chapter deals with the possible implications of gene duplication on the qualitative behaviour of a simple gene regulatory system. First a model of a small network formed by GATA-type transcription factors, central in nitrogen catabolite repression in yeast, is created and validated to obtain approximate parameter values. A model of a sub-module of this network - a single autoactivating GATA type transcription factor - is then used to study the effects of gene duplication and dosage effects. Further, topologies of potential gene regulatory networks and modules consisting of GATA-type transcription factors in other fungi are derived using sequence-based approaches and compared.

In the fourth chapter a novel *in silico* cell model is introduced and discussed. The model is fully self-contained and could be used for the study of evolution of gene-regulatory networks.

2 Repressilator-like Gene Regulatory Networks

2.1 Introduction

In recent years further interest in the understanding of gene regulatory systems has stemmed from Synthetic Biology [110, 178]. One of the targets of Synthetic Biology is the implementation of genetic circuits that fulfil certain design specifications and functions. Theoretical analysis of the behaviour of such genetic circuits is essential to prune possible designs and to find the most appropriate and promising ones.

Using transcriptional feedback, both switches and oscillators have been implemented in the last decade [108, 111, 138, 221, 381, 397]. These systems not only prove the usefulness of theoretical analysis, but also pose an excellent testing ground for the validation of mathematical models and their predictions. Based on theoretical predictions such a negative feedback cycle was designed by Elowitz and Leibler [108]. This cycle consisted of three transcriptional inhibitors encoded on a single plasmid. When transformed into *E. coli* this so called "repressilator" gave rise to oscillations in living cells.

Inspired by this work, an in depth and generalized analysis of repressilator-like systems was performed as part of this thesis. In general terms, a repressilator can be defined as a cyclical gene regulatory network in which each gene G_i encodes for a transcription factor that represses the expression of its succeeding gene G_{i+1} in the cycle:

$$G_1 \dashv G_2 \dashv G_3 \dashv \ldots \dashv G_n \dashv G_1 \tag{24}$$

This study builds on previous analytical efforts by extending the cycles of transcriptional inhibitors to an arbitrary number of genes with varying strength of repressor binding. To simplify the equations only symmetrical cases of genes and inhibitors, having identical expression and binding parameters, were considered. Also protein dilution by cell growth and varying gene copy numbers over the cell cycle were ignored. In the equations, both mRNA and protein concentrations were explicitly modelled. Repressor binding is modelled as an equilibrium process under the assumption that it is relatively fast compared to mRNA and protein

synthesis and decay. However, by explicitly correcting for bound and unbound protein, the common assumption of excess transcription factors was avoided, thus ensuring the validity of our expressions at low protein concentrations.

In this analysis two repressilator systems, with different types of regulation, were considered. First, a model similar to the ones described before in the literature [24, 108, 384, 417], with repressors exhibiting cooperative binding characteristics and leaky repression, was analysed. Second, a system with essential auto activation and non-leaky repression was studied.

2.2 Mathematical Formulation and Basic Assumptions

The systems considered can be regarded as mutually dependent reactions. Each gene possesses a regulatory region to which modulators of transcription can bind. Depending on the nature and position of the transcription factors bound to the regulatory region, RNA polymerase may be recruited to the promoter region of the gene to initiate mRNA transcription. mRNA serves as a template for protein translation, and subsequently both mRNA and proteins are degraded via various processes and are diluted by cell growth and division.

These reactions can be modelled to different degrees of detail. However, to obtain a reasonably accurate, yet still tractable mathematical representation, some simplifying assumptions have to be made:

(a) Gene copy numbers are assumed to be constant.

(b) The rates of transcription and translation are not limited by the availability of polymerases, ribosomes, nucleotides or amino acids.

(c) The binding of proteins to the regulatory region of genes is much faster than transcription and translation and is assumed to be near equilibrium.

(d) The multi-step processes of protein translation and degradation can be assumed to be reactions with first order kinetics. Proteins in complex with DNA are considered to be protected from degradation.

(e) Transcription velocity has a clearly defined upper limit and its actual value depends on the binding state of the regulatory region only.

Gene expression is inherently a stochastic and discrete process [64, 301], however, to allow for easier general analysis, a continuous deterministic approach in the form of first order ODEs was employed here. Under these simplifying assumptions the expression of gene i can be described by the following system of equations:

$$\frac{\mathrm{d}m_i}{\mathrm{d}t} = k_{TS,i}\, a_i - k_{Dm,i}\, m_i \tag{25a}$$

$$\frac{\mathrm{d}\bar{p}_i}{\mathrm{d}t} = k_{TL,i}\, m_i - k_{Dp,i}\, p_i \tag{25b}$$

with

$$a_i = a_i(\mathbf{p}) \tag{25c}$$

$$\bar{p}_i = \bar{p}_i(\mathbf{p}) \tag{25d}$$

Here the concentration of mRNA transcribed from gene i is denoted by m_i. The total concentration of the translated protein is symbolised by $\bar{p}_i$ and its unbound, freely available fraction by p_i. $k_{TL,i}$ and $k_{TS,i}$ are the rate constants of transcription and translation respectively, while the rate constants for mRNA and protein degradation are denoted by $k_{Dm,i}$ and $k_{Dp,i}$. Finally a_i represents the transcriptional activity of gene i.

Both the transcriptional activities, a_i, and the total protein concentrations, $\bar{p}_i$, depend on the concentrations of all free proteins, due to transcription factors binding to the regulatory regions of each gene. The number of genes per cell is discrete and the transcription of each gene occurs, at most, in as many levels as there are different binding states of the regulatory region. A continuous function for transcriptional activity can be justified by viewing it as an average over time periods which are long compared to the time scale of regulator binding, but which are short in relation to the duration of transcription. It has to be mentioned, however, that, while providing easily tractable mathematical expressions, this approach neglects noise and stochasticity, which have been found to be of major importance in the regulation of gene expression [109].

The more interconnected the gene regulatory network, the more distinct proteins influence each transcriptional activity. For the simple cyclical repression systems

analysed in this work, a_i depends on p_{i-1} for the pure repressilator and on p_i and p_{i-1} for the system with auto-activation. The total protein concentration, $\bar{p}_i$, on the other hand is a function of p_i in the first example and of p_{i-1}, p_i and p_{i+1} in the latter case.

Following the example of Elowitz and Leibler [108] in their mathematical model of the repressilator, and to further simplify the equations, only cases of genes with identical parameter values were considered in the current study:

$$
\begin{aligned}
k_{TS,i} &= k_{TS}, \quad k_{TL,i} = k_{TL} \\
k_{Dm,i} &= k_{Dm}, \quad k_{Dp,i} = k_{Dp} \\
\bar{g}_i &= \bar{g}
\end{aligned}
\tag{26}
$$

Using the framework set by Eqs. (25) two different cases of regulation are considered:

RepLeaky : a repressilator with leaky repression (described in detail in sect. 2.2.1 page 50)

Transcriptional repression in this system is not complete, but only lowers the transcriptional activity to a basal rate - a process described as "leaky transcription". Transcription occurs at the highest rate when the gene is free. Repressor binding is modelled to be cooperative using Hill functions [183].

RepAuto : a repressilator with auto activation (described in detail in sect. 2.2.2 54)

Transcriptional repression in this system is considered to be complete, such that genes bound by the repressor are not transcribed at all. Furthermore, each protein functions as an essential transcriptional activator for its encoding gene, a process known as "auto activation". Consequently each gene is only transcribed when it is *both* unoccupied by repressors *and* bound by its own gene product. The binding of the auto-activator and inhibitor can affect each other, exhibiting cooperativity.

2.2.1 *RepLeaky*

In this case the regulatory region of each gene is exclusively bound by the proteins encoded by its predecessor in the cyclical gene regulatory network. Each gene G_i can be bound by up to b transcriptional inhibitors P_{i-1} to form a gene-repressor complex $G_i \cdots (P_{i-1})_b$ or $C_i^{(b)}$:

$$G_i + b\,P_{i-1} \rightleftharpoons C_i^{(b)} \tag{27}$$

Complex formation can be envisaged by different mechanisms, the extreme cases being either a stepwise process:

$$G + b\,P \rightleftharpoons C^{(1)} + (b-1)\,P \rightleftharpoons C^{(2)} + (b-2)\,P \rightleftharpoons \dots \rightleftharpoons C^{(b)} \tag{28}$$
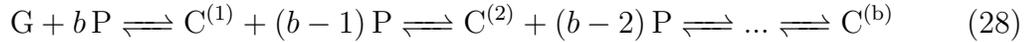
or the formation of a protein multimer with successive DNA binding:

$$b\,P \rightleftharpoons P_2 + (b-2)\,P \rightleftharpoons P_3 + (b-3)\,P \rightleftharpoons \dots \rightleftharpoons P_b$$
$$G + P_b \rightleftharpoons C^{(b)} \tag{29}$$
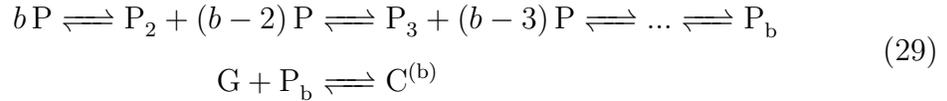
Under the assumption (d) (section 2.2), that regulator binding is fast compared with the rates of gene transcription and mRNA translation, a mathematical expression for the relative amount of gene G bound to b molecules P can be derived using rapid equilibrium kinetics. In general the amount $c^{(b)}$ of complex $C^{(b)}$ is a rational function of the total gene concentration, $\bar{g}$, and the free protein concentration p:

$$c^{(b)}(p) = \bar{g}\,\frac{A(p)}{B(p)} \tag{30}$$
$$\text{with:} \quad B(p) \geq A(p)$$

where A and B are polynomials in p of degree b.

By contrast, the simplest mechanism for complex formation would be the case

of single-step binding, following a reaction equation as given in Eq. (27). With the dissociation constant K for this reaction defined as:

$$K = \frac{g_i p_{i-1}^b}{c_i^{(b)}} \tag{31}$$

and using the mass conservation relation for the constant total gene concentration:

$$\bar{g} = g_i + c_i^{(b)} \tag{32}$$

a functional relationship between $c_i^{(b)}$ and p_{i-1} can be derived:

$$c_i^{(b)}(p_{i-1}) = \bar{g} \frac{p_{i-1}^b}{K + p_{i-1}^b} \tag{33}$$

While the reaction mechanism underlying this expression is rather improbable, it can be regarded as a useful approximation to more complex and realistic mechanisms, especially in its purely empirical form as derived by Hill [183] in the context of oxygen binding to haemoglobin. In this form the exponent b is replaced by the Hill coefficient h, which can be non-integral and, in general, has the number of proteins in the complex, b, as an upper limit. The overall dissociation constant K is replaced by its hth root, $\tilde{K}$, which has the units of a concentration.

Using this Hill function the concentration of $C_i^{(b)}$ as a function of P_{i-1} can be written as follows

$$c_i^{(b)}(p_{i-1}) = \bar{g}_i s\left(\frac{p_{i-1}}{\tilde{K}}\right) \tag{34}$$

in which:

$$s(x) = \frac{x^h}{1 + x^h} \tag{35}$$

The amount of free and bound gene, g_i and $c_i^{(b)}$ respectively, together determine the transcriptional activity of gene i, a_i. Defining the ratio of the activities of the fully repressed gene, $C^{(b)}$, to the repressor free gene, G, as the leakiness δ,

the following formulation can be derived for the relative transcriptional activity:

$$a_i = g_i + \delta c_i^{(b)}$$
$$= (1 - \delta)g_i + \delta \bar{g}_i \tag{36}$$

Mass conservation for the binding proteins P_{i-1} in the case of single-step binding amounts to:

$$\bar{p}_{i-1} = p_{i-1}bc_i^{(b)} \tag{37}$$

Using equations (34), (36) and (37) both the transcriptional activation, a_i as well as protein concentrations can be expressed as follows:

$$a_i = \bar{g}\left((1 - \delta)\left(1 - s\left(\frac{p_{i-1}}{\tilde{K}}\right)\right) + \delta\right) \tag{38}$$
$$\bar{p}_i = p_i + b\bar{g}s\left(\frac{p_i}{\tilde{K}}\right) \tag{39}$$

This gives a complete model definition and allows the system, previously described by (25), to be solved. However, in order to simplify the model analysis, the system needs to be transformed into a more manageable form. This is achieved, as in Elowitz *et al.* 2000 [108], by rescaling the variables, namely time, mRNA and protein concentrations, and by the definition of new lumped parameters.

Time is rescaled to units of the average mRNA lifetime, $1/k_{Dm}$. Proteins are scaled in equivalents x of their half repression concentration, $\tilde{K}$. mRNA concentrations, y, are expressed in units of the maximal steady state concentrations - again rescaled by $\tilde{K}$ - of the proteins they translate into:

$$\tau = t/k_{Dm}, \quad x = \frac{p}{\tilde{K}}, \quad y = \frac{m\, k_{TL}}{\tilde{K}\, k_{Dp}} \tag{40}$$

The general parameters, $k_{Dm}, k_{Dp}, k_{TL}, k_{TS}$ defining gene expression can also be used to define some characteristic values of the system. For the rescaled system

only two parameters are required to replace these four rate constants. β, the ratio of the life times, or degradation rates, of the mRNAs and proteins, and σ, the ratio of production to degradation rates. These two alone are sufficient to calculate the ratio of mRNAs to proteins.

$$\beta = \frac{k_{Dp}}{k_{Dm}}, \quad \sigma = \frac{k_{TS}\, k_{TL}}{k_{Dm}\, k_{Dp}} \tag{41}$$

Finally, the promoter characteristics are described by the binding equivalents per gene, γ, and the maximal rescaled transcription rate, α:

$$\gamma = b\,\frac{\bar{g}}{K}, \quad \alpha = \gamma\,\sigma \tag{42}$$

With these variable transformations and lumped parameters the ODEs for the expression of each gene of the *RepLeaky* system can be put into a concise and simple form:

$$\frac{\mathrm{d}\bar{x}_i}{\mathrm{d}\tau} = \beta\,(y_i - x_i) \tag{43a}$$

$$\frac{\mathrm{d}y_i}{\mathrm{d}\tau} = \alpha\, f(x_{i-1}) - y_i \tag{43b}$$

where:

$$f(x_{i-1}) = (1 - \delta)(1 - s(x_{i-1})) + \delta \tag{43c}$$

$$\bar{x}_i = x_i + \gamma\, s(x_i) \tag{43d}$$

$$s(x) = \frac{1}{1 + x_i^h} \tag{43e}$$

For the limiting case $\gamma = 0$ this system mathematically corresponds to the ones applied in [24, 384, 417] and [108], though it differs in its derivation from some of these.

2.2.2 *RepAuto*

The expression of each gene, G_i, in the repressilator system with auto activation depends on both its own gene product, P_i as well the product of the preceding gene P_{i-1}. Binding of P_i activates gene transcription, while P_{i-1} - as in the *RepLeaky* system - functions as a repressor. If only a single activator and a single repressor bind to the regulatory regions of the gene, the following two cases have to be distinguished.

(1) Mutually exclusive binding of the two types of regulators

(2) Possibility of formation of a trimeric complex, containing an activator, a repressor as well as the regulatory region.

The mutually exclusive form (1) corresponds to only one shared binding site for activator and repressor per gene. Therefore, only one gene-activator, C_i^A and one gene-repressor complex, C_i^R, can be formed:

$$G_i + P_i \xrightleftharpoons{K_A} C_i^A \tag{44}$$

$$G_i + P_{i-1} \xrightleftharpoons{K_R} C_i^R \tag{45}$$

K_A and K_R represent the dissociation constants of the gene-activator and the gene-repressor complexes respectively.

In the alternative case, (2), where there are two distinct binding sites for either the activator or the repressor, an additional complex, C_i^{AR}, containing both regulators can be formed:

$$C_i^A + P_{i-1} \xrightleftharpoons{K_{AR}} C_i^{AR} \xrightleftharpoons{K_{RA}} C_i^R + P_i \tag{46}$$
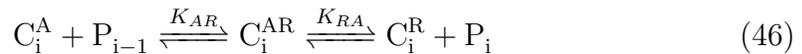
Again K_{AR} and K_{RA} denote the dissociation constants of the complex C_i^{AR} in respect to dissociation of either the repressor or the activator. Following Hess' law, which states that the overall free energy difference ΔG between two substances is independent from the reaction path, the following relationship between

the dissociation constants can be derived:

$$K_A\,K_{AR} = K_R\,K_{RA} \tag{47}$$

This allows for the definition of a cooperativity constant κ, which describes the mutual effect of the regulators on their binding affinities:

$$\kappa = \frac{K_A}{K_{RA}} = \frac{K_R}{K_{RA}} \tag{48}$$

The value of κ represents the interaction of the two proteins in the complex. $\kappa = 1$ indicates that the binding of the two regulators is completely independent, while values greater or smaller than 1 signify positive and negative cooperativity respectively.

Negative cooperativity means that the binding of one of the transcription factors decreases the affinity of the other, with $\kappa = 0$ constituting mutually exclusive binding. Therefore the case of a single binding site can, and will be, treated as a special sub-case of the more general one with two binding sites. As a consequence, different degrees of negative cooperativity can be regarded as varying degrees of steric hindrance by partly overlapping, or closely neighbouring, binding sites. Positive cooperativity - or the enhancement of regulator affinity - can be due to the presence of stabilising interactions between the proteins, either unspecific or via dedicated binding motifs.

Assuming quasi-equilibrium conditions for the binding reactions, the following expressions describe the concentrations of the complexes as functions of the free protein concentrations:

$$c_i^A = \frac{g_i\,p_i}{K_A} \tag{49}$$

$$c_i^R = \frac{g_i\,p_{i-1}}{K_R} \tag{50}$$

and for $\kappa > 0$:

$$c_i^{AR} = \frac{c_i^A \, p_{i-1}}{K_{AR}} = \frac{c_i^R \, p_i}{K_{RA}} = \frac{g_i \, p_i \, p_{i-1}}{K_A \, K_{AR}} \tag{51}$$

Mass conservation of genes and proteins gives the following relations:

$$\bar{g} = g_i + c_i^A + c_i^R + c_i^{AR} \tag{52}$$

$$\bar{p}_i = p_i + c_i^A + c_{i+1}^R + c_i^{AR} + c_{i+1}^{AR} \tag{53}$$

The transcriptional activity of gene G_i in the *RepAuto* system only depends on the amount of activator-gene complex, c^A, formed, since all other variants of gene are assumed to be transcriptionally inactive. Eq. (25d) therefore follows as:

$$a_i = c^A \tag{54}$$

c^A can be described as a function of the free protein concentrations using Eqs. (49),(50),(51) and (52):

$$c_i^A = \bar{g} \, \frac{p_i}{K_A} \left(1 + \frac{p_i}{K_A} + \frac{p_{i-1}}{K_R} + \frac{p_i \, p_{i-1}}{K_A K_{AR}}\right)^{-1} \tag{55}$$

Inserting Eq. (55) and the similar ones for the repressor containing complexes into the mass conservation for P_i (Eq. (53)), the following nonlinear relation between the total and free protein concentrations can be derived:

$$\bar{p}_i = p_i \left[1 + \frac{\bar{g}}{K_A} \left(1 + \frac{p_{i-1}}{K_{AR}}\right) \left(1 + \frac{p_i}{K_A} + \frac{p_{i-1}}{K_R} + \frac{p_i \, p_{i-1}}{K_A K_{AR}}\right)^{-1} \right.$$
$$\left. + \frac{\bar{g}}{K_R} \left(1 + \frac{p_{i+1}}{K_{RA}}\right) \left(1 + \frac{p_{i+1}}{K_A} + \frac{p_i}{K_R} + \frac{p_{i+1} \, p_i}{K_A K_{AR}}\right)^{-1}\right] \tag{56}$$

For the two special cases of either mutually exclusive regulator binding or two completely independent binding sites the equations simplify considerably:

Mutually exclusive binding, ie. the availability of only one binding site for both auto-activator and repressor, implies $\kappa = 0$. In this case all terms divided by K_{AR} and K_{RA} vanish in Eqs. (55) and (56).

By contrast, if the binding of repressor and activator are completely independent from each other, ie. $\kappa = 1$, K_{AR} and K_{RA} become K_R and K_A respectively. The equations describing the total protein concentrations decouple as the p_{i-1} and p_{i+1} terms cancel out:

$$
\begin{aligned}
\bar{p}_i = p_i \Big[1 &+ \frac{\bar{g}}{K_A}\,(1 + \frac{p_{i-1}}{K_R})\,(1 + \frac{p_i}{K_A} + \frac{p_{i-1}}{K_R} + \frac{p_i\,p_{i-1}}{K_A K_R})^{-1} \\
&+ \frac{\bar{g}}{K_R}\,(1 + \frac{p_{i+1}}{K_A})\,(1 + \frac{p_{i+1}}{K_A} + \frac{p_i}{K_R} + \frac{p_{i+1}\,p_i}{K_A K_R})^{-1} \Big] = \\
= p_i \Big[1 &+ \frac{\bar{g}}{K_A}\,(1 + \frac{p_i}{K_A}) + \frac{\bar{g}}{K_R}\,(1 + \frac{p_i}{K_R}) \Big]
\end{aligned}
\tag{57}
$$

Again, to simplify the resulting system of differential equations, both time and the time depending variables have to be rescaled and parameters have to be lumped together in a similar way as previously described with the system *RepLeaky* (Sect. 2.2.1). Time is rescaled by the average mRNA life time, $1/k_{Dm}$, and the parameters β and σ are introduced as above in Eqs. (40) and (41). Protein concentrations, similar to Eq. (40), are expressed in units of the activator binding constant, K_A, and mRNA concentrations are rescaled appropriately:

$$
x = p/K_A \quad , m = \frac{m}{K_A}\,\frac{k_{TL}}{k_{Dp}}
\tag{58}
$$

Additionally the binding equivalents per gene γ and the maximal rescaled transcription rate α are introduced as in Eq. (42) and the dissociation constants are related to each other by the relative repressor affinity, ρ, and the cooperativity constant κ (Eq. (48)).

$$
\gamma = \bar{g}/K_A \quad , \rho = K_A/K_R
\tag{59}
$$

These transformations allow the system *RepAuto* to be expressed succinctly similar to Eqs. (43). For the expression of each gene G_i the following rescaled equations can be derived:

$$\frac{\mathrm{d}\bar{x}_i}{\mathrm{d}\tau} = \beta\,(y_i - x_i) \tag{60a}$$

$$\frac{\mathrm{d}y_i}{\mathrm{d}\tau} = \alpha\, f(x_i, x_{i-1}) - y_i \tag{60b}$$

where:

$$f(x_i, x_{i-1}) = \frac{x_i}{1 + x_i + \rho\,x_{i-1} + \kappa\rho\,x_i\,x_{i-1}} \tag{60c}$$

$$\bar{x}_i = x_i \left[1 + \gamma \left(\frac{1 + \kappa\,\rho\,x_{i-1}}{1 + x_i + \rho\,x_{i-1} + \kappa\,\rho\,x_i\,x_{i-1}} + \frac{\rho\,(1 + \kappa\,x_{i+1})}{1 + x_{i+1} + \rho\,x_i + \kappa\,\rho\,x_{i+1}\,x_i}\right)\right] \tag{60d}$$

In the case of binding two independent binding sites without cooperativity, $\kappa = 1$, the relation for the protein concentration, Eq. (60d), simplifies to the uncoupled form:

$$\bar{x}_i = x_i \left[1 + \gamma \left(\frac{1}{1 + x_i} + \frac{\rho}{1 + \rho\,x_i}\right)\right] \tag{61}$$

2.2.3 Elimination of the Total Protein Concentrations

In the form described above, *RepLeaky* and *RepAuto* are not pure ODE systems. Instead they also contain algebraic relations between the free and total protein concentrations which need to be solved simultaneously. The amount of bound protein becomes negligible, if either protein concentrations are high compared with the gene copy numbers, or in the case of low protein binding - or high dissociation - constants, making the distinction between free and total protein unnecessary. For all other cases it can be demonstrated, that the total protein concentrations can be eliminated by replacing them with the free protein ones. In both systems in vector notation, the total protein concentration depends on the concentration of the free protein in the following way (the dot-notation indicates time derivatives):

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \beta\,(y - x) \\ \alpha\,F(x) - y \end{pmatrix} \tag{62}$$

with:

$$F(x)_i = \begin{cases} f(x_{i-1}) & \text{in system } RepLeaky, \text{ see Eq. (43c)} \\ f(x_i, x_{i-1}) & \text{in system } RepAuto, \text{ see Eq. (60c)} \end{cases} \tag{63}$$

While it would be complicated, or even impossible, to analytically express the free protein concentration in terms of the quantity of total protein, the opposite can be achieved via a linear transformation using mass conservation relations. By partially differentiating these relations to the protein concentration the following relation can be obtained:

$$\dot{\bar{x}} = \frac{\partial \bar{x}}{\partial x} \dot{x} = M(x)\, \dot{x} \tag{64}$$

in which $M(x)$ is called the mass transformation matrix. If the mass transformation matrix is invertible, the system can be rearranged to form a true ODE system expressed in terms of the free protein and total mRNA concentrations:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \beta\, M^{-1}(x)(y - x) \\ \alpha\, F(x) - y \end{pmatrix} \tag{65}$$

The mass transformation matrix $M(x)$ can be derived from Eq. (43d) for the system $RepLeaky$ and from Eq. (60d) for $RepAuto$ by partial differentiation. For $RepLeaky$ each entry $M_{ij}(x)$ takes the form:

$$M(x)_{i,j} = \begin{cases} 1 + \gamma\, \frac{h\, x_i^{h-1}}{(1+x_i^h)^2} & \text{if } \quad j = i \\ 0 & \text{otherwise} \end{cases} \tag{66}$$

This matrix is diagonal and therefore invertible.

For the system $RepAuto$ a more complicated, cyclically tri-diagonal transforma-

tion matrix can be obtained:

$$
M(x)_{i,j} = \begin{cases}
1 + \gamma \left(\dfrac{(1+\kappa\,\rho\,x_{i-1})\,(1+\rho\,x_{i-1})}{(1+x_i+\rho\,x_{i-1}+\kappa\,\rho\,x_i\,x_{i-1})^2} \right. \\[2ex]
\qquad \left. + \dfrac{\rho\,(1+\kappa\,x_{i+1})\,(1+x_{i+1})}{(1+x_{i+1}+\rho\,x_i+\kappa\,\rho\,x_{i+1}\,x_i)^2} \right) & \text{if} \quad j = i \\[3ex]
\gamma\,\dfrac{\rho\,(\kappa-1)\,x_i}{(1+x_i+\rho\,x_{i-1}+\kappa\,\rho\,x_i\,x_{i-1})^2} & \text{if} \quad j = i-1 \\[3ex]
\gamma\,\dfrac{\rho\,(\kappa-1)\,x_i}{(1+x_{i+1}+\rho\,x_i+\kappa\,\rho\,x_{i+1}\,x_i)^2} & \text{if} \quad j = i+1 \\[3ex]
0 & \text{otherwise}
\end{cases}
\tag{67}
$$

It can be shown [290], that the diagonal elements of this matrix fulfil the inequality $M_{ii} > |M_{i-1,i}| + |M_{i+1,i}|$ and that $M(x)$ therefore is diagonally dominant and hence invertible. It is also possible to proof, that x maps to $\bar{x}$ in a one-to-one fashion for all $x_n \geqq 0$, allowing the system *RepAuto* to be described in a concise form.

As stated above, for weak repressor binding, $\gamma << 1$, or high free protein concentrations, the mass transformation matrix approaches the identity matrix $\mathbf{I}$, thereby reducing the systems to the simpler form:

$$
\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \beta\,(y - x) \\ \alpha\,F(x) - y \end{pmatrix}
\tag{68}
$$

It can be shown that the stationary points of the exact system, Eq. (65), do not depend on the transformation matrix and therefore are identical between the full and the simplified systems. Equilibrium points have to satisfy the following condition:

$$
\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \beta\,M^{-1}(x)(y - x) \\ \alpha\,F(x) - y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\tag{69}
$$

Multiplication of the equations describing the development of protein concentrations, x, with $M(x)$, allows the system being solved to become independent of the transformation matrix. This means that the consideration of mass conservation is not necessary for the derivation of the number and position of the system's

equilibria. Also, as will be later demonstrated, the influence of $M(x)$ on the stability of these equilibria is often small or negligible, meaning that the reactions and stoichiometries of transcription factor binding do not need to be known in detail to describe the qualitative behaviour of a gene regulatory network. In general the simplified system has been used as a sufficient approximation in most studies on gene regulatory networks, though some interesting differences stem from the exact treatment, as will be described later (Sect. 2.3.2).

2.3 Detailed Analysis of *RepLeaky*

Recapitulating the derivations given above, the system *RepLeaky* is a special instance of the general system described by Eq. (65). The interactions of the different genes are described by $F_i(x) = f(x_{i-1})$, which, in turn, takes the form of Eq. (43c). The transformation matrix $M(x)$ is described by Eq. (66).

2.3.1 Equilibrium Points

As a starting point for a qualitative analysis of systems behaviour a general expression for the stationary, or equilibrium points, has to be derived. After setting the time derivatives of all variables to zero, as in Eq. (69), and multiplying the differential equations describing the development of the protein concentrations by $M(x)$ the following system needs to be solved:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \beta(y - x) \\ \alpha\, f(x_{i-1}) - y \end{pmatrix} \tag{70}$$

Exploiting the cyclical nature of the system and with $x_0 = x_n$ being the last in the sequence of n repressors, the following relation can be derived from the condition $x_i = y_i = \alpha\, f(x_{i-1})$:

$$x_i = (\alpha f(x_i))^n \tag{71}$$

Due to the use of the sigmoid function $s(x)$, $f(x)$ has certain properties which can be exploited to derive the number and values of potential equilibrium points. As x and y represent, albeit rescaled, concentrations, only solutions in the positive

orthant need to be considered. For positive values, $x \geq 0$, $f(x)$ is a monotone decreasing, bounded function with at most, a single inflection point. This is important for the determination of the number of the resulting solutions of Eq. (70), as both monotonicity and boundedness, as well as the maximal number of inflection points, are preserved during exponentiation.

Both boundedness and monotonicity of $(\alpha\, f(x))^n$ follow from the definition of $f(x)$, Eq. (43c), and its derivatives:

$$f(x) = (1 - \delta)\,(1 - s(x)) + \delta \tag{72}$$

$$s(x) = \frac{x^h}{1 + x^h} \tag{73}$$

$$f'(x) = (\delta - 1)\,s'(x) \tag{74}$$

$$s'(x) = \frac{h\,x^{h-1}}{(1 + x^h)^2} \tag{75}$$

For $\delta \leq 1$ and $h > 0$, $f(x)$ starts from $f(0) = 1$ and, with $f'(x) < 0$ decreases monotonically to $\lim_{x\to\infty} f(x) = \delta$. The boundedness does not change by multiplication with α or exponentiation. By contrast, for the monotonicity the sign of the first derivative $((\alpha\, f(x))^n)' = \alpha^n\, n(f'(x))^n$ changes according to n. It is *negative* for odd n, and $(\alpha\, f(x))^n$ therefore monotonically *decreasing*, and vice versa for even n. This behaviour also fits the nature of the feedback loops, which are negative for an odd, and positive for an even number of genes in the cycle.

To prove that $(\alpha\, f(x))^n$ can have no more than a single inflection point, the Schwarzian Derivative, **SD**, of $s(x)$ has to be calculated:

$$\mathbf{SD}(f) = \frac{f'''(x)}{f'(x)} - \frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2 \tag{76}$$

and therefore:

$$\mathbf{SD}(s) = -\frac{n\,h^2 - 1}{2\,x^2} \tag{77}$$

The Schwarzian Derivative of $s(x)$ and therefore of $\alpha f(x)$ is negative, a property which is transferred to the nth iterate, $(\alpha f(x))^n$. As shown above, depending on the value of n, the nth iterate is either monotonically decreasing or increasing for odd and even values of n respectively, such that its derivative is strictly negative

or positive. From the definition of the Schwarzian Derivative it follows that the only extremum of the derivative of $(\alpha f(x))^n$ is a minimum for odd and a maximum for even n. In other words, the nth iterate can only possess, at most, one inflection point.

As for odd values of n the nth iterate of $\alpha f(x)$ is monotonically decreasing, only one critical concentration $x_c = (\alpha f(x_c))^n = \alpha f(x_c)$ can exist. The corresponding equilibrium point is termed the central equilibrium, E_c, as all rescaled concentrations are identical, $x_i = y_i = x_c$, and all genes are equally regulated. Incidentally, E_c is also the fixed point of the first iterate (see figure 7).

For even n, as for the odd case, there exists only one central equilibrium E_c, with the critical concentrations $x_c = x_i = y_i$, satisfying $x_c = \alpha f(x_c)$. Depending on the value of the critical acclivity at E_c, $A_c = |\alpha f'(x_c)|$, two further critical points x_d and x_u appear, with $x_d < x_c < x_u$. These two additional critical points exist when $A_c > 1$. As can be shown by reinserting the solutions for these equilibria into the fixed point equation, these additional critical points are solutions of the second, but not the first, iterate, $(\alpha f(x_{d/u}))^2 = x_{d/u}$ (see figure 7). The corresponding equilibria E_{odd} and E_{even} are given by $x_i = y_i = x_u$ for i odd, and $x_i = y_i = x_d$ for i even for E_{odd}, and vice versa for E_{even}. That is, at E_{odd} all odd genes are upregulated while all even genes are repressed and vice versa at E_{even}.
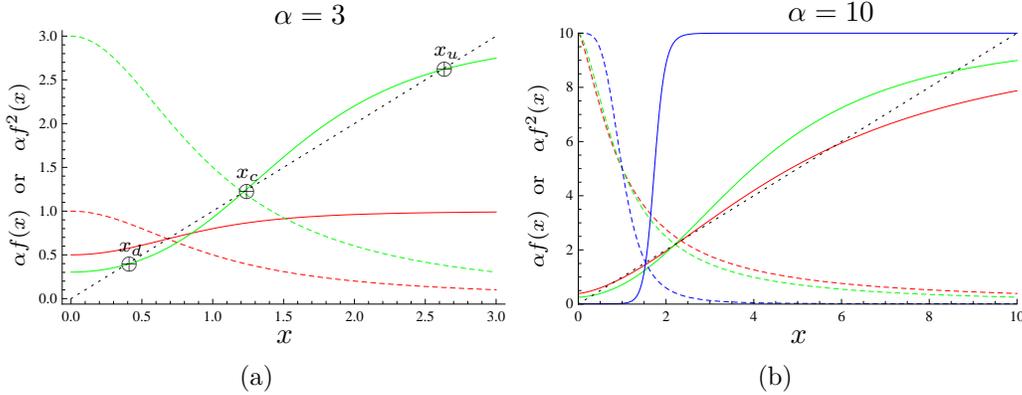
Figure 7: The first, (solid), and second (dashed) iterate of $\alpha f(x)$, for different values of α (a: $\alpha = 3$; b: $\alpha = 10$) and h (a: red: $h = 1.5$, green: $h = 2$; b: blue: $h = 1.4$, red: $h = 1.8$, green: $h = 4$). The intersection points between the first median (dotted) and the iterates indicate the fixed points. At the central fixed point, x_c, the first and second iterate coincide with the median. For $|\alpha f'(x_c)| > 1$ two additional fixed points, x_u and x_d appear, which with increasing cooperativity move towards 0 and α.

As shown in figure 8, A_c shows a strong dependency on α, h and δ. For $\delta = 0$, A_c

increases monotonically with α and reaches a supremum with $\lim_{\alpha \to \infty} A_c(\alpha) = h$. In the case of leaky repression and $\delta > 0$, $A_c(\alpha)$, shows maxima for given values of h with the following properties:

$$A_c^{max} = h \frac{1 - \delta^{1/2}}{1 + \delta^{1/2}} \quad \text{at} \quad \alpha^{max} = \delta^{-(h+1)/2h} \tag{78}$$

This means multiple steady states, and therefore potential bistability, in this system can only exist with at least some sort of cooperativity in repressor binding. Especially for real systems, in which repression is never absolute, even higher values of h, the Hill factor, are required to create a multistable system.
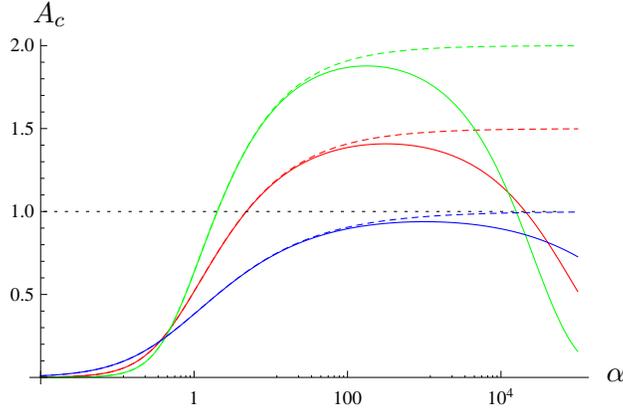
Figure 8: Value of the critical acclivity A_c as a function of α for different values of binding cooperativity (blue: $h = 1$, red: $h = 1.5$, green: $h = 2$) and repression leakiness (solid: $\delta = 10^{-3}$, dashed: $delta = 0$. The dotted line at 1 indicates the criterium for multistability.

2.3.2 Stability Analysis

To analyse the qualitative behaviour of the system *RepLeaky*, the stability of the identified equilibrium points has to be evaluated. This can be achieved by analysing the Jacobian matrix, $J(x)$, and its eigenvalues at these points.

The Jacobian matrix, $J(x)$, using Eqs. (70) and (71), takes the following form at an equilibrium point:

$$\mathbf{J}(x) = \frac{\partial(\dot{x}, \dot{y})}{\partial(x, y)} = \begin{pmatrix} -\beta\, M(x)^{-1} & \beta\, M(x)^{-1} \\ A(x) & -I \end{pmatrix} \tag{79}$$

64

where:

$$\mathbf{A}(x) = \alpha \, \frac{\partial f(x_{i-1})}{\partial x_j} \tag{80}$$

As described in the appendix A on page 161, the eigenvalues of this type of matrix, can be calculated, by solving the characteristic equation of the form:

$$|\mathbf{T}(x)| = 0 \tag{81a}$$

with:

$$\mathbf{T}(x) = \beta \, (1 + \lambda) \, \mathbf{I} + \lambda \, (1 + \lambda) \, \mathbf{M}(x) - \beta \, \mathbf{A}(x) \tag{81b}$$

For the central equilibrium, E_c, $\mathbf{T}(x)$ becomes a bi-diagonal and circulant matrix:

$$T_{i,j} = \begin{cases} T_\Delta = \beta \, (1 + \lambda) + \lambda \, (1 + \lambda) \, M_c & \text{if} \quad j = i \\ T_- = \beta \, A_c & \text{if} \quad j = i - 1 \\ 0 & \text{otherwise} \end{cases} \tag{82a}$$

where:

$$A_c = -A_{i,i-1} = -\alpha \, f'(x_c) = \alpha \, (1 - \delta) \, s'(x_c) \tag{82b}$$

$$M_c = M_{i,i} = 1 + \gamma \, s'(x_c) \tag{82c}$$

Using the nth root of unity, $z_k = e^{i2\pi k/n}$, the resulting equations for the n eigenvalues can be solved to give:

$$\lambda_{k,\pm} = -\frac{1 + \beta/M_c}{2} \pm \sqrt{\left(\frac{1 + \beta/M_c}{2}\right)^2 + (\beta/M_c)\,(A_c\,z_k - 1)} \tag{83}$$

The stability of the central equilibrium is determined by the eigenvalue with the

largest real part, λ_{max}.

$$\lambda_{max} = -\frac{1+\beta/M_c}{2} + \sqrt{\left(\frac{1+\beta/M_c}{2}\right)^2 + (\beta/M_c)\,(A_c\,z - 1)} \tag{84a}$$

in which:

$$z = \begin{cases} 1 & n \text{ even} \\ e^{i\pi/n} & n \text{ odd} \end{cases} \tag{84b}$$

Here again a difference between the behaviour of cycles with even or odd members can be seen.

In the case of an even number of genes in the cycle, the largest eigenvalue is always real. The stability of the central equilibrium depends on the value of the critical quantity A_c. As shown above for the existence of multiple equilibria, the stability of E_c switches at $A_c = 1$. For $A_c < 1$ the largest eigenvalue λ is negative, E_c is the only equilibrium and it is asymptotically stable. As A_c increases, the two additional asymptotically stable equilibria E_{even} and E_{odd} appear and E_c becomes unstable for $A_c > 1$. At $A_c = 1$ the system undergoes a supercritical pitchfork bifurcation. The point of bifurcation exclusively depends on A_c, which is solely a function of α, h and δ (see figure 9). This means that both the repressor binding strength, as well as the correction for repressor binding, do not influence the stability of the system.

It can be proven, that for even n the criterion $A_c < 1$ determines the global behaviour of the system, by showing that it falls within the group of systems for which Theorem 2.1 in [384] holds. To this end, the variables x_i and y_i are divided into two distinct groups, one containing all even, and the other all odd, indexes. The intra-group interactions of the variables can be shown to be positive, while the inter-group interactions are negative:

$$\frac{\partial \dot{x}_i}{\partial y_i} = \beta/M_{ii}(x_i) > 0 \quad \text{and} \quad \frac{\partial \dot{y}_i}{\partial x_{i-1}} = \alpha\,f'(x_{i-1}) < 0 \tag{85}$$

All other off-diagonal entries of the Jacobian matrix (Eq. (79)) are zero. This means, that according to [385], the system converges to certain equilibria. Re-
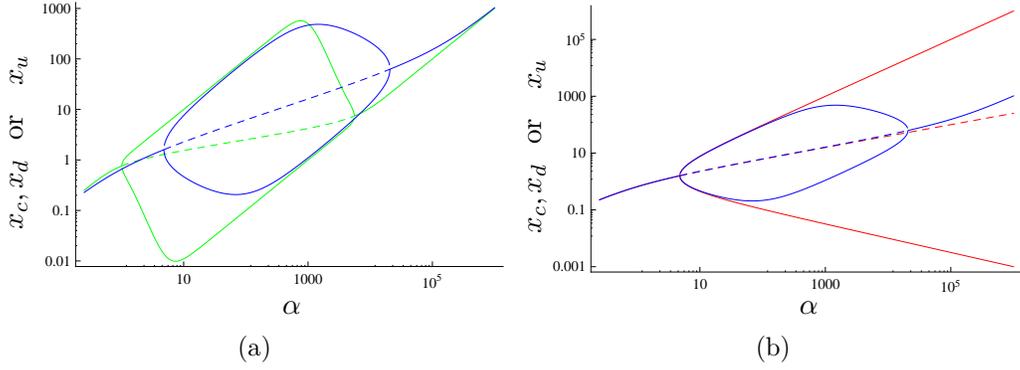
Figure 9: Bifurcation diagram of *RepLeaky* with an even number of genes in dependence of α. (a) shows the effects different degrees of cooperativity h (blue: $h = 1.5$, green: $h = 4$, both $\delta = 0.001$), (b) of leakiness δ (blue: $\delta = 0.001$, red: $\delta = 0$, both $h = 1.5$). Solid lines indicate stable, dashed lines unstable equilibria.

capitulating, for even n and $Ac < 1$, the central equilibrium, E_c, is globally asymptotically stable and all orbits converge to it, while for $A_c > 1$ it becomes unstable and almost all orbits tend to either E_{even} or E_{odd}.

For systems with odd n, as shown above, E_c is the only fixed point. The eigenvalues with the largest real part are a pair of complex conjugates and the system undergoes a supercritical Hopf bifurcation. As E_c becomes unstable, a stable limit cycle appears and the system starts to display oscillations. The sign of the largest Eigenvalues' real parts determines the stability of the system and the following condition can be derived:

$$\Re(\lambda_{max}) < 0 \quad \Leftrightarrow \quad \frac{\beta/M_c}{(1 + \beta/M_c)^2} < \frac{1 - A_c \cos(\pi/n)}{A_c^2 \sin^2(\pi/n)} \tag{86}$$

In this form the condition for stability depends on all the parameters of the system, n, α, δ and h via A_c, β and γ through M_c, making it difficult to interpret. As the left hand side of the inequality in Eq. (86) is positive and always smaller than 1, two simpler, but still sufficient, conditions for stability and instability can be derived:

$$\text{stable: } A_c < \frac{2}{1 + \cos(\pi/n)} \quad \text{resp. unstable: } \quad A_c > \frac{1}{\cos(\pi/n)} \tag{87}$$

For large systems with $n >> 3$, these conditions converge to $A_c \approx 1$ as the bifurcation criterium. This means that it is harder to maintain an equally expressed
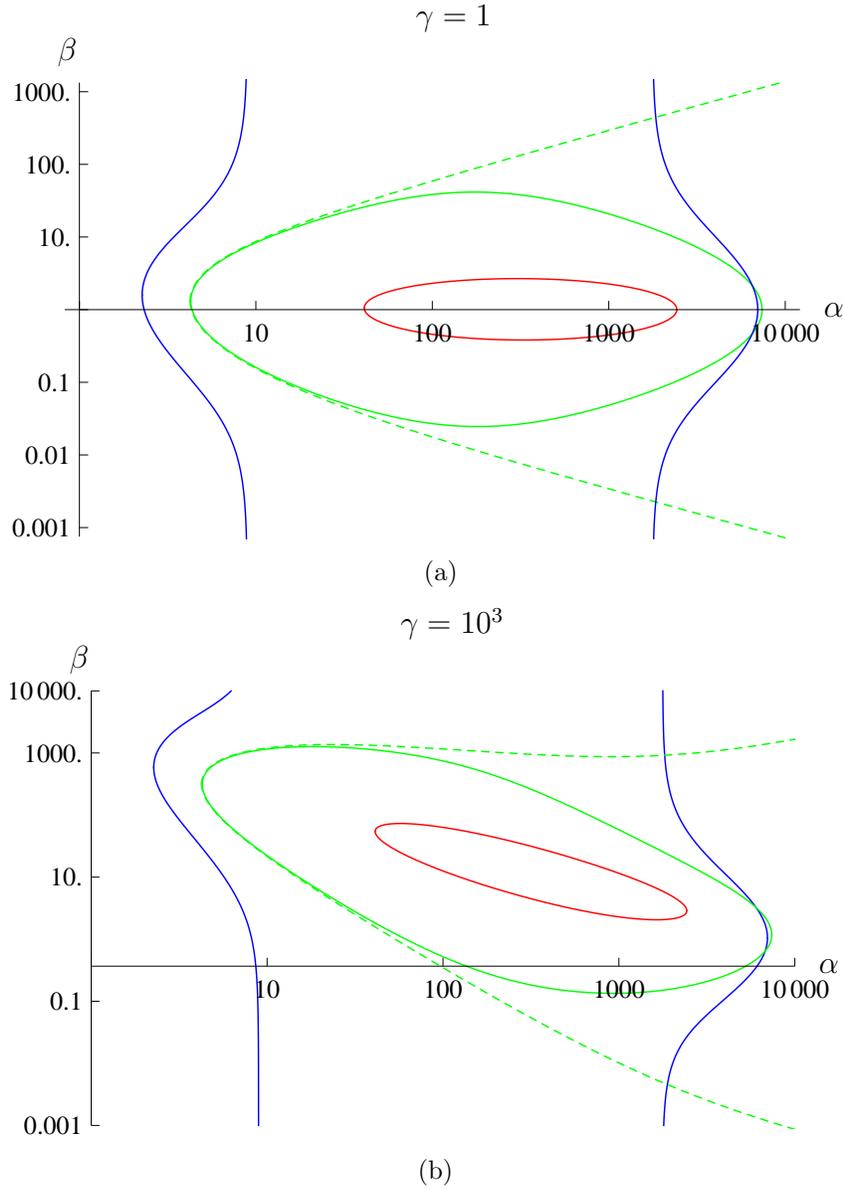
Figure 10: Stability diagram for *RepLeaky* with the smallest odd number of genes, $n = 3$. The dependence of stability of the central (only) equilibrium on α and β is shown for weak ((a) $\gamma = 1$), and strong ((b) $\gamma = 10^3$) repressor binding. Stability boundaries enclose the regions of instability, that is, the regimes of oscillation. Solid lines denote leaky ($\delta = 10^{-3}$), dashed non-leaky ($\delta = 0$) repression, whereas colours indicate different Hill coefficients (red: $h = 1.5$, green: $h = 2$, blue: $h = 2$). (figure as in [290])

state for increasing sizes of negative feed back cycles. Also, it is easier to obtain stable oscillations with larger systems.

Near a Hopf bifurcation the frequency of oscillations, ω, can be approximated by

the imaginary part of the conjugated complex eigenvalue pair, λ_{max}:

$$\frac{\omega}{k_{Dm}} = \Im(\lambda_{\max}) = \frac{\beta/M_c}{1 + \beta/M_c} A_c \sin(\pi/n) \tag{88}$$

This expression contains three multiplicative terms, each depending on different variables. As is to be expected, the bigger n, that is the bigger the system, the slower the oscillations. Higher transcription and translation rates, as well as higher Hill coefficients, h, and on the other hand lead to faster oscillations. The leakiness δ has little to negative influence on frequencies, while the influence of the degradation rates for mRNA and proteins depends on the other parameters values and can be negative or positive. Strong repressor binding leads to slower oscillations.

The Poincaré Bendixon Theorem for monotone cyclic feedback systems derived by Mallet-Paret and Smith [264] can be used to prove that the central equilibrium and the limit cycle determine the global behaviour of the system. If the system is written in terms of the variables $(z_1, z_2, z_3, z_4, ...) = (y_1, x_1, y_2, x_2, ...)$, it is easy to see that it is a monotone cyclic feedback system. First each $\dot{z}_i$ only depends on itself z_i and its predecessor in the cycle z_{i-1}. Second the values of $\frac{\partial \dot{z}_i}{\partial z_{i-1}}$ are iteratively positive or negative, depending on i being even or odd. As the criteria for a monotonicity and cyclical feedback are fulfilled every orbit either tends to the central equilibrium, E_c, or if it is unstable, to a periodic orbit.

2.4 Detailed Analysis of *RepAuto*

The gene-regulatory network with auto activation and cyclical repression is, as the *RepLeaky* system, an instance of the general system (Eq. (65)). However, the transcriptional regulation of each gene, $F_i(x)$, is a function of both x_{i-1} and x_i, $f((x_{i-1}, x_i))$. Further, the entries of the mass transformation matrix, $M(x)$, are more interconnected, as the free protein concentration x_i directly depends on the binding states of the genes G_i and G_{i+1} and therefore on x_{i-1}, x_i and x_{i+1}. This increased complexity gives rise to additional forms of dynamics; $f((x_{i-1}, x_i))$ and $M(x)$ are described by Eqs. (60c) and (67), respectively.

2.4.1 Equilibrium Points

Similar to *RepLeaky*, by setting all time derivatives to zero, the full system (Eq. (65)), can be reduced to a form which is independent of $M(x)$. By solving this system, again a simple condition for all fixed points can be derived:

$$x_i = y_i = F_i(x) = \alpha\, f(x_{i-1}, x_i) =$$
$$= \alpha\, \frac{x_i}{1 + x_i + \rho\, x_{i-1} + \kappa\rho\, x_i\, x_{i-1}} \tag{89}$$

In this system, genes require their own gene product for their expression and no basal rate of transcription is assumed. Contrary to the classical repressilator, *RepLeaky*, this means that a gene can be completely switched off by repression and that all switched off genes stay silent. Therefore solutions of Eq. (89) encompass the origin, O, and stationary points at which some genes are silenced, $x_i = y_i = 0$. All solutions can be described by the following conditions:

$$x_i = 0 \quad \vee \quad x_i = g(x_{i-1}) \tag{90a}$$

where:

$$g(x) = \frac{\alpha - 1 - \rho\, x}{1 + \kappa\, \rho\, x} \tag{90b}$$

The existence of equilibria in addition to the origin, O, depends on the parameter α. For $\alpha < 1$ degradation prevails, and the origin, O, is the only, stable equilibrium. Genes can only be transiently expressed and all concentrations tend to 0. For $\alpha > 1$, O becomes unstable, and additional equilibria come into existence.

If all genes are expressed at equilibrium, that is all $x_{i...n} > 0$, this condition can be resolved as for the leaky repressilator in Eq. (71) via iterative insertion to derive the following criterion:

$$x_{1...n} > 0 \quad \wedge \quad x_i = g(x_{i-1})^n \tag{91}$$

$g(x)$ becomes negative for values of $x > (\alpha - 1)/\rho$, therefore meaningful solutions can only be obtained in the interval $[0, (\alpha - 1)/\rho]$. Within this interval, $g(x)$ is bounded, monotonic decreasing and has no inflection points, properties which also propagate to its higher iterates. As with the system *RepLeaky*, this implies that only one critical concentration x_c can exist on this interval, with $g(x_c) = x_c$. x_c can be derived as the positive solution of the equation

$$\alpha = 1 + (1 + \rho)x_c + \kappa \rho x_c^2 \tag{92}$$

or

$$x_c = \frac{-(\rho + 1) + \sqrt{(\rho + 1)^2 + 4\kappa\rho(\alpha - 1)}}{2 (\kappa \rho)^2} \tag{93}$$

The corresponding central equilibrium, E_c, again has all genes under equal expression with $x_i = y_i = x_c > 0$.

Apart from the central equilibrium, E_c, for $\alpha > 1$ a number of boundary equilibria exist with varying genes G_i switched off, $x_i = y_i = 0$. To describe them, the support $\mathcal{S}$ of a point in concentration space is introduced, which is the set of indexes i of a point with $x_i > 0$. For the boundary equilibria this means that there is a group of silenced genes with $x_i = y_i = 0 \quad \forall i \notin \mathcal{S}$ and genes which are expressed at a level $x_i = y_i = g(x_{i-1}) > 0$. Otherwise, depending on the value of relative repressor affinity, ρ, two possible cases of boundary equilibria, E_b, arise.

For $\rho < 1$, meaning each protein has a higher affinity to bind its own gene as an auto-activator than its succeeding gene as a repressor, each gene still allows its successor to be expressed. In this case the boundary equilibria, E_b, are all

points whose support $\mathcal{S}$ are proper subsets of $N_n = 1, 2, \ldots, n$. For each $\mathcal{S}$ there exists a unique equilibrium $E_{\mathcal{S}}$, with all expressed genes having steady state concentrations of $x_i = y_i = g(x_{i-1})$. There exist $2^n - 1$ such equilibria.

If $\rho > 1$, that is, each transcription factor binds stronger to the successive gene than to its own, each unrepressed gene can effectively switch off its successor. This can be deduced from Eqs. (90): assuming a gene G_{i-1} is not expressed, $x_{i-1} = 0$, its successor G_i is unrepressed and therefore at its maximal expression level $x_i = \alpha - 1$. This expression level is higher than the sustainable critical concentration $x_c = (\alpha - 1)/\rho$ for $\rho > 1$, and does not allow a positive solution to $x_{i+1} = g(\alpha - 1)$, ie. the succeeding gene G_{i+1} is completely repressed. In this case only boundary equilibria $E_{\mathcal{S}}$ exist in which $\mathcal{S}$ is a subset of non-consecutive numbers from N_n. For each of these equilibria $E_{\mathcal{S}}$ no more than half of the genes can be expressed and their expression level is $\alpha - 1$. In total there are L_n of such boundary equilibria, with L_n being the nth Lucas number ($L_1 = 1, L_2 = 3, L_n = L_{n-1} + L_{n-2}$).

Due to the cyclical nature of the systems considered here, for even n two equilibria exist, in which half of the genes are turned on, ie. with maximal support: E_{even}, in which all genes G_i with even, and E_{odd}, in which all genes with an odd index i are expressed. For odd values of n, the maximal number of expressed genes in boundary equilibria is $\frac{n}{2} - 1$. There exist n such maximal boundary equilibria.

2.4.2 Stability Analysis

The Jacobian matrix $\mathbf{J}(x)$ of the system $RepAuto$ is similar to the one described previously by Eq. (79). The differences between them lie in the mass transformation matrix $M(x)$ and also in the partial regulatory response matrix $A(x)$. $M(x)$ is described by Eq. (60d) while $A(x)$ can be derived from Eq. (60c):

$$A(x)_{i,j} = \alpha \frac{\partial f(x_{i-1}, x_i)}{\partial x_j} = \begin{cases} \alpha \frac{1 + \rho\, x_{i-1}}{(1 + x_i + \rho\, x_{i-1} + \kappa\, \rho\, x_i\, x_{i-1})^2} & \text{if} \quad j = i \\[2ex] \alpha \frac{-\rho\, x_i\, (1 + \kappa\, x_i)}{(1 + x_i + \rho\, x_{i-1} + \kappa\, \rho\, x_i\, x_{i-1})^2} & \text{if} \quad j = i - 1 \\[2ex] 0 & \text{otherwise} \end{cases} \tag{94}$$

The stability of the origin, O, as an equilibrium, as mentioned before, depends

solely on the value of the parameter α. The eigenvalue with the largest real part, λ_{max}, can be derived to be the positive solution of the following equation:

$$\lambda_{max} = -\frac{1+\beta/M_0}{2} \sqrt{\left(\frac{1+\beta/M_0}{2}\right)^2 + \beta/M_0 \, (\alpha - 1)} \tag{95a}$$

with:

$$M_0 = 1 + \gamma \frac{1+\rho}{\alpha^2} \tag{95b}$$

For values of $\alpha < 1$, λ_{max} is negative, and O is the only equilibrium point in the positive orthant and constitutes a stable global attractor. As α increases above 1, additional fixed points appear in the positive orthant, and O loses its stability.

To determine the stability of the central equilibrium, E_c, the characteristic equation (81), needs to be solved to derive expressions for the eigenvalues. The matrix $\mathbf{T}(x)$ is circulant at E_c:

$$T_{i,j} = \begin{cases} T_\Delta = \beta\,(1+\lambda) + \lambda\,(1+\lambda)\,M_\Delta - \beta\,A_\Delta & \text{if} \quad j = i \\ T_- = \lambda\,(1+\lambda)\,M_\pm + \beta\,A_- & \text{if} \quad j = i - 1 \\ T_+ = \lambda\,(1+\lambda)\,M_\pm & \text{if} \quad j = i + 1 \\ 0 & \text{otherwise} \end{cases} \tag{96a}$$

where:

$$A_\Delta = A_{i,i} = \frac{1 + \rho\,x_c}{\alpha} \tag{96b}$$

$$A_- = -A_{i,i-1} = \frac{\rho\,x_c\,(1 + \kappa\,x_c)}{\alpha} \tag{96c}$$

$$M_\Delta = M_{i,i} = 1 + \gamma \frac{(1 + \kappa\,\rho\,x_c)\,(1 + \rho\,x_c) + \rho\,(1 + \kappa\,x_c)\,(1 + x_c)}{\alpha^2} \tag{96d}$$

$$M_\pm = M_{i,i-1} = M_{i,i+1} = \gamma \frac{\rho\,(\kappa - 1)\,x_c}{\alpha^2} \tag{96e}$$

Again using the formula for circulant determinants, the characteristic equation

can be factorised:

$$|T| = \prod_{k=0}^{n-1} (T_\Delta + T_- \, z_k + T_+ \, z_k^{-1}) = 0$$

with:

$$z_k = e^{i2\pi k/n}$$

(97)

Setting each factor individually to zero, a quadratic equation for each eigenvalue is obtained:

$$\lambda_{k,\pm} = -\frac{1 + \beta/M_k}{2} \pm \sqrt{\left(\frac{1 + \beta/M_k}{2}\right)^2 + (\beta/M_k)\,(A_k - 1)}$$

(98a)

with:

$$A_k = A_\Delta - z_k \, A_-$$

(98b)

$$M_k = M_\Delta - M_\pm \, 2 \cos(2\pi k/n)$$

(98c)

For the principal stability analysis only the real-parts of the eigenvalues, and of these the maximal ones, are of interest. The solutions of Eq. (98) show a pattern in that $\Re(\lambda_{k,-}) = \Re(\lambda_{n-k,+})$, with the exception of n even and $k = n/2$, where nevertheless $\lambda_{k,-} < \lambda_{k,+}$. This means that the negative branch of the square root does not need to be considered and in the following calculations $\lambda_{k,+}$ is shortened to λ_k.

As in the system *RepLeaky*, the stability of the central equilibrium, E_c, depends on whether the number of genes, n, is odd or even. The critical eigenvalue is the one with the largest real part and the following criterion has to be fulfilled for E_c to be stable:

$$\Re(\lambda_{critical}) < 0 \quad \Leftrightarrow \quad \forall k : \Re(\lambda_k) < 0$$

(99)

If n is even, the expression under the square root in Eq. (98) is maximal for $k = n/2$ as $z_{n/2} = -1$. The critical eigenvalue for even n, λ_{even}, therefore is $\lambda_{n/2}$ and is purely real. The sign of λ_{even}, and with it the stability of E_c, depends on

the sign of $A_{n/2} - 1$.

$$\Re(\lambda_{even}) < 0 \quad \Leftrightarrow \quad A_{n/2} < 1 \tag{100a}$$

with:

$$A_{n/2} = A_\Delta - z_k A_- = 1 + \frac{(\rho - 1) x_c}{\alpha} \tag{100b}$$

In other words for even n, the stability of the central equilibrium, E_c, solely depends on the parameter ρ, the ratio of binding affinities. E_c is asymptotically stable for $\rho < 1$, and unstable for $\rho > 1$. This criterion is identical to the one previously defined for the occurrence of boundary equilibria with only non-consecutive expressed genes (see page 71). E_c undergoes a highly degenerate bifurcation, since the number of boundary equilibria simultaneously decreases from $2^n - 1$ for $\rho < 1$ to L_n for $\rho > 1$. The reason for this abrupt change again lies in the complete switching off of successive genes, which occurs, if the affinity for binding as a repressor becomes stronger than for binding as an auto-activator.

As shown previously for the leaky repressilator with even n (see page 66), for the simplified system without mass conservation for regulator binding, the system's variables x_i and y_i can be divided into two disjoint groups with even and odd indexes. As the inter- and intra-group variables interact negatively and positively respectively, the system falls under the category considered in Theorem 2.1. in [384]. This means that, at least for the simplified system with even n and $\alpha > 1$, the central equilibrium, E_c, is globally asymptotically stable for $\rho < 1$, and almost all orbits converge to the boundary equilibria for $\rho > 1$.

The critical eigenvalue for odd n, λ_{odd}, does not have to be the eigenvalue with the largest absolute real part, as each eigenvalue's sign depends on both A_k and β/M_k. Still it can be shown that the real parts of the eigenvalues λ_k with both $k = (n+1)/2$ and $k = (n-1)/2$ change signs first and therefore determine stability. The sign of the real part always depends on the real part of the solution of the full square root in Eq. (98), which cannot be solved analytically in a general way. As β/M_k is always positive and real, however, the following conditions describe the dependence of the sign of the eigenvalue on A_k exclusively:

$$\Re(\sqrt{A_k}) < 1 \Rightarrow \Re(\lambda_k) < 0 \tag{101}$$

$$\Re(A_k) > 1 \Rightarrow \Re(\lambda_k) > 0 \tag{102}$$

Using the expression for α at x_c, Eq. (92), clearly $0 < A_- < 1$. Together with the definition of A_k, Eq. (98b), $A_\Delta > 0$ and the $\Re(z_k)$ being minimal for $k = (n+1)/2$ and $k = (n-1)/2$ this implies that the following conditions hold for $\lambda_{odd,\pm}$ representing both $\lambda_{(n+1)/2}$ and $\lambda_{(n-1)/2}$:

$$\Re(\sqrt{A_k}) \leq \Re(\sqrt{A_{odd,\pm}}) \tag{103}$$

$$\Re(A_k) \leq \Re(A_{odd,\pm}) \tag{104}$$

Therefore $\lambda_{odd,\pm}$, a pair of complex conjugates, determines the stability of the central equilibrium, E_c, for odd values of n.

$$\Re(\lambda_{odd,\pm}) < 0 \quad \Leftrightarrow \quad \forall k : \Re(\lambda_k) < 0 \tag{105}$$

The values of $\lambda_{odd,\pm}$ are given by Eq. (98) with the following expressions for $A_{odd,\pm}$ and $M_{odd,\pm}$:

$$A_{odd,\pm} = A_{(n+1)/2 \vee (n-1)/2} = A_\Delta + A_- \, e^{\pm i\pi/n} \tag{106}$$

$$M_{odd,\pm} = M_{(n+1)/2 \vee (n-1)/2} = M_\Delta - M_\pm \, 2 \cos(\pi/n) \tag{107}$$

Using these expressions, the stability criterion can be derived by analysing the square root in Eq. (98):

$$\Re(\lambda_{odd,\pm}) < 0 \quad \Leftrightarrow \quad \frac{\beta/M_{odd}}{(1 + \beta/M_{odd})} < \frac{1 - \Re(A_{odd})}{(\Im(A_{odd}))^2} \tag{108}$$

As previously mentioned, the two eigenvalues $\lambda_{odd,\pm}$ are a pair of complex conjugates. When their real part crosses from a negative to a positive sign, the system

undergoes a Hopf bifurcation and a periodic orbit appears. Similar to the leaky repressilator (Eq. (88)), the scaled angular frequency ω of the oscillations close to the bifurcation is proportional to $\Im(\lambda_{odd,\pm})$:

$$\frac{\omega}{k_{Dm}} = \Im(\lambda_{odd}) = \frac{\beta/M_{odd}}{1 + \beta/M_{odd}} \Im(A_{odd}) \tag{109}$$

As seen with the leaky repressilator (Eq. (86)), the stability of the central equilibrium depends on all parameters - with $A_{odd,\pm}$ on n, α, ρ and κ and additionally on β and γ. Again based on the fact that the left hand side of the right inequality in Eq. (108) is always positive and ≤ 1, the following sufficient criteria, independent of β/M_{odd}, for stability can be derived:

$$\Re(\sqrt{A_{odd,\pm}}) < 1 \Rightarrow \Re(\lambda_{odd,\pm}) < 0 \tag{110}$$

$$\Re(A_{odd,\pm}) > 1 \Rightarrow \Re(\lambda_{odd,\pm}) > 0 \tag{111}$$

This means that for a system with a given odd number of genes, n, while the stability of the central equilibrium, E_c, in general depends on all parameters, there exist regions in parameter space in which the parameters α, ρ and κ exclusively determine its stability. These regions of definite stability and instability are bounded by functions defined by Eqs. (110) and (111).

The expression $\Re(\sqrt{A_{odd,\pm}}) < 1$ unfortunately cannot be simplified, so that the boundary of the definite unstable domain, $\Re(\sqrt{A_{odd,\pm}}) = 1$ is given by an implicit function in n, α, ρ and κ:

$$\Re(\sqrt{A_{odd}}) = 1 \quad \Leftrightarrow \quad A_\Delta + \cos(\pi/n)\,A_- + \frac{1}{4}\,\sin^2(\pi/n)\,A_-^2 = 1 \tag{112}$$

The region of definite instability, however, can be explicitly described by the following expressions after some reorganisation:

$$\Re(A_{odd}) > 1 \quad \Leftrightarrow \quad \frac{(\rho - 1/\cos(\pi/n))\,(\rho - \cos(\pi/n))}{\rho} > \frac{(1 - \cos(\pi/n))^2}{\cos(\pi/n)}\,\kappa\,(\alpha - 1)$$

$$\wedge \quad \rho > 1/\cos(\pi/n) \tag{113}$$

For the boundary equilibria at least some solutions of the characteristic equation, $|\mathbf{T}(x)| = 0$, can readily be found. Each silenced gene i produces a row in the matrix $\mathbf{T}(x)$, in which only the diagonal entry T_{ii} is non-zero. The same is true for each gene surrounded by shut off genes. This means, for an equilibrium with support S, that each gene G_i with either $i \notin S$ or $i \in S$ and $(i-1)(i+1) \notin S$ produces a factor in the characteristic equation from which the following pair of eigenvalues can be derived:

$$T_{i,i} = \beta\,(1+\lambda) + \lambda\,(1+\lambda)\,M_{i,i} - \beta\,A_{i,i} = 0 \tag{114}$$

$$\lambda_{i,\pm} = -\frac{1+\beta/M_{i,i}}{2} \pm \sqrt{\left(\frac{1+\beta/M_{i,i}}{2}\right)^2 + (\beta/M_{i,i})\,(A_{i,i}-1)} \tag{115}$$

with for each $i \notin S$:

$$A_{i,i} = \frac{\alpha}{1+\rho\,x_{i-1}} \tag{116}$$

$$M_{i,i} = 1 + \gamma\left(\frac{1+\kappa\,\rho\,x_{i-1}}{1+\rho\,x_{i-1}} + \frac{\rho\,(1+\kappa\,x_{i+1})}{1+x_{i+1}}\right) \tag{117}$$

All these eigenvalues are purely real. Furthermore, for the stability analysis only the positive branch of the square root is of interest, as it is always greater than the negative one.

The sign of the eigenvalue $\lambda_{i,+}$ is determined by the sign of the expression $A_{ii}-1$. As A_{ii} depends on the values of x_i and x_{i-1}, different possible cases have to be considered. For $\rho < 1$ all combinations of switched off and on genes are possible, with x_i ranging from 0 to $\alpha - 1$. For any shut off gene G_i with $x_i = 0$ and $x_{i-1} \leq \alpha - 1$ that implies:

$$A_{i,i} \geq \frac{\alpha}{1 - \rho\,(\alpha - 1)} \tag{118}$$

With $\alpha > 1$ and $\rho < 1$, definitely $A_{ii} > 1$ and therefore there exists at least one positive eigenvalue, $\lambda_{i,+}$, for each silenced gene. In other words, all boundary equilibria are unstable for $\rho < 1$.

In the case of $\rho > 1$, as previously derived for the boundary equilibria, every

expressed gene is surrounded by switched off genes. So three different combinations of x_i and x_{i-1} need to be considered. One for expressed genes, $x_i = \alpha - 1$ and $x_{i-1} = 0$ and two for repressed genes, with $x_{i-1} = 0 \vee \alpha - 1$:

$$
\begin{array}{|c|c||c|}
\hline
x_{i-1} & x_i & A_{i,i} \\
\hline
0 & 0 & \alpha \\
\hline
\alpha - 1 & 0 & \frac{\alpha}{1 + \rho\,(\alpha - 1)} \\
\hline
0 & \alpha - 1 & \frac{1}{\alpha} \\
\hline
\end{array}
\tag{119}
$$

For $\rho > 0$ only the first case needs to be considered, that of two successive, completely shut off genes, gives $A_{ii} > 1$ and therefore yields a positive eigenvalue. This occurs at least once in all boundary equilibria for odd n, indicating that they are unstable. For even values of n, the two equilibria with maximal support S, E_{even} and E_{odd} are the only ones without consecutive silenced genes. These two are the only stable equilibria for $\rho > 1$ and almost all orbits starting from $x_i > 0$ tend towards either of them.

Between the different unstable boundary equilibria, there exist heteroclinic orbits for α and $\rho > 1$. This means that all trajectories starting at, or close to, a boundary converge to a boundary equilibrium. This can be shown for $\gamma = 0$ by considering a starting point with $x_{i-1} = y_{i-1} = 0$ and $x_i + y_i > 0$. Unrepressed both x_i and y_i will tend to $\alpha - 1$. Once $x_i(t) > (\alpha - 1)/\rho$, x_{i+1} will tend towards zero as insertion in Eq. (68) using (60c) shows:

$$
\begin{aligned}
\frac{\mathrm{d}y_{i+1}}{\mathrm{d}t} &= \alpha\, f(x_i, x_{i+1}) - y_{i+1} = \\
&= \frac{\alpha}{\alpha + x_{i+1}\,(1 + \kappa\,\rho\,(\alpha - 1))}\, x_{i+1} - y_{i+1} < x_{i+1} - y_{i+1}
\end{aligned}
\tag{120}
$$

and hence:
$$
\frac{\mathrm{d}(x_{i+1} + \beta\, y_{i+1})}{\mathrm{d}t} = \beta\,(y_{i+1} - x_{i+1}) + \beta\,\frac{\mathrm{d}y_{i+1}}{\mathrm{d}t} < 0
\tag{121}
$$

For $\gamma > 0$ the case is equivalent and both x_{i+1} and y_{i+1} tend to zero. If the gene G_{i+2} is expressed, it tends to its maximal expression $\alpha - 1$, thereby switching off the next gene and so on. In the end, the alternating pattern of expression propagates until the cycle is closed or an unexpressed gene is encountered. From

a starting point at which all genes are expressed, if the gene G_i is turned off the system converges to a boundary equilibrium with the (maximal) support $\{\ldots, i-3, i-1, i+1, i+3, \ldots\}$. In general, for $n \geq 3$, all systems expressing only the genes i and $i+1$ converge against the boundary equilibrium E_i. All such orbits are confined to a 4 dimensional - x_i, y_i, x_{i+1} and y_{i+1} - boundary face, whose closure contains two additional equilibria, the origin, O, functioning as a repellor at $\alpha > 1$, and the saddle E_{i+1}. Moreover, there exists a heteroclinic connection from $E_{i+1} \to E_i$, which runs along a one dimensional manifold. This connection can be imagined as a trajectory starting from E_{i+1} after the addition of an infinitesimal small amount of x_i. Together these heteroclinic trajectories connect all the boundary equilibria of a given type in a heteroclinic cycle, eg. $E_1 \to E_n \to E_{n-1} \cdots \to E_2 \to E_1$.
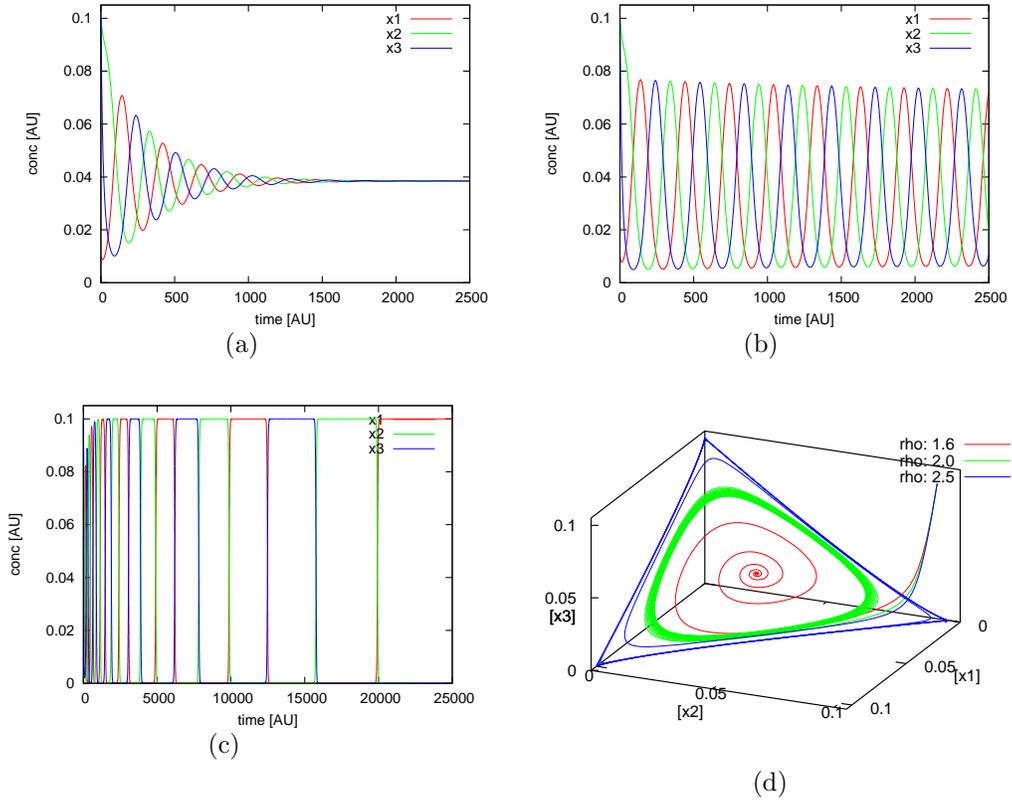
Figure 11: The different dynamical behaviours of the system *RepAuto* for varying values of ρ. Subfigures (a), (b), and (c) show time courses of the system exhibiting mono-stability, limit cycle oscillations, and a stable heteroclinic cycle, respectively. (d) shows a projection of all three trajectories into the protein subspace of phase space. Parameter values: $\alpha = 1.1, \beta = 1.0, \kappa = 0$ and (a): $\rho = 1.6$; (b): $\rho = 2.0$, (c): $\rho = 2.5$.

For even values of n, all heteroclinic cycles are generally unstable. Stable het-

eroclinic cycles can only exist for odd values of n. In general they connect the equilibria with maximal support S of the form $\{i, i+2, i+4, \ldots i+n-3\}$, $E_{i,i+2,\ldots i+n-3}$. This means that starting from an expressed gene G_i every second gene is expressed, with the gene before G_i, G_{i-1}, being repressed, that is $\frac{n-1}{2}$ genes are expressed. Each boundary equilibrium is stable in the $2n-1$ dimensional manifold in which G_{i-1} is turned off and connected to the equilibrium $E_{i+2,i+3,\ldots,i+n-3,i+n-1}$ by a heteroclinic orbit. For $n=3$, this would be $E_1 \to E_3 \to E_2 \to E_1$, in the case of $n=5$ the only stable heteroclinic cycle consists of $E_{1,3} \to E_{3,5} \to E_{5,2} \to E_{2,4} \to E_{4,1} \to E_{1,3}$.

(a)

(b)

(c)
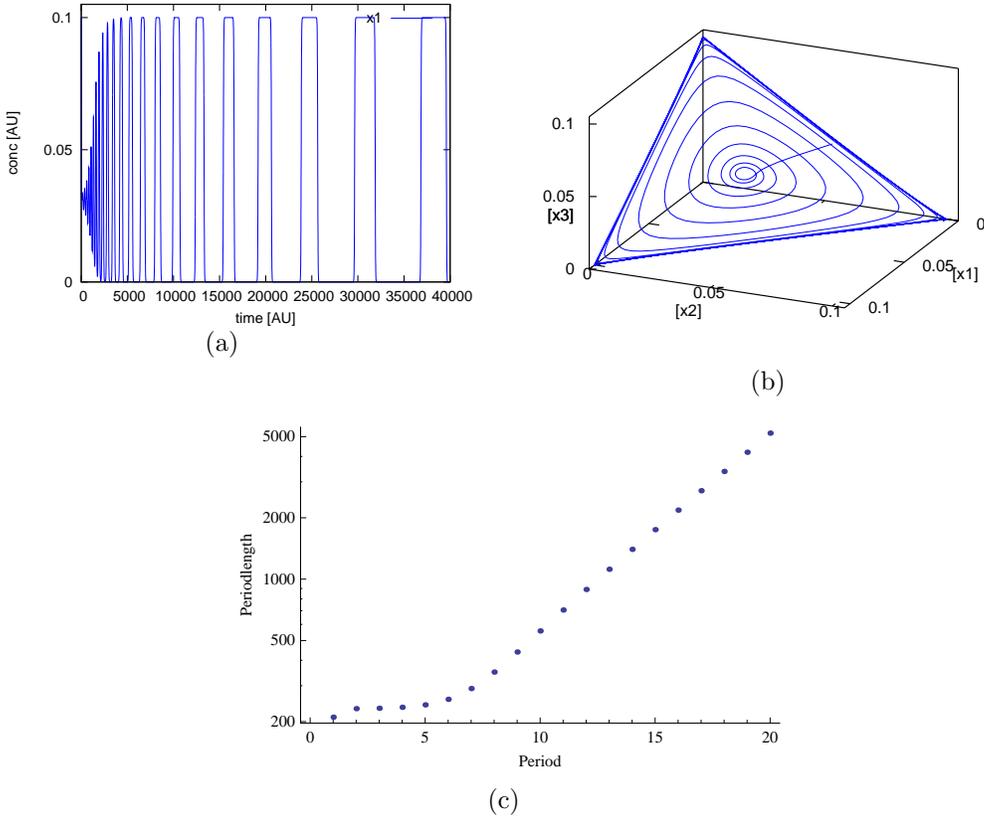
Figure 12: Timecourse (a), and trajectory in the protein space (b) of a system approaching a stable heteroclinic cycle. (c) shows the exponentially growing period length. Periods where calculated by taking every second passage through the plane $x1 = x2$ in protein concentration space. Parameter values: $\alpha = 1.1, \beta = 1.0, \kappa = 0$ and $\rho = 2.25$.

The existence of stable heteroclinic orbits for the system *RepAuto* has been proven by Stefan Müller and Josef Hofbauer [290]. Their proof uses the two principal eigenvalues λ and μ of each boundary equilibrium. λ is the largest

positive eigenvalue, and its corresponding eigenvector follows the leaving hetero-clinic orbit, while μ is the negative eigenvalue corresponding to the eigenvector pointing in the direction of the incoming heteroclinic orbit. From their proof a criterium for the stability of the heteroclinic cycle can be derived. The system is permanent and the heteroclinic cycle unstable if $\lambda + \frac{n-1}{2}\mu > 0$, while the heteroclinic cycle becomes asymptotically stable for $\lambda + \frac{n-1}{2}\mu < 0$. At $\lambda + \frac{n-1}{2}\mu = 0$ an invariant set is created in a heteroclinic bifurcation. In general this is a periodic orbit with very large periods. The principal eigenvalues at a boundary equilibrium take this form for $\gamma = 0$:

$$\lambda = -\frac{1+\beta}{2} + \sqrt{\left(\frac{1+\beta}{2}\right)^2 + \beta\left(\alpha - 1\right)} > 0 \tag{122}$$

$$\mu = -\frac{1+\beta}{2} + \sqrt{\left(\frac{1+\beta}{2}\right)^2 + \beta\left(\frac{\alpha}{1+\rho\left(\alpha - 1\right)} - 1\right)} < 0 \tag{123}$$

As can be seen, these eigenvalues do not depend on κ. Both exclusive and independent binding of transcription factors show the same pattern of stability of the heteroclinic cycle as different levels of cooperativity. Elimination of the other parameters from the stability criterion, though, is not possible. However, the eigenvalues λ and μ, have an inherent symmetry around $\beta = 1$. It can be shown that $\beta \cdot \lambda(1/\beta) = \lambda(\beta)$ and equivalent for μ. For the stability criterion this means that it is symmetric around the plane $\beta = 1$ on a logarithmic scale. Furthermore, for a given set of parameters other than β, the value of $\lambda + \frac{n-1}{2}\mu$ is minimal for $\beta = 1$ and maximal for $\beta = \infty = 0$. This allows the derivation of two sufficient criteria for definite stability and instability, independent of the value of β.

$$\left(\lambda + \frac{n-1}{2}\mu\right)_{\beta \to \infty} < 0 \quad \Rightarrow \quad \lambda + \frac{n-1}{2}\mu < 0 \tag{124}$$

$$\left(\lambda + \frac{n-1}{2}\mu\right)_{\beta = 1} > 0 \quad \Rightarrow \quad \lambda + \frac{n-1}{2}\mu > 0 \tag{125}$$

Using the central and limit cases $\beta = 1$ and $\beta = \infty = 0$, the following inequalities for definite stability and instability, respectively, dependant on the parameters
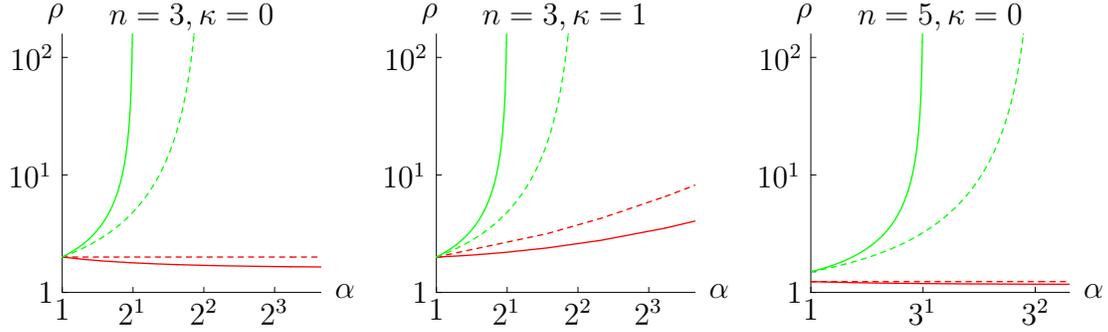
Figure 13: System *RepAuto* for n odd. Bifurcation diagrams in the (α, ρ)-plane for different numbers of genes ($n = 3$ or $n = 5$) and different types of regulator binding ($\kappa = 0$ or $\kappa = 1$). For $\alpha < 1$, the only attractor of system *RepAuto* is the origin. For $\alpha > 1$, there are three possible attractors: the central equilibrium, a limit cycle, and a heteroclinic cycle. The diagram shows the stability boundaries of the central equilibrium (red) and the heteroclinic cycle (green). The central equilibrium is stable below the solid red line (and unstable above the dashed red line), whereas the heteroclinic cycle is stable above the solid green line (and unstable below the dashed green line). Between the dashed lines there is a stable limit cycle. Between solid and dashed lines the stability (of the central equilibrium or the heteroclinic cycle) also depends on β and γ. (figure as in [290])

n, α and ρ can be derived:

$$(-1 + \alpha) + \frac{n-1}{2}\left(-1 + \frac{\alpha}{1 + \rho\,(\alpha - 1)}\right) < 0 \quad \text{(stability)} \tag{126}$$

$$(-1 + \sqrt{\alpha}) + \frac{n-1}{2}\left(-1 + \sqrt{\frac{\alpha}{1 + \rho\,(\alpha - 1)}}\right) > 0 \quad \text{(instability)} \tag{127}$$

Separated for ρ and α in dependency of the cycle size n this gives for the regions of definite stability and instability respectively:

$$\rho > \frac{\frac{n+1}{2}}{\frac{n+1}{2} - \alpha} \quad \wedge \quad \alpha < \frac{n+1}{2} \quad \text{(stability)} \tag{128}$$

$$\rho > \frac{\frac{n+1}{2}\left(\frac{n+1}{2} + \frac{n-3}{2}\sqrt{\alpha}\right)}{\left(\frac{n+1}{2} - \sqrt{\alpha}\right)^2 \left(1 + \sqrt{\alpha}\right)} \quad \vee \quad \alpha > \left(\frac{n+1}{2}\right)^2 \quad \text{(instability)} \tag{129}$$
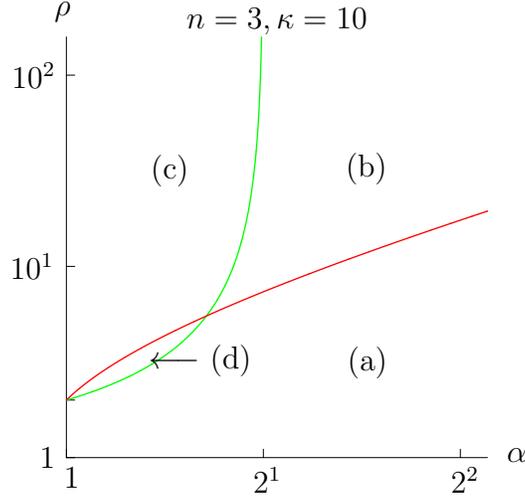
Figure 14: Bifurcation diagram in the (α, ρ)-plane for the smallest odd number of genes ($n = 3$), high cooperativity ($\kappa = 10$), and high degradation ratio ($\beta \to \infty$). For $\alpha < 1$, the only attractor of system *RepAuto* is the origin. For $\alpha > 1$, there are three possible attractors: the central equilibrium, a limit cycle, and a heteroclinic cycle. The diagram shows the stability boundaries of the central equilibrium (red) and the heteroclinic cycle (green). Below the red line the central equilibrium is stable, and above the green line the heteroclinic cycle is stable. As a consequence, in region (a) there is a stable central equilibrium and an unstable heteroclinic cycle, whereas in region (c) there is a stable heteroclinic cycle and an unstable central equilibrium. In region (b) there is a stable limit cycle. Finally, in region (d) both the central equilibrium and the heteroclinic cycle are stable. (figure as in [290])

2.5   Stochastic Simulation

To test whether the *RepAuto* system can also exhibit oscillatory behaviour under the more realistic assumption of fluctuating and discrete molecule numbers, a stochastic kinetic model was developed. For activator and repressor binding, fully detailed mechanisms were assumed following eqs. (44), (45), and (46). Binding reactions were split into association and dissociation steps, and all association rate constants assumed to be identical. For mutual exclusive binding of activator and repressor, $\kappa = 0$, the reactions leading to the ternary complex are omitted. Variation in binding affinities and cooperativity is achieved by varying the dissociation rates. In all other aspects the model follows the description in Elowitz *et al.* (2000) [108], both in formularisation and in the basic parameter values for transcription, translation, and mRNA and protein stability. All reactions describing mRNA and protein synthesis and decay are modelled as first order processes. The detailed model equations and parameter values are given

in appendix B.

(a)   (b)
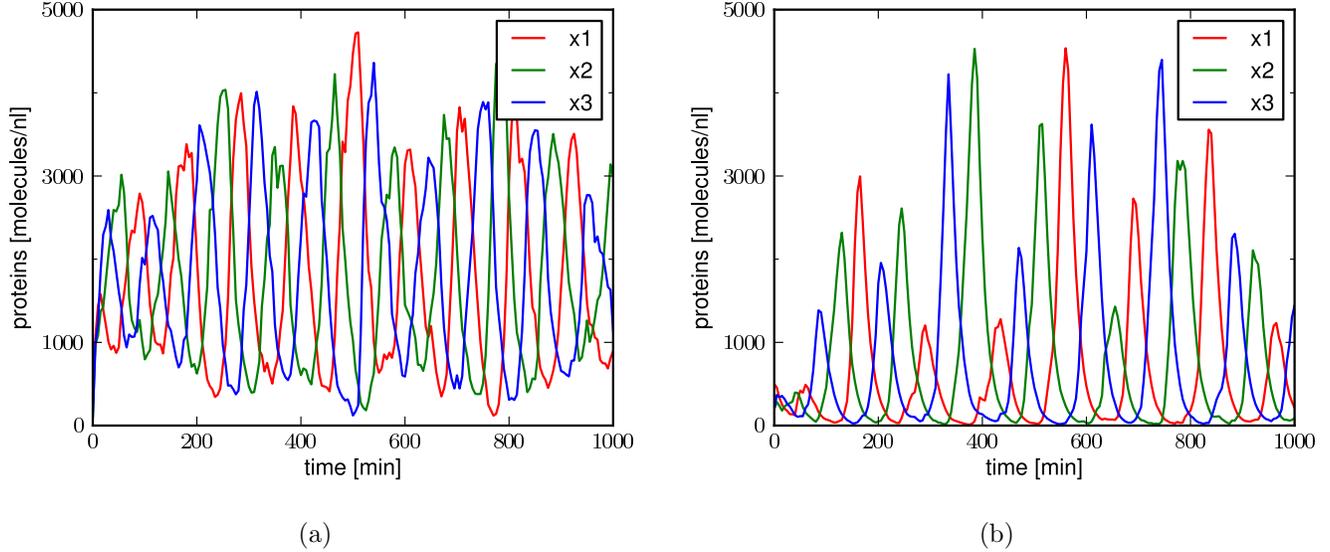
Figure 15: Single stochastic simulations of the classical repressilator, *RepLeaky*, (a), and the repressilator with autoactivation, *RepAuto* (b). The simulations were performed using the direct method implemented in the tool `Copasi` [196] with a duration of 2500 min and 500 output intervals. The parameters were chosen as in appendix B. The systems differ in the values for α (*RepLeaky*: $\alpha = 216$, *RepAuto*: $\alpha = 113$) so that they exhibit comparable mean protein numbers and frequencies. For *RepAuto* mutual exclusive binding of activator and repressor, $\kappa = 0$, was assumed, and the ratio of repressor to activator binding affinity, ρ, set to 2. In both systems β equals 0.2.

The three gene versions of the *RepAuto*, and, for comparison, the *RepLeaky* system, were implemented in SBML [200] and simulated using the direct method implemented in Copasi [147, 196]. Figure 15 shows representative time courses for both systems for similar parameter values.

To compare the different systems, the autocorrelation functions of one protein species were computed and averaged over a 1000 runs (see figure 16). From these autocorrelation functions approximate autocorrelation times, τ_A were calculated, by fitting their maxima to an exponential decay function:

$$D(t) = D_{t=0} \cdot e^{-t/\tau_a} \tag{130}$$

For the parameters used, the classical repressilator shows a slightly shorter autocorrelation time of 210 min or 1.6 periods compared to the repressilator with auto-activation with 352 min or 3.4 periods. The average amplitude spectra, de-

85

(a)

(b)
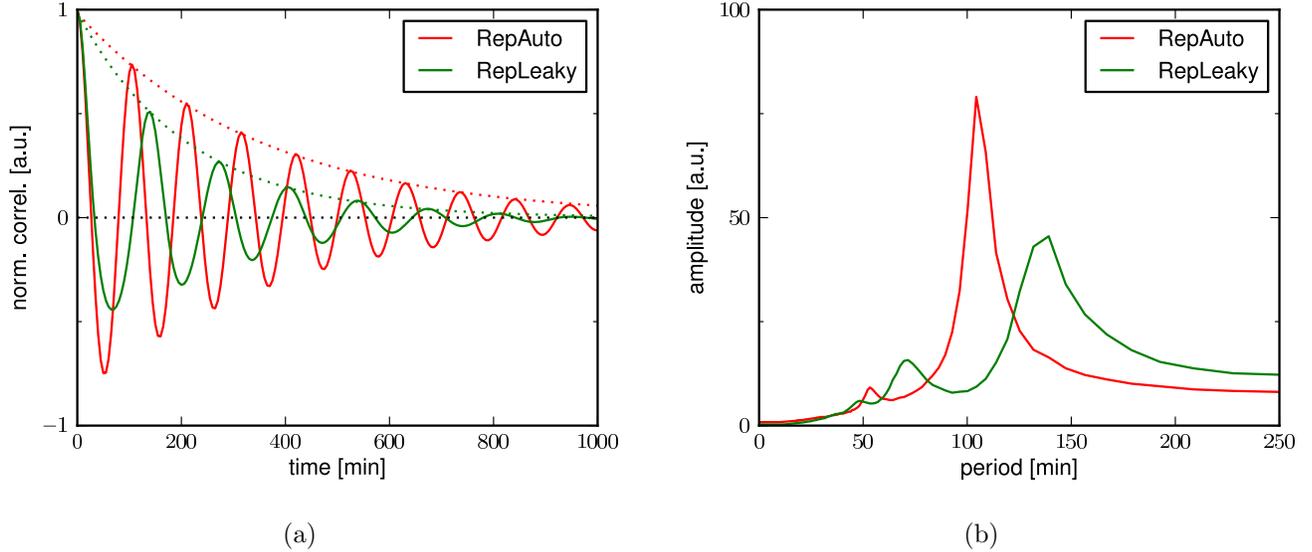
Figure 16: Normalised autocorrelation functions (a), and amplitude spectra (b) for one protein averaged over 1000 stochastic simulations (endpoint: 2500 min, 500 print intervals). The dotted lines in (a) show the fitted exponential decay functions used for calculating autocorrelation times. Both the autocorrelations and the amplitude spectra were calculated using the python library `numpy` [15, 208] for correlation and discrete Fourier transformation, respectively. Parameter values and simulations as in fig. 15.

rived by discrete Fourier transformation (see fig. 16(b)), also show that the peak around the main period is slightly wider in *RepLeaky* than in *RepAuto*. At least for these parameter values, the combination of positive and negative feedbacks seems to give oscillations with more robust and uniform periods in the stochastic framework.

As *RepAuto* was implemented with tight repression, and the requirement of auto-activation, the oscillations exhibited by the stochastic version have an increased tendency for dying out at parameter values that lead to low minimal protein or mRNA amounts per cell. As soon as one species goes extinct, the system tends to one of the corner equilibria, that is uniform expression of one species alone, and the oscillations abort. For mutually exclusive binding of activators and repressors, that is $\kappa = 0$, repressor affinity double that of activators, $\rho = 2.0$, and parameter values as in figure 15, abortion of oscillation occurs in $\approx 6\%$ of simulations over 2500 minutes. For higher values of ρ the abortion frequency increases rapidly reaching $\approx 80\%$ at $\rho = 2.5$. Above $\rho = 3.0$ nearly all simulations show only a few oscillations before two of the three species go extinct.

For values of $\kappa > 0$ and $\alpha > 100$, the deterministic model predicts that higher values of ρ are needed for to achieve stable oscillations. This leads to high rates of species extinction in the stochastic model, and made it impossible to find sustained oscillations for mechanisms other than mutually exclusive activator and repressor binding.

By including a low transcription rate from free promoters this effect can be mitigated. For this, similar to the leakiness in *RepLeaky*, basal transcription from the unoccupied promoter was assumed at a rate of $k_{ts}^{bas} = \delta \cdot k_{ts}^{act}$. While this inclusion hardly influences the deterministic behaviour at the inspected parameter ranges, and for small values of the leakiness δ, it reduces the abortions of oscillations in the stochastic interpretation.

2.6   Discussion

The results presented in this chapter affirm and expand the behaviours previously suggested in [129] and proven by [384], for cyclic gene regulatory networks with negative feedback for two generalised repressilator systems. For both models of regulatory control, the repressilator with leaky repression and the one with autoactivation, multistability for even and oscillatory behaviour for odd numbers of genes could be found.

Detailed analytical investigation of gene regulatory systems not only provides us with an idea of which behaviours can be displayed by a given network layout, but it also allows us to identify the key parameters that are required for a specific behaviour to arise, and the ranges over which it is displayed. These parameter ranges are of particular importance as some parameter values can fluctuate considerably due to individual differences and noise in cellular environments, or they can be sensitive to mutational changes in both the coding and the regulatory gene sequences. Analysing them can help our understanding of the evolution and robustness of existing systems, and provide clues as to how we can synthetically create new ones.

As only approximate values are known for most parameters in biological systems, the derived bifurcation diagrams are especially helpful in finding the combinations which are most likely to lead to the desired behaviour. While the current analysis of the established *RepLeaky* system did not reveal any unexpected
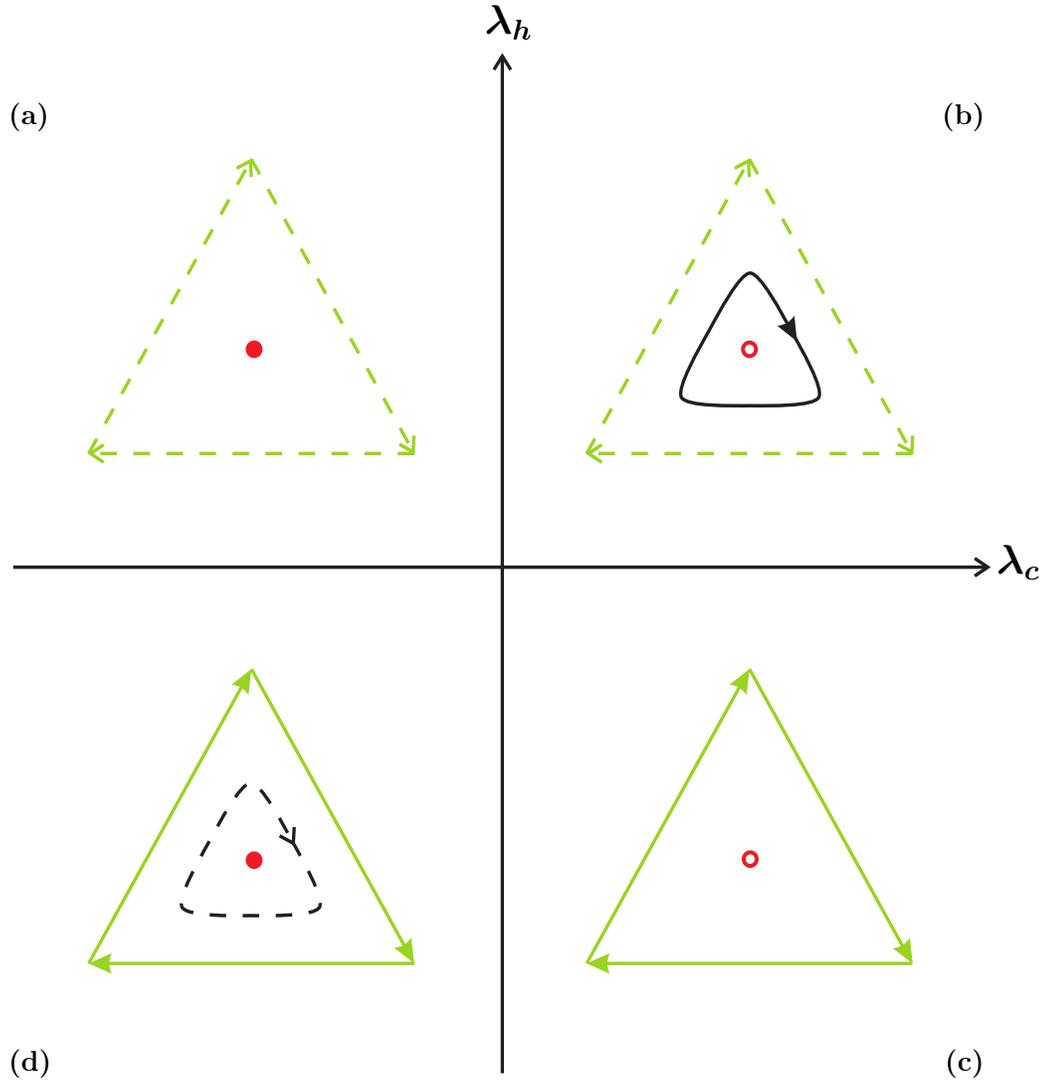
Figure 17: System *RepAuto* for n odd. A sketch of the four dynamical scenarios for the smallest number of genes ($n = 3$) and high cooperativity ($\kappa \gg 1$). There are three possible attractors (denoted by colors): the central equilibrium (red), the limit cycle (black), and the heteroclinic cycle (green). Stable orbits are shown as filled circles or solid lines, unstable orbits as empty circles or dashed lines. The "eigenvalue" of the central equilibrium, $\lambda_c = \Re(\lambda_{\text{odd}})$, and the "eigenvalue" of the heteroclinic orbit, $\lambda_h = \lambda + \mu$, are indicated on the coordinate axes. In case (a) there is a stable central equilibrium and an unstable heteroclinic cycle, whereas in case (c) there is a stable heteroclinic cycle and an unstable central equilibrium. In case (b) both eigenvalues are positive and there is a stable limit cycle. Finally, in case (d) both eigenvalues are negative and there is a stable central equilibrium and a stable heteroclinic cycle together with an unstable limit cycle. (figure as in [290])

behaviour, the detailed study of the dependence of previously characterised behaviours on the various parameters may help to find the ideal combinations of transcription factors and regulatory sequences when trying to engineer robust

oscillatory systems with specific properties.

Unexpectedly the inclusion of mass conservation for regulator binding did not change the position of the equilibria of the considered systems, and only affected the emergence and stability of periodic and aperiodic attractors. While over large ranges of parameter space, the stability of these attractors has been shown to be independent of the correction for transcription factor binding, it shows some influence on the frequency of the resulting oscillations.

In general, mass conversation for transcription factor binding only becomes important for high copy numbers of genes, strong binding, and weak promoters. While chromosomal genes in bacteria commonly only exist in 1 to 2 copies, plasmids can have copy numbers in the tens to even thousands, which could make such a correction necessary [304, 360].

The analysis of the repressilator with auto-activation, *RepAuto*, revealed some interesting novel behaviours. This system shows a broader range of behaviours than the classical repressilator, due to the existence of the heteroclinic cycle (figure 17). It can exhibit flow to a central equilibrium, stable periodic oscillations, or oscillations with increasing period lengths (figures 11 and 12). In this *RepAuto* is similar to the system of three competing populations described by May and Leonard [274, 326].

In real gene regulatory networks this kind of behaviour does not have much direct significance, as protein and mRNA amounts per cell are discrete numbers. The system would stop after a few oscillations in the nearest corner equilibrium, once the copy number of one type of transcription factor becomes too low. Also, the stable heteroclinic orbits require relatively low values of α in combination with high values of ρ (fig. 13), which would be unrealistic in real systems, and did not show any oscillations in the stochastic system. However, the existence of the heteroclinic orbit by itself has some implications. It gives an upper limit for repressor binding, above which oscillations die out. For the three gene case, this means that the system tends to a state in which one gene is actively transcribed, while the other two are silenced.

The stochastic simulations showed that, for comparable parameter values, *RepAuto* has a slighter higher period stability than the classical repressilator. This could be due to the, on average, higher minimal protein numbers during oscillations.

In the classical repressilator during oscillations one gene is repressed nearly to the limit of leaky transcription, which could lead to more fluctuation in the times between peaks.

Another important difference between the two systems analysed lies in the dependency of oscillations on cooperativity in transcription factor binding. While the classical repressilator requires a minimal degree of cooperativity in repressor binding, the combination of independently binding activators and repressors suffices to give stable oscillations.

In the deterministic version of *RepAuto* sustained oscillations can be achieved both with independent and cooperative activator and repressor binding, with the region of oscillatory being greatest for mutually exclusive binding, $\kappa = 0$ (see fig. 13 and eq. (113)). For realistic parameter values sustained oscillations were only achievable with strong negative cooperativity with values of $\kappa \leq 0.1$. Further, while the analysis of the deterministic system suggests that high values of the ratio of repressor to activator affinity, ρ, together with high values of α would robustly lie in the oscillating region, in the stochastic simulations the strong repression leads to the abortion of oscillations and the expression of only one of the genes.

In this case the deterministic treatment alone gives insufficient information for choosing the right combination of parameters. By complementing it with stochastic simulations, a better picture of the requirements for a *in vivo* system exhibiting sustained oscillations can be provided.

For such a system, an architecture with mutually exclusive binding of activators and repressors should be chosen. This could for example be achieved by having overlapping binding sites close to the promoter region of each gene. If each gene encodes for an activating transcription factor, its product could function as an activator or repressor, depending on the position of the binding site [337].

The main practical problems to tackle would be the mutual exclusivity of binding of repressors and activators, $\kappa \ll 1$, and the ratio of affinities of the each transcription factor to the promoter of the repressed and the activated gene, ρ. A higher value of ρ leads to higher amplitudes, and is also required for oscillations in combination with the high values of α necessary to achieve protein concentrations high enough to minimise fluctuations. ρ could be varied to some extent by

changing the DNA sequences the transcription factors bind to, although more than a factor of 10 could be hard to achieve without losing binding specificity completely. Similar to the classical repressilator it is beneficial if the protein and the mRNA half-life are similar, that is $\beta \approx 1$. Both the half life of the mRNA and the protein can be altered, for example using destabilising sequence tags [70, 108]. Altering the half-lives of proteins should also be one of the easier means to change the frequency of oscillations.

3 Gene Regulatory Networks and Gene Duplication

3.1 Introduction

Studying the effects of mutation on gene regulatory networks is important to improve our understanding of both their evolution and the robustness of their behaviour. While many mutations only lead to subtle changes in a single interaction, some can completely alter the topology of a network by adding or removing new transcriptional regulators and by rearranging their targets. Gene duplication, followed by mutation of the paralogous genes, is one such radical mutational process, and its various effects on gene regulatory networks with feedback loops are intriguing. In this chapter the effects of gene duplication on auto-regulatory GATA-type transcription factors are studied using analytical and numerical methods. Potential network topologies, deduced from yeast data, are taken as starting points, and are subsequently combined with physiological parameter ranges from the literature. The chosen parameters are refined and validated by creating a core model of nitrogen catabolite repression in *S. cerevisiae* and comparing it to experimental data.

3.1.1 Gene Duplication

One of the driving forces of evolution is the emergence of new genes. While point mutations and small-scale sequence alterations can have an important impact on the genetic variability of a species, sequencing projects have shown that a high percentage of expressed genes in higher eucaryotes stems from duplication events, and the reuse of existing sequences [81, 102, 169, 401]. During the recent evolution of primates, gene gain and loss have been found to be particularly accelerated relative to the mammalian average, while nucleotide substitutions appear to be less common [169]. These copy number variations could be one of the factors underlying the large morphological differences between some primates, which are present despite their highly conserved nucleotide sequences. For example, brain related gene families in humans have been found have doubled in size in comparison to other mammals, which could be linked to the increased relative brain mass. Further, the intra-species variability seems, to a high degree,

92

to stem from variations in gene copy numbers [363].

Gene duplication has been postulated to be a paramount factor in the development of complex organisms and new genes by Ohno and Kimura [223, 310] and over the years, this view has been corroborated [39, 401]. Several different processes can lead to gene duplication, ranging from duplication of the whole genome, over single chromosomes, to smaller DNA fragments, such as tandem gene duplications. The mechanisms underlying such duplications are diverse, ranging from polyploidization, over incomplete separation of chromosomes and unequal crossovers during meiosis, to the action of transposable elements and retroviruses.

Whole genome duplications (WGD) are assumed to have occurred multiple times during the evolution of most eukaryotes. In the evolution of vertebrates alone, at least three potential WGDs have been identified [81, 215]. Budding yeast, *S. cerevisae*, seems to have undergone at least one round of genome, and numerous tandem gene duplications in its divergence from the other hemiascomycetes [102] (see figure 19).

As found with *S. cerevisae*, massive gene duplication events can be followed by wide ranging loss of gene copies. The processes of gene loss and retention following duplication have been studied extensively and found to vary between different types of gene functions and families. One potential explanation for the different rates of gene copy loss and retention is that some classes of genes are very much dependent on the number of copies in the genome, that is they posses a higher gene dosage sensitivity. Duplications can therefore lead to a selective pressure to either the loss of a copy, or divergence of the paralogous genes by sequence mutation. Genes involved in signal transduction, transcriptional regulation, or those which encode parts of macromolecular complexes depending on a strict stoichiometry have been speculated to be particularly dosage sensitive [47, 318].

It has been suggested, that gene duplications can lead to a certain degree of redundancy, meaning that less selective pressure is exerted on the paralogous gene copies. Such conditions would allow the genes to accumulate mutations and diverge in function over time. The classical view for many years was that of neofunctionalization, meaning that one paralog diverges to acquire novel functions, while the other retains the original role [311]. Another possibility for the

retention of both gene copies is the division of the original functions between the paralogous genes, an evolutionary process termed subfunctionalization [128].

The effect of gene duplication on transcriptional networks is especially intriguing, as a duplication of a transcription factor followed by a few alterations in regulation and/or binding specificity could lead to completely new network topologies and complex behaviours. An example of such a duplication and subsequent divergence of regulatory sequences has been described in the genetic switch controlling the *GAL* pathway in *S. cerevisiae* [187].

3.1.2   The GATA Family of Transcription Factors

GATA-type transcription factors constitute an ubiquitous family of DNA-binding proteins found in almost all eucaryotes from fungi to vertebrates [256, 320]. They are characterised by containing one or two Zn-finger domains with an adjacent, highly conserved, basic region. Most GATA factors bind to a consensus sequence containing the name-giving `(A/T)GATA(A/G)` motif. Outside the DNA binding region little sequence conservation is observed and GATA factors can function as either transcriptional activators or repressors, and sometimes even as both, depending on their context and co-factors [256]. Distinct from other widely spread families of transcription factors, such as winged helix or homeo-box domain containing factors, of which dozens to hundreds of different factors have been identified in some genomes, most species only posses a few closely related forms of GATA factors.

GATA-type transcription factors are involved in the regulation of diverse processes ranging from basic metabolism to developmental processes, cell differentiation and immune response. They have been found to be part of feedback loops and feed-forward motifs of varying size. Small auto-regulatory loops have been identified [350, 412], as well as larger loops in combination with other GATA type [160, 161] or different transcription factors [59, 61, 199]. In *C. elegans* a complex, three tiered cascade of GATA-type transcription factors has been implicated in endomesoderm differentiation [261]. At least one of these GATA factors, ELT-2, furthermore seems to be auto-activating.

Transcriptional feedback loops are fascinating, as they can give rise to phenomena such as bistability and oscillation, which have been proposed to underlie

biological phenomena such as cell differentiation, rhythmic processes, and tissue morphogenesis [118, 159, 403, 409]. The involvement of GATA-type transcription factors in one such phenomenon, the differentiation of T-helper (T_h) cells into T_h2 cells, has been studied in detail with both experimental methods and mathematical models [192, 292]. In this process, the stimulation of T_h cells by the cytokine IL4 leads to sustained activation of GATA-3 via the STAT-6 transcription factor. Assuming a positive, autoregulatory feedback loop for GATA-3, bistability in the expression of GATA-3 was found, giving a possible mechanism for T_h2 cell differentiation and memory cell formation.

3.1.3   GATA Factors in *S. cerevisae*

In the yeast *S. cerevisae* GATA factors play a central role in the regulation of amino acid uptake and metabolism. Nitrogen catabolite repression (NCR)-sensitive gene expression is, to a large extent, regulated by four members of the GATA family: two transcriptional activators, Gln3p and Gat1p/Nil1p, and two repressors, Dal80p and Deh1p/Gzf3p. These four GATA factors are interconnected by transcriptional regulations, forming an intricate network with positive and negative feedback loops [83, 84] (see figure 18). The term NCR describes the observation, that during the growth of yeast on good nitrogen sources, such as ammonium or glutamate, certain genes needed for the assimilation of nitrogen from poor sources, such as proline or arginine, are not transcribed, or only to a lesser degree - hence the slightly missleading term 'repression'. NCR-sensitive genes comprise enzymes and permeases mainly required for growth and survival on substrates with poor nitrogen availability.

As shown in figure 18, expression of NCR-sensitive genes is tightly controlled. The whole system is regulated by at least two types of ubiquitous eukaryotic transducers of nutrition status signals. In budding yeast these are the TOR kinases, Tor1p and Tor2p, and the AMP-activated kinase Snf1p [84]. Of the GATA factors involved in NCR, only GLN3 transcription seems not to be regulated by other GATA factors [83], and instead is expressed constitutively. In presence of a good nitrogen source, both Gln3p and Gat1p are exported from the nucleus into the cytoplasm, where they are subsequently sequestered by association with the prion-like Ure2p protein. While the exact mechanisms are unknown, upon deactivation of TOR, Gln3p and Gat1p dissociate from Ure2p and translocate
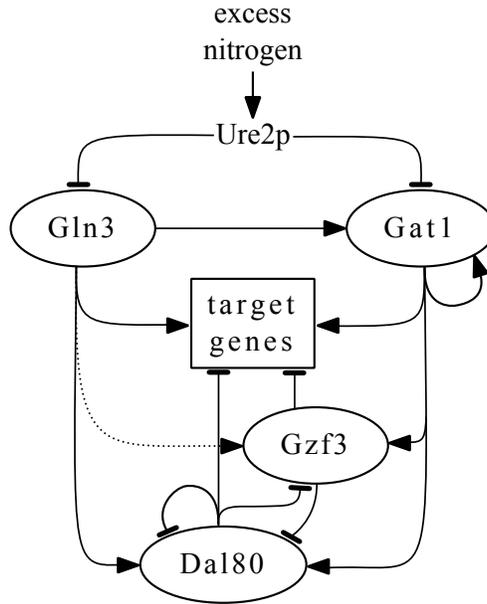
Figure 18: Schematic representation of the regulatory interactions of GATA type transcription factors involved in NCR in the yeast *S. cerevisiae*. GATA factors are shown as elipses, bars and arrowheads stand for positive and negative regulation, respectively. Dotted lines indicate weaker or speculative interactions. After Cooper (2002)[84].

into the nucleus. Under conditions promoting NCR, Gln3p, and possibly also Ure2p, has been found to be phosphorylated, leading to the assumption that Tor1/2p could either directly phosphorylate Gln3p, or, alternatively, inhibit a phosphatase acting on it [84].

Once Gln3p has been translocated to the nucleus, it activates GAT1 transcription. Most other GATA regulated NCR sensitive genes require both Gln3p and Gat1p for their expression, which is indicative of their finely tuned regulation. With increased Gat1p expression, DAL80 and GZF3 are expressed, leading to subsequent down-regulation of Gat1p and NCR-sensitive genes. Dal80p, and most probably also Gzf3p, has been found to compete with Gln3p and Gat1p for available (A/T)GATA(A/G) sequences and been implicated in the fine-tuning of final gene expression. Directly after the release from NCR, the GATA regulated genes are only transcribed at a high level for a short period, until Dal80p concentrations increase and down-regulate expression to the final steady-state levels [92]. Such interplay between a positive and negative feedback is a common feature in homoeostasis and can help to overcome gene dosage effects [403]. In combination with a time delay - such as the one present in this system and as

seen in the repressilator (see chapter 2) - it can even lead to oscillations.

In several fungal species, excluding *S. cerevisiae*, another family of GATA-type transcription factors is involved in the transcriptional regulation of genes encoding parts of iron import systems [167]. This family of transcription factors contains a small iron binding domain, conferring sensitivity to iron dependency, and functions as an inhibitor of transcription. Whilst none of these repressors has been shown to regulate their own expression, their DNA binding domain shows significant similarities with those of GATA factors involved in NCR regulation [256]. This could point to a common origin and also opens up the possibility of a potential cross-talk between the regulation of iron uptake and nitrogen metabolism.

3.1.4 Duplication and Mutation in GATA Type Gene Regulatory Networks

The development of different GATA-type transcription factors has been tightly linked to various duplication events in many organisms. A study on the phylogeny of GATA genes suggested that the numerous GATA factors of the GATA123 and the GATA456 families in vertebrates stem from two ancestoral GATA genes in the invertebrate deuterostomal progenitor. Furthermore, it was noted that they were mainly derived in two - or three in the case of teleost fish - rounds of whole genome duplications [151, 174]. There is also evidence of tandem gene duplications, as was speculated in the development of the GATA genes containing two Zn-fingers [256]. Another example has been found in the nematode *C. elegans*, in which the elt-4 gene most likely stems from a tandem duplication of elt-2, a gene encoding a GATA factor involved in gut development. While no function has been ascribed to the elt-4 gene, it has survived longer than the average duplicated gene in *C. elegans*, leading to the assumption that it does - or once did - confer some selective advantage [132].

The involvement of GATA type transcription factors in intricately cross-regulated networks, such as the one governing NCR-sensitive genes in yeast, leads to the question of how such complexity could evolve. One possibility would be rounds of gene duplication, followed by mutation of the paralogous genes and their regulatory sequences. This type of transcription factor could be predestined to this kind of network generation for various reasons. For one, they can function both

as activators as well as repressors of transcription. Further, they bind similar DNA motifs, and, in at least some cases, both inhibitory and activating GATA factors have been found to compete for the same sequences. Starting from a possible *ur*-GATA factor, a gene regulatory network, such as the one shown in figure 18, could have evolved by a gene duplication, loss of the transactivator domain in one paralog, another duplication event, and loss or mutation, respectively, of the gene regulatory sequences.

This reasoning follows the line of the classical theory of neofunctionalization of paralogous genes [401]. The case of GLN3 and GAT1 could also be envisioned to stem from subfunctionalization after a duplication event. In this case the progenitor could have been more like GAT1, auto-activated and with a positive feedback on its own expression. After a duplication one paralog could have lost the regulatory regions and become consitutively expressed at a low level, while the other would have stayed strongly regulated, together leading to a more efficient switching behaviour in response to a stimulus.

3.2 Derivation of Potential Network Topologies

`PSI-BLAST` [9] was used to retrieve 106 open reading frames (ORFs) from 15 completely sequenced fungal genomes downloaded from the NCBI webpage using the sequences of the core Zn-finger domains of the *S. cerevisiae* GATA factor Gat1p as a basis for the search. To narrow down the results to only closely related factors, first the core Zn-finger domains comprising around 60 amino acids were identified using the program `DIALIGN` [287]. These partial sequences were subsequently aligned using `Clustal W` [410].

The multiple alignment demonstrated that only 49 of the 106 potential proteins were closely related, while the remaining proteins displayed little sequence similarity. These 49 related proteins were then further subdivided into different classes. First, they were grouped according to the similarity of their core Zn-finger regions to activators and inhibitors involved in *S. cerevisiae* NCR, Gat1p, Gln3p, and Gzf3p, and to the iron-sensitive inhibitors SreA of *A. nidulans* [168] and URBS1 of *U. maydis* [10]. The iron sensitive GATA factors were further subdivided according to the number of Zn-fingers. In ambiguous cases the full length alignments to the seed sequences were used to find the most closely related
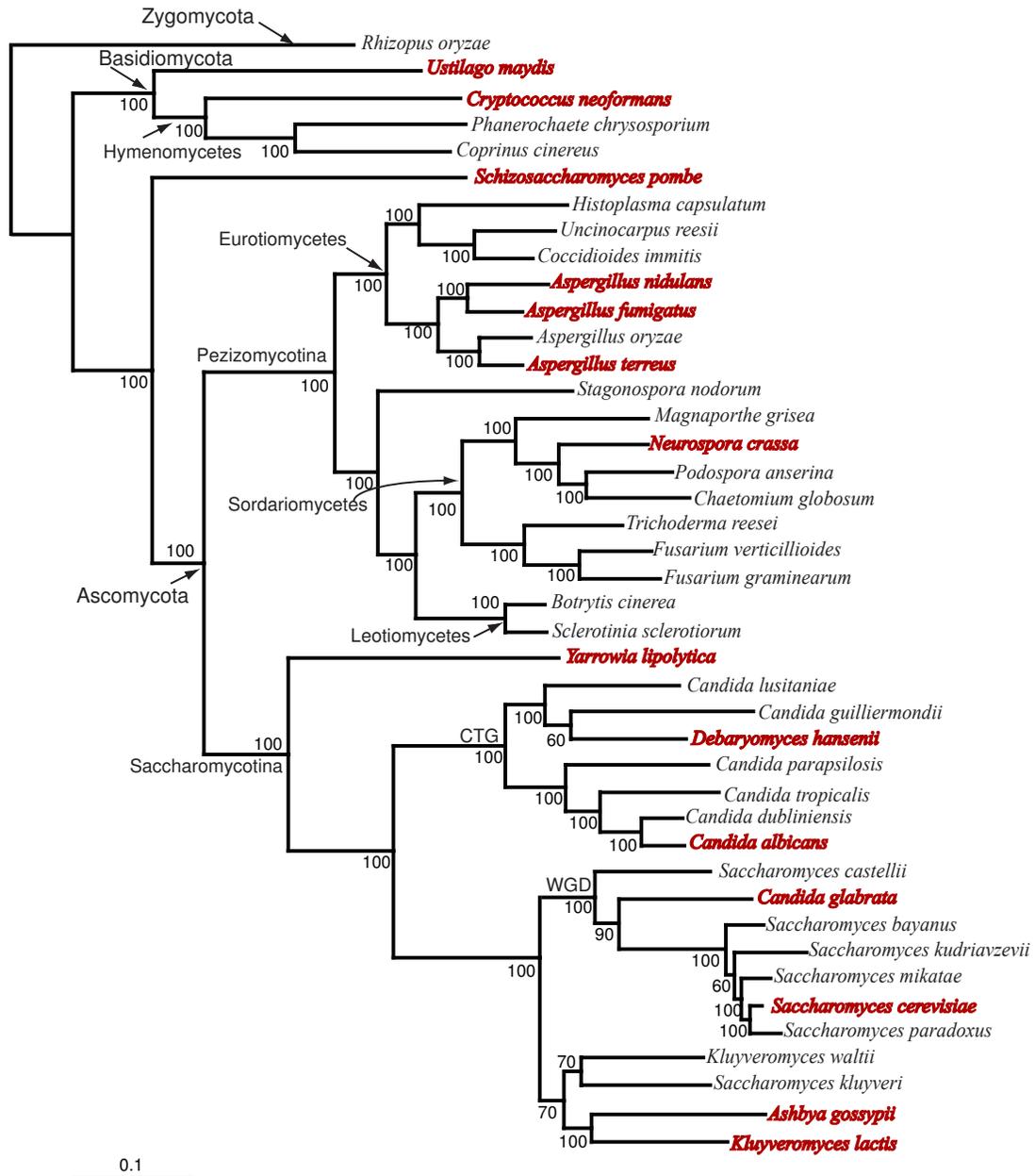
Figure 19: Maximum likelihood phylogeny of 42 fungal species based on 153 genes found in all taxa (taken from Fitzpatrick *et al.* BMC Evolutionary Biology 2006 6:99 [122], doi:10.1186/1471-2148-6-99, Creative Commons Attribution License). Taxa in the branch under WGD have undergone a whole genome duplication, the clade under CTG translate this codon as serine instead of leucine. The bar inciates the distance in expected substiutions per site. The figure was slightly altered to highlight the species used in this study in red.

factor (see table 1).

While the binding specificities of these potential GATA factors have not been analysed in detail, as mentioned above most GATA factors show similar consen-

	Total	ZN2FB	ZNFB	ZN2	GAT1	GAT1$_{inh}$	GZF3	GLN3
Encephalitozoon cuniculi	1					1		
Cryptococcus neoformans	2		1/**1**		1			
Ustilago maydis	3	1		1	1			
Schizosaccharomyces pombe	3	1		1/**1**	1/**1**			
Aspergillus terreus	2	1/**1**			1/**1**			
Aspergillus nidulans	3	1			1/**1**		1/**1**	
Aspergillus fumigatus	3	1			1/**1**		1/**1**	
Neurospora crassa	3	1			1		1/**1**	
Yarrowia lipolytica	4	1			1	1	1/**1**	
Debaryomyces hansenii	5	1			2*/**1**		1/**1**	1
Candida albicans	6	1			2*/**1**		2/**2**	1
Ashbya gossypii	3				1/**1**		1/**1**	1
Kluyveromyces lactis	3				1		1/**1**	1
Candida glabrata	4				1/**1**		2/**2**	1
Saccharomyces cerevisiae	4				1/**1**		2/**2**	1
Total	**49**	**9**	**1**	**2**	**16**	**2**	**13**	**6**

Table 1: Numbers per organism and classifications of the 49 potential GATA factors retrieved from 16 fungal genomes. Boldface numbers indicate the number of genes with multiple adjacent `HGATAR` motifs in their 500 bp upstream regions.
GATA factors similar to iron-sensitive inhibitors: ZN2FB: GATA factors with 2 Zn-finger domains and iron-binding domain, ZNFB: iron-binding domain, but only C-terminal Zn-finger, ZN2: 2 Zn-fingers without iron-binding domain
GATA factors potentially involved in NCR: GAT1: Gat1/AreA-like activating GATA factors, GAT1$_{inh}$: Gat1/AreA-like GATA factors without the N-terminal trans-activation domain, Gzf3: Gzf3/Dal80/AreB-like inhibiting GATA factors, Gln3: Gln3-like GATA factors.
∗) the Zn-finger domain of one of the proteins differs significantly, but, according to the full length alignment, the protein is most closely related to Gat1.

sus binding motifs [256]. The retrieved core sequences have been mapped onto an NMR structure of an AreAp:DNA complex [392], and most residues involved in specific interactions with nucleotides were found to be fully conserved in all 49 proteins (Rainer Machné, personal communication). This further supports the hypothesis that this set of proteins have similar, or maybe even identical, DNA binding specificities.

To find genes, whose transcription is potentially regulated by GATA-type transcription factors, in this set of 49 proteins, the 500 bp upstream of their start codons were screened for potential GATA binding sites using a `HGATAR` consensus sequence [341] where H stands for A, T or C and R for A or G. There has been

evidence that repeats of `HGATAR` motifs are required for efficient transcriptional regulation by GATA factors alone [158, 339], so as an additional constraint a distance of at most 37 or 51 bp between adjacent motifs was used. Table 2 shows the individual results for each of the 49 proteins, and table 3 gives an overview of the basic architecture and potential regulation of the proteins found.

Gene	GATA sites total num.	# repeats dist. $\leq$ 36/50bp	Gene	GATA sites total num.	# repeats dist. $\leq$ 36/50bp
Zn2FB			**Gat1**		
Nc_SREP			Kl_50312349	1	
Yl_50552360			Ag_45199031	2	1/1
Sp_19113848	2		Sc_GAT1	6	4/5
Af_SreA	2		Dh_50421709	2	
At_SREP	3	1/2	Ca_46438200	3	1/1
An_40746893	2		Cn_58258269		
Ca_SFU1	2		Um_46100068	1	
Dh_50420129	2		At_AreA	4	3/3
Um_Urbs1	1		Af_AreA	3	1/2
ZnFB			An_AreA	4	3/3
Cn_SREP	4	1/1	Nc_28925530		
Zn2			Sp_63054447		
Sp_19075466	5	1/2	Yl_50556296		
Um_46099653			Cg_50292241	6	1/3
Gzf3			Ca_46443763	2	
Ca_46437465	8	1/4	Dh_50427591	4	1/1
Ca_46437412	7	1/4	**Gat1$_{inh}$**		
Dh_50418791	9	6/7	Yl_50551201		
Sc_DAL80	8	4/5	Ec_GATA	1	
Af_AreB	8	4/5	**Gln3**		
An_40738445	7	3/3	Dh_50424457	1	
Yl_50549355	5	3/3	Ca_46434483	2	
Nc_28923776	9	6/6	Kl_50302249	2	
Kl_50312009	6	2/2	Ag_45198755	2	
Cg_50288243	8	5/5	Sc_GLN3	2	
Ag_45198587	2	1/1	Cg_50285693		
Sc_GZF3	6	3/3			
Cg_50292953	8	5/6			

Table 2: Total number of GATA binding sites (`HGATAR`) and number of adjacent sites with a distance of maximal 36 (left) and 50 bp (right), respectively, for the 500 nucleotide upstream sequences of each of the 49 putative GATA factors. Protein classes (bold) as defined in table 1.

3.2.1   Evolution of GATA Factors in Fungi

While the species most distantly related to the other examined fungi, the microsporidium *E. cuniculi*, only possesses a single truncated GAT1 related gene, GAT1-like and iron-sensitive GATA factors are found in both *Basidiomycotes*

and *Ascomycetes*, pointing to an ancient common origin of these genes. Of the *Ascomycetes* only *S. pombe* and *A. terreus* do not possess a GZF3-like inhibitor, which indicates that this innovation might have arisen after the branching off of the *Taphrinomycotina*, and lost in *A. terreus*.

Class	Structure	Genes	GATA sites
Zn2FB	`----//--ZnN-FeB-ZnC-------C-`	9	1
ZnFB	`------FeB-ZnC------C-`	1	1
Zn2	`--(TAD)----ZnN-ZnC-------`	2	1
Gat1	`---TAD--//-----ZnCe--//-H`	16	8
Gat1$_{inh}$	`-ZnCe-----H`	2	0
Gzf3	`--ZnC---//---C`	13	13
Gln3	`---TAD--//-----ZnCe--//-H`	6	0

Table 3: Protein classes as defined in table 1. Domains: **ZnN/C**: N- and C-terminal Zn-fingers, **e:** nuclear export signal, **FeB:** iron-binding domain, **TAD:** (putative) TransActivation Domain, **C:** coiled-coil domain, **H:** small C-terminal domain/helix, **-:** strongly disordered regions, **//:** length variations, **GENES:** number of genes assigned to this type, **GATA sites:** number of genes with multiple `HGATAR` motifs in their 500 bp upstream region.

Nearly all of the closely related *Saccharomycetales* possess a GLN3-like gene, apart from *Y. lipolytica*, which separated quite early from the other *Saccharomycetales*. By contrast, *Y. lipolytica* possesses a gene similar to GAT1, only without the trans-activation domain, potentially functioning as a competitive inhibitor to its GAT1-like gene. While *S. cerevisiae* and *C. glabrata* supposedly underwent a WGD before they diverged [102], only the GZF3-like inhibitors are found to exist in two closely related forms. The duplicates of GLN3 and GAT1 seem to have been lost, which might be indicative of a potential disadvantageous gene dosage effect of these activating GATA factors. Interestingly all 13 GZF3-like genes, putative transcriptional inhibitors, possess at least one repeated `HGATAR` motif, which points to a strong conservation of regulation of repressor expression by GATA type transcription factors in all the yeasts inspected. Strict transcriptional regulation could be a reason for the retention of these genes after duplication, as it could mitigate dosage effects.

Similarly a considerable number of the GAT1/AreA type genes show multiple `HGATAR` motifs in their upstream sequences. Among the *Ascomycetes K. lactis* and *Y. lipolytica* seem to have lost these motifs, and hence maybe regulation by other GATA factors. The GLN3-like genes on the other hand, only posses one or

two separated GATA binding motifs, which fits with the assumption that their transcription is neither auto-regulated nor influenced by other GATA factors [84]. In summary, the core architecture of the GATA factors network involved in NCR appears to be reasonably conserved over all the *Saccharomycetes*.

The three *Aspergillus* species and *N. crassa* each possess one GATA factor similar to AreA and GAT1, and with the exception of *A. terreus*, one AreB-like inhibitory factor similar to GZF3. The lack of a GLN3-like protein is in agreement with the findings on the differences in regulation of the NCR-sensitive genes to *S. cerevisiae*. It seems that the activity of AreA in *A. nidulans* and *N. crassa* is mainly controlled at the post-transcriptional level via mRNA stability [288, 328], although it has been suggested that AreB is capable of inhibiting AreA expression under certain conditions [440].

3.2.2 Potential Network Structures

The different GATA factors related to types involved in the regulation of nitrogen catabolic gene transcription inspired the derivation of the topologies for the underlying GATA factor gene regulatory networks. Additional information for the regulation of the activity of the GAT1/AreA-like transcription factors was taken from *S. cerevisiae*, *N. crassa* or *A. nidulans*, depending on which species the regulation of a particular gene is best characterised in.

The evolution of the assumed network starts with a single activating GAT1-like transcription factor, as in the *Basidiomycotes*. Then the network is altered by gain of auto-activation, emergence of inhibitors, and, in the case of *Hemiascomycetes*, of an constitutively expressed GLN3-like activating factor (see figure 20). As mentioned above, the latter two cases can easily be envisioned as gene duplication events followed either by the loss of the trans-activation domain, or subsequent mutation of the upstream regulatory sequences, respectively.

3.3 Base Model of NCR in *S. cerevisiae*

As parameter values and mathematical relations are of paramount importance for the behaviour of a dynamical system, at first a basal model of the core regulation in yeast NCR was created for validation and to find suitable parameter ranges.
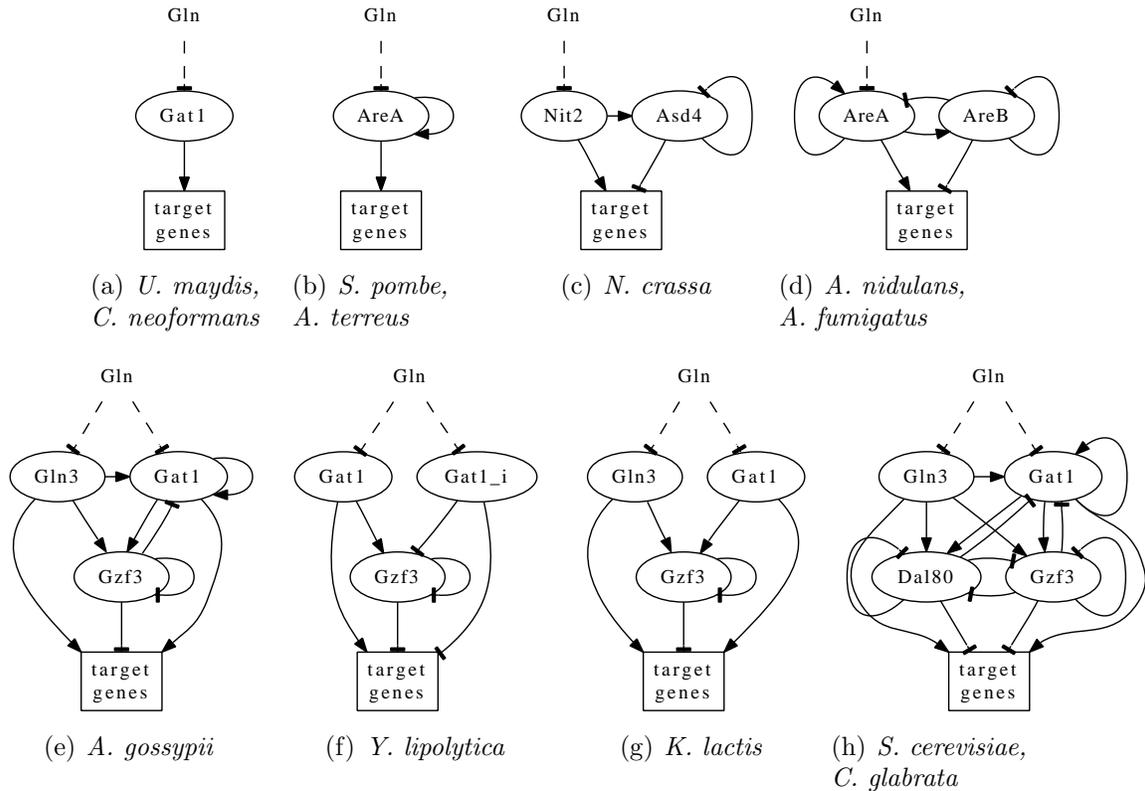
Figure 20: Potential topologies of regulatory networks derived from the sequence analysis. *Gln* stands for the level of readily available nitrogen sources, e.g. Glutamine. Gat1/AreA and Gln3-like factors, were assumed to be trans-activators with an activity regulated indirectly by Gln. All other factors were taken as inhibitors. The name of the closest related factor from either *S. cerevisiae*, *N. crassa*, or *A. nidulans* was taken. Solid lines represent transcriptional regulations, dashed ones stand for general up or down regulation of activity. Blunt arrows represent inhibitory, normal arrows activating interactions. For the topology maximal connectivity was assumed, even for cases as in the *Aspergilli*, and *S. cerevisiae*, in which some of the depicted regulatory interactions have not been found.

As a starting point for the model, the model of GATA-3 regulation in human Th2 cells as described in Höfer *et al.* [192] was taken and adapted.

The modelled cell consists of a nucleic and cytoplasmic compartment, with mRNA transcription and transcription factor binding confined to the nucleus, and translation taking place in the cytoplasm. mRNA is exported to the cytoplasm, translated, and the resulting proteins, the transcription factors, can be imported into the nucleus. Import of the activating transcription factors Gat1p and Gln3p from the cytosol into the nucleus is hindered by reversible complex formation with Ure2p (see figure 21).

Transcription of genes is modelled with low basal level in the absence of transcriptional activators, and a higher level if the regulatory region is bound by transcriptional activators. As a simplification, two independent binding sites that can be bound by both activators and inhibitors are assumed. Transcriptional activation is modelled to occur only if both sequences are bound by an activator. The underlying assumption for this is that GATA factors can bind as monomers, and that Dal80p and Gzf3p act as competitive inhibitors to Gat1p and Gln3p [84, 142].

As experimental data have indicated that only GAT1 and Dal80 expression are regulated by other GATA factors during NCR [50, 68, 84, 153, 359] (see Tab. 6), transcription factor binding was only considered for these two genes.

The level and quality of the nitrogen source was subsumed in a single parameter, *Gln*, ranging from 0 to 100. To account for the dependence of Gat1p and Gln3p sequestration in *S. cerevisae* on the nitrogen availability, the rate of complex formation, and with it the dissociation constants of the complexes Gln3p-Ure2p and Gat1p-Ure2p, is linearly varied by the value of *Gln*.

While actual complex formation has only been experimentally verified between Gln3p and Ure2p [34, 38, 263], there exists evidence that Ure2p also regulates Gat1p by similar mechanisms: transcriptional activation by Gat1p as well as localisation of an EGFP-Gat1p construct have been found to depend on relative expression levels of GAT1 and URE2 [91], and interaction between Ure2p and Gat1p has been observed in two-hybrid screens [378]. Therefore, in the model an identical mechanism was assumed for Gat1p and Gln3p sequestration. There seems to exist at least an additional, Ure2p independent, mechanism for Gat1p sequestration [142], which for the sake of simplicity and lack of mechanistic detail is neglected in this model. A graphical representation of the finally created model is shown in figure 21.

3.3.1 Parameter Derivation

The values of parameters and initial concentrations have a great influence on the behaviours shown by a dynamical model. Therefore the literature was searched to obtain physiologically feasible ranges of parameters. As the model is analysed over ranges of parameter values and not meant to directly reproduce quantitative
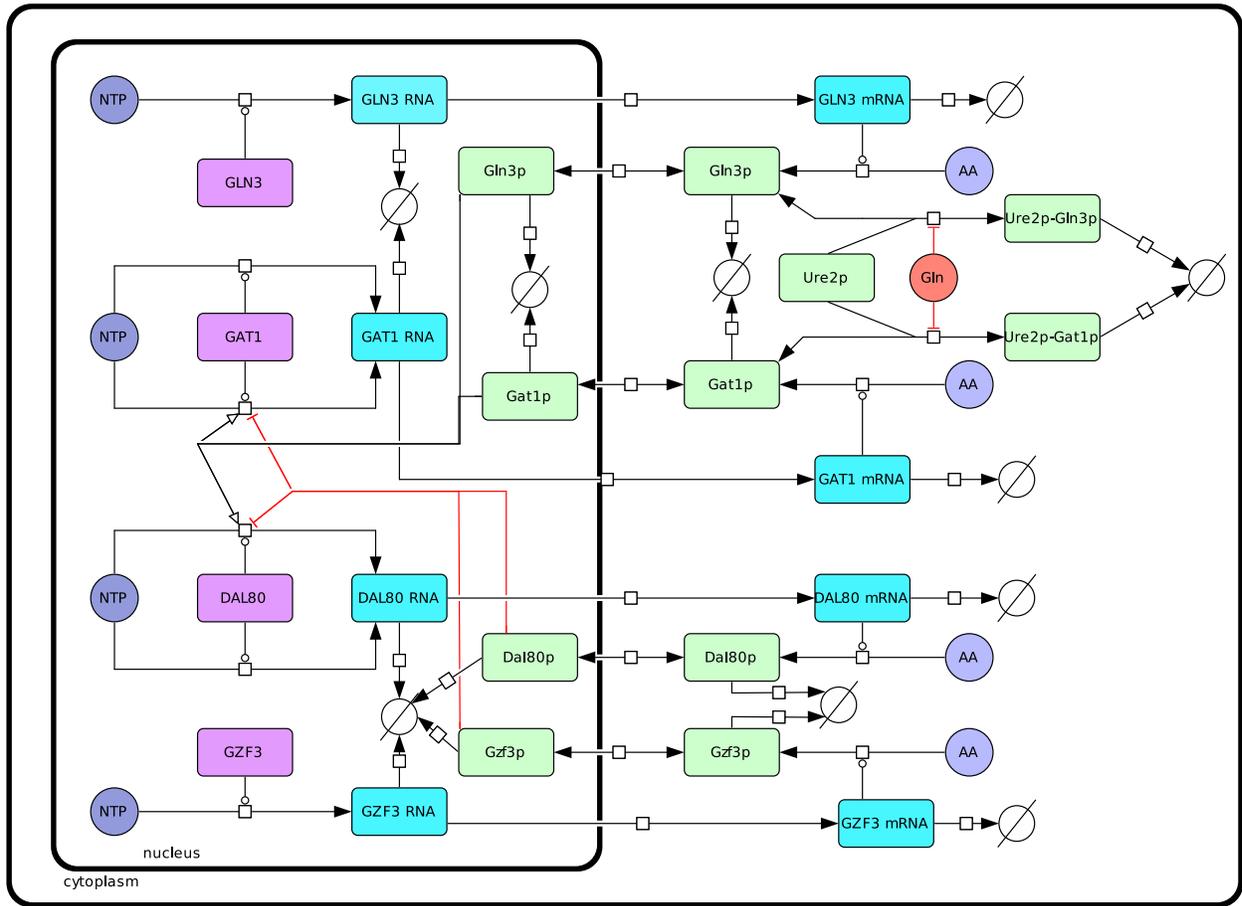
Figure 21: Reaction diagram of the base model of GATA factor regulation involved in NCR in *S. cerevisiae*. NTP and AA represent nucleotides and aminoacids as the building blocks for RNAs and protein, respectively. Amongst macromolecules, purple stands for genes, blue for RNAs, and green for proteins. *Gln* represents the level of available nitrogen, or glutamine, and inhibits the complex formation between Ure2p and Gln3p or Gat1p, respectively. The figure was created as a process diagram compliant to the Systems Biology Graphical Notation (SBGN [239]) using CellDesigner 4.2.

experimental results, the values retrieved from the literature were only taken as starting points for approximations. Whenever possible, values for *S. cerevisiae* were used. The cell size was assumed to be 33 femtolitre, with a nuclear compartment of 3 fl and a cytoplasm of 24 fl (total cell volume minus the volumes of mitochondria, vacuole, and nucleus) [411, 446]. All amounts of involved components and parameter values were converted into molecules, femtolitre, and minutes for easier comparison.

As no data for the dissociation constants of GATA transcription factors in *S.*

cerevisiae were available, data from various other species were compared (Table 4). Dissociation constants were estimated to range from 10 to 100 nM. For binding of Ure2p to Gln3p, and Gat1p respectively, a strong binding with a K_D of 50 nM was assumed under high nitrogen availability. Basal transcription rates, mRNA half-lifes and abundances as well as protein numbers per cell were obtained from published single cell studies (see table 5) [144, 195], and an average chosen for all GATA factors. Due to lack of data, a half-life of 45 min was assumed for proteins. Based on the single cell data for protein and mRNA abundance, an average translation rate was calculated using the models differential equations for mRNA translation and protein degradation under a steady state assumption.

DNA:GATA factor binding

protein	K_D [nM]	K_D [mol/fl]	source
AreA	18	10	[76]
GATA3	73	44	[237]
Fep1	63	38	[323]

Table 4: Data on GATA factor DNA binding constants.

The predicted time-courses of mRNA expression after a switch from a good to a bad nitrogen source were compared to a comprehensive set of measured time-courses for all four GATA factors [50] (figure 22). As neither the temporal sequence nor the relative levels of expression coincided, the basal and maximal transcription rates of the inhibiting GATA factors Gzf3p and Dal80p, as well as their binding affinity were manually adjusted.

After lowering basal transcription and affinity of the inhibiting GATA factors, the resulting time courses after switching from a good to bad source of nitrogen (figure 23(d)) reproduce the temporal sequence of activation and the relative expression levels at least qualitatively. The key characteristics of the experimental time series, a much faster increase of DAL80 than GAT1 mRNA levels, and a slow decline after reaching a maximum, are also reproduced by the model. A global parameter fit using the tool COPASI proved to produce better matches, but due to the little data available and the significant measurement errors, this strategy was abolished. The parameters finally chosen are given in appendix C.1 page 165 together with the kinetic laws and initial conditions. An SBML version of the model is available under `http://www.tbi.univie.ac.at/~luen/Diss/GATA/models/`.
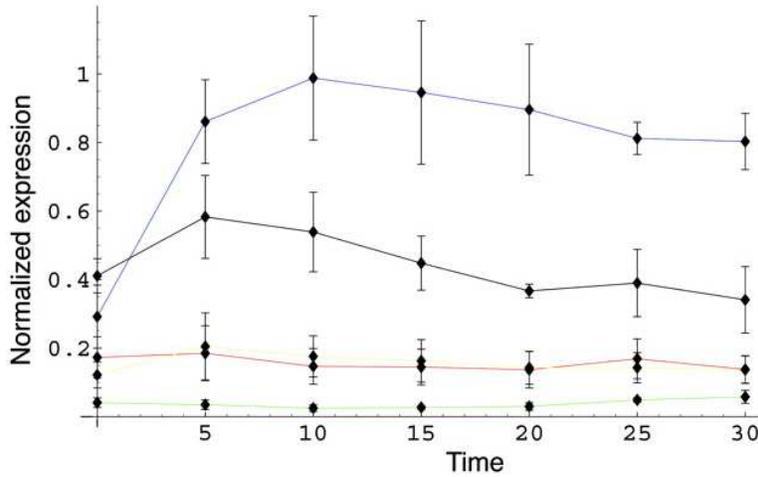
Figure 22: Experimental mRNA time courses obtained for the yeast strain BT4743 after a switch from glutamine, a good, to proline, a bad source of nitrogen. Time is measured in minutes. DAL80: blue, GAT1: black, GLN3: red, GZF3: yellow (next to GLN3), URE2: green. (Reproduction of figure 2B from Boczko *et al.* 2005 [50] with kind permission of the author.)

Transcription rates and mRNA half-lives			
Gene	transcr. freq.	mRNA half-life	source
GZF3	0.012 mol/min	27 min	[195]
GAT1	0.024 mol/min	13 min	[195]
GLN3	0.007 mol/min	25 min	[195]
ADH1	2.1 mol/min	15 min	[195]
Proteins numbers per cell			
Protein	molecules per cell	localisation	source
Gat1p	1180	c&n	[144]
Gzf3p	319	c&n	[144]
Ure2p	7060	c	[144]
Gln3p	589	c&n	[144]

Table 5: Transcription frequencies (in molecules per minute per cell) and mRNA half-lifes as well as protein numbers per cell in *S. cerevisiae* grown in medium with glutamine.

3.3.2 Model Validation

To validate the model predictions for the steady-state behaviour of the transcriptional activation patterns of various deletion and over-expression strains were compared to experimental data from the literature. For the *in silico* experiments gene-knockouts were simulated by setting the respective gene number to zero. The steady-state concentrations for high nitrogen availability were used as initial conditions for deriving the steady state at low nitrogen availability.

(a)

(b)

(c)

(d)
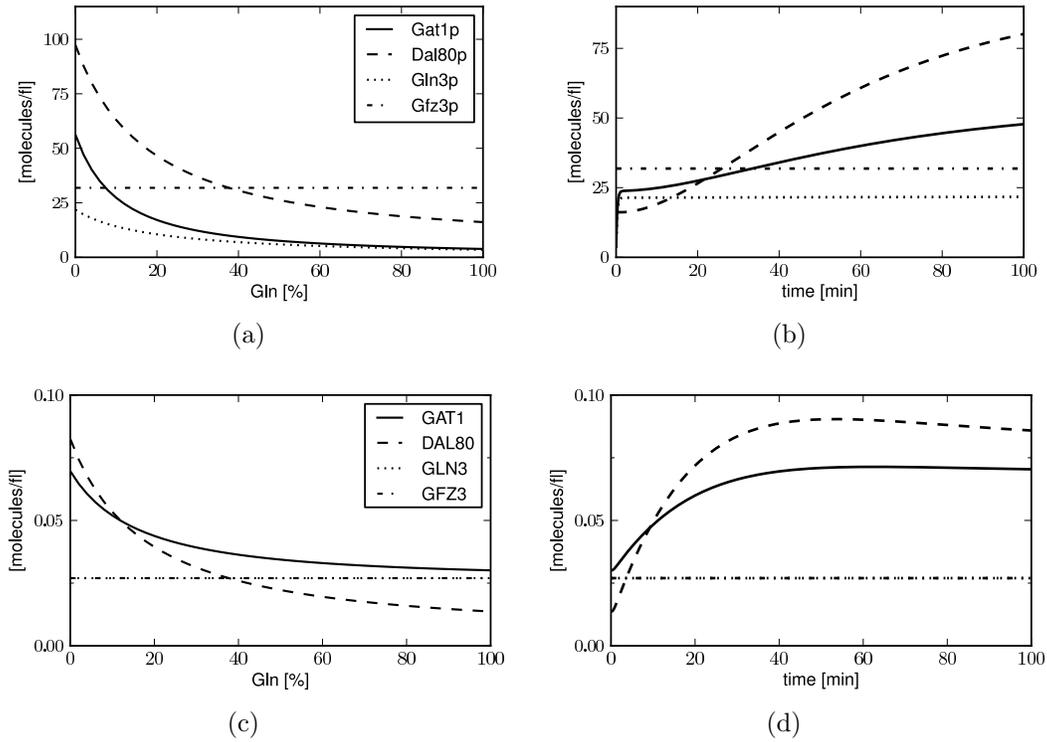
Figure 23: Steady state (left) results under varying levels of nitrogen availability (Gln) and time course (right) after a shift from a medium of high nitrogen availability ($Gln = 100$)to low ($Gln = 0$). The upper row shows the concentrations of transcription factors in the nucleus, the lower the cytosolic concentrations of mRNAs.

	GAT1	DAL80	GZF3	GLN3	DAL5	source
+ Rap	3.1	8.2	-	-	-	[68]
+ Rap	3.5	4.1	~1.5	~1.5	48.9	[173]
Pro/Gln	2.3	7.7	1.1	-	25.9	[153]
Pro/Gln	2.3	1.5	-	-	14.8	[359]
Pro/Gln	~1.4	~3	~1	~1	-	[50]
± N	2.5	6.5	1.8	1.1	27.1	[51]

Table 6: Maximal fold change of mRNA transcription of genes in *S. cerevisiae* after addition of rapamycin (+ Rap), on growth on proline as a nitrogen source in relation to growth on glutamine containing medium (Pro/Gln), and with or without NH_4^+ (± N) as a sole nitrogen source. The data from [50] is estimated from a graph.

Predictions for deletions of all four GATA factors, Δgat1, Δgln3, Δgzf3, and Δdal80, were compared to data from Georis *et al.* 2009 [141] (GAT1 mRNA: wt, Δgln3, Δdal80 in figure 1B, Δgzf3 in figure 7B; DAL80 mRNA: wt, Δgln3, Δgat1 figure 5B). Furthermore, ten fold over-expression of URE2, 10·Ure2p, was compared to the results in Cunningham *et al.* 2000 [91].

AreA mRNA stability $\pm$ glutamine [288]		
glutamine	half life $\tau_{1/2}$	degradation rate k_d
-	40 min	0.017 min^{-1}
+	7 min	0.1 min^{-1}

Table 7: Stability of AreA mRNA in *A. nidulans* grown in medium with and without glutamine, a good nitrogen source (from [288]).
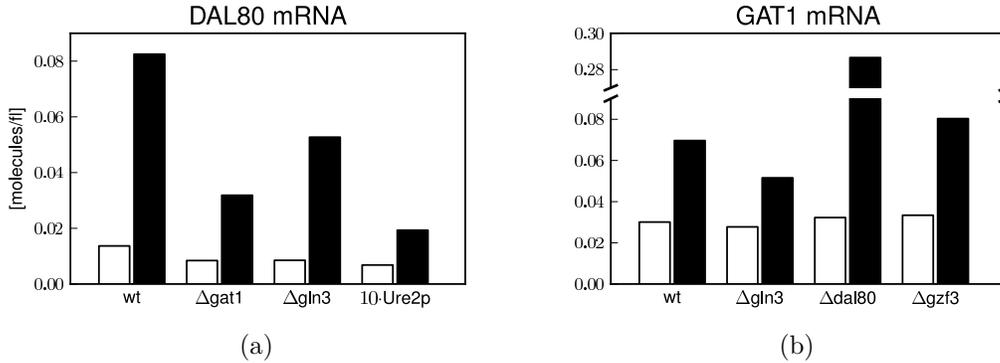
Figure 24: Steady state concentrations for wild-type, various gene knockouts, and 10 fold over-expression of Ure2p (10·Ure2p) as predicted by the NCR model for DAL80 (a) and GAT1 (b) mRNA. The white bars indicate high ($Gln = 100$), the black bars low ($Gln = 0$) nitrogen availability.

In accordance the experimental data, for growth on good nitrogen sources the model predicts hardly any change in expression for the various deletions and the over-expression (figure 24). Deletion of either one of the activating GATA factors, GAT1 or GLN3, leads to diminished expression in comparison with the wild-type (WT) after shift to a bad nitrogen source, with GAT1 having a much bigger effect on DAL80 expression (figure 24(a)). Conversely, deletion of the repressors DAL80 or GZF3 increases expression of GAT1 relatively to the WT after changing to a low nitrogen medium (figure 24(b)). As shown in Georis *et al.* 2009 [141], deletion of DAL80 shows a much bigger effect compared to deletion of GZF3, although the difference is much more pronounced in the model's prediction than in the experimental observations. In accordance to Cunningham *et al.* 2000 [91], over-expression of Ure2p leads to a greatly diminished response to low nitrogen availability.

All in all the results of the validation show that, while the model does not reproduce experimental results quantitatively, it can predict behaviours of a variety of mutations at least qualitatively.

3.3.3 The Function of the Negative Feedback

(a) nuclear Gat1p   (b) GAT1 mRNA
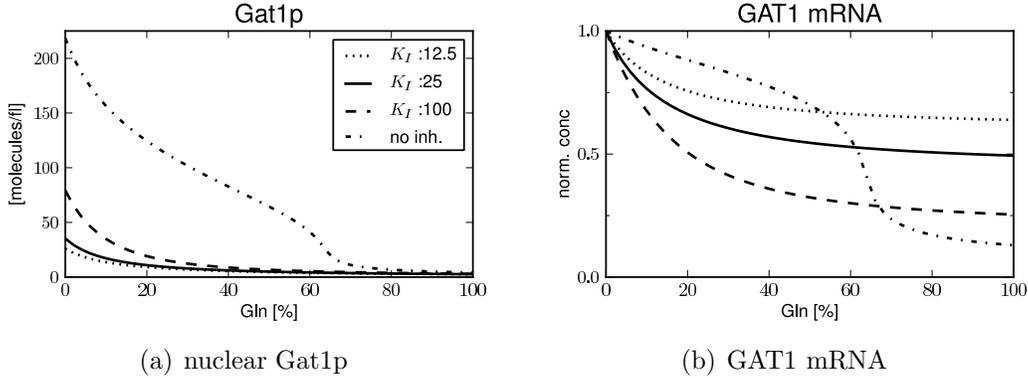
Figure 25: Effect of varying strength of repressor binding on steady-state Gat1p concentration and GAT1 expression in dependence of the nitrogen availability, *Gln*. The affinity of the repressor is indirectly proportional to the value of the dissociation constant K_I, a lower K_I-value means stronger, a higher weaker binding.

To better understand the function of the negative feedback loop of the activating GATA transcription factor Gat1p via Dal80p, the influence of inhibitor binding on GAT1 expression was inspected. Decreasing DNA binding of the inhibitor by increasing its dissociation constant K_I leads to increased sensitivity to the nitrogen availability, *Gln* (figure 25). Without any inhibition, the response curve becomes sigmoid, while increasing inhibition creates a more gradual response.

The form of the response of course depends highly on the way nitrogen availability influences the sequestration of Gln3p and Gat1p in the cytoplasm. In this model a linear dependency was assumed, while Boczko *et al.* 2005 [50] use an all-or-nothing approach. Experimental results show that genes controlled by NCR exhibit differential expression in dependence on the nitrogen source [153], fitting well to a gradual activation in dependence of nitrogen availability.

3.4 A Single Auto-activating GATA Factor

As a basic unit before gene duplication, a simplified system with only one auto-activating GATA factor, A, was assumed. The parameters and equations for the model where adapted from the base model for NCR in *S. cerevisiae*, though the system was extended slightly.
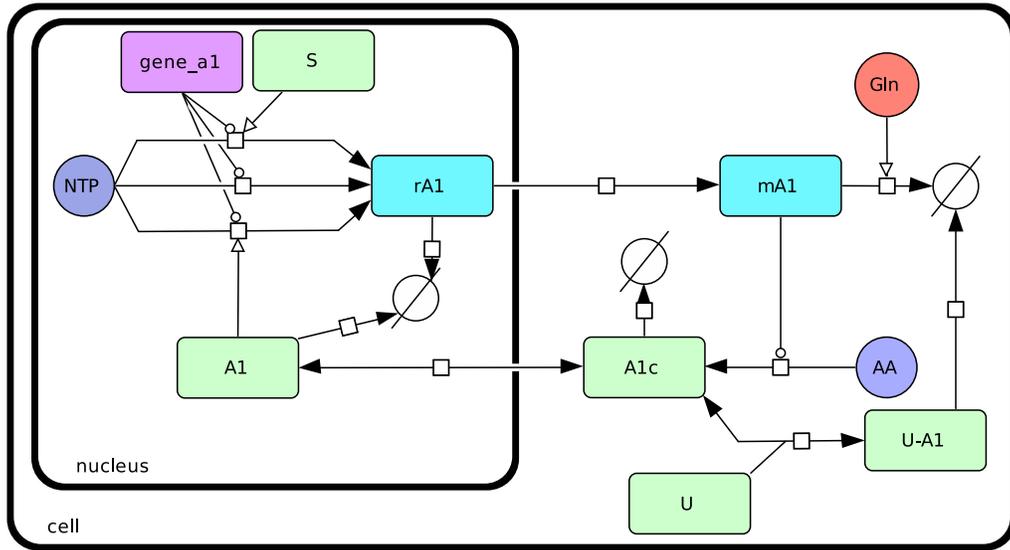
Figure 26: Reaction diagram of a single auto-activating GATA factor, A_1, with both mRNA degradation and sequestration in cytoplasm as post-transcriptional regulation mechanisms. The figure was created using CellDesigner 4.2.

To model the direct influence of nitrogen availability on the expression of A, the half-life of the A mRNA is assumed to depend on the concentration of glutamine, similar to a mechanism influencing AreA mRNA stability found in *A. nidulans* [328]. In *A. nidulans* an element in the 3'-UTR of the AreA gene has been identified to be necessary and sufficient to change the half-life of the mRNA from 40 min under low nitrogen availability to 7 min under high levels of glutamine (see Table 7) [288, 289]. Similar sequence elements have been identified in genes encoding related GATA factors in several other *Aspergilli* ([134] and Rainer Machné, pers. communication), indicating that this constitutes a general mechanism in regulation of this type of GATA factor.

As another mechanism of post-transcriptional control of A activity, the A protein is assumed to be bound and sequestered in the cytoplasm by a protein U acting in a similar fashion to Ure2p in the basal NCR model. To be able to separate the sequestration from the degradation mechanism, high and low nitrogen availability is simulated by both high and low concentrations of U and Gln.

While both mechanisms have been not been observed to occur together in the same organism, it is convenient for technical reasons to include them in the same model. By setting the total concentration of U to 0 or by equating the decay rates of mRNA bound to Gln and free mRNA, the model can be restricted to

one of the two mechanisms.

Finally, an additional activating transcription factor, S, altering expression of A was introduced, to allow for yet another signalling input on the system. The regulatory region of GAT1 features several binding sites for the transcription factor Swi5p that activates expression of genes in the M/G1 and G1 phase [259], and has furthermore be suggested to be controlled by the forkhead proteins Fkh1p and Fkh2p [450]. As all of these transcription factors are involved in cell cycle specific gene regulation, this could constitute at a potential link between cell cycle and nitrogen metabolism, for example to account for higher demands of nitrogen in specific phases. The main motivation, though, for including this additional signal, was to make the model a more general representation of gene-regulatory modules with auto-regulation.

Not to impose any further non-linearities, S is modelled to bind non-cooperatively at a single, independent site. Also, the maximal rate of transcription from a promoter activated by S was chosen to be only $\frac{1}{75}^{\text{th}}$ of the rate of a promoter bound by A, limiting the influence of S. To further broaden the scope of behaviours, the basal transcription rate of A was lowered, so that S can also be viewed as a change of basal transcription, for example by chromatin remodelling.

A reaction diagram of the simplified model of an auto-activating GATA factor is shown in figure 26. The parameter values for this model were assumed to be similar to the values used for the basal NCR model, with only slight changes to move it into a bistable region (see appendix C.2 page 167 for details and reactions). An SBML version of the model is available under `http://www.tbi.univie.ac.at/~luen/Diss/GATA/models/`.

3.4.1 Steady States and Regions of Bistability

In the case of the system with a single transcription factor it is possible to derive a closed analytical form for its steady states. The parameters for transcription, expression, and decay can be combined into two variables, α and δ. δ subsumes induced and basal transcription and depends on the signal S, while α subsumes auto-activation, and expression against degradation and retention in the cytoplasm, and depends on the concentration of the cytoplasmic factor U, as well as the influence of the nitrogen level, *Gln*, on mRNA stability.
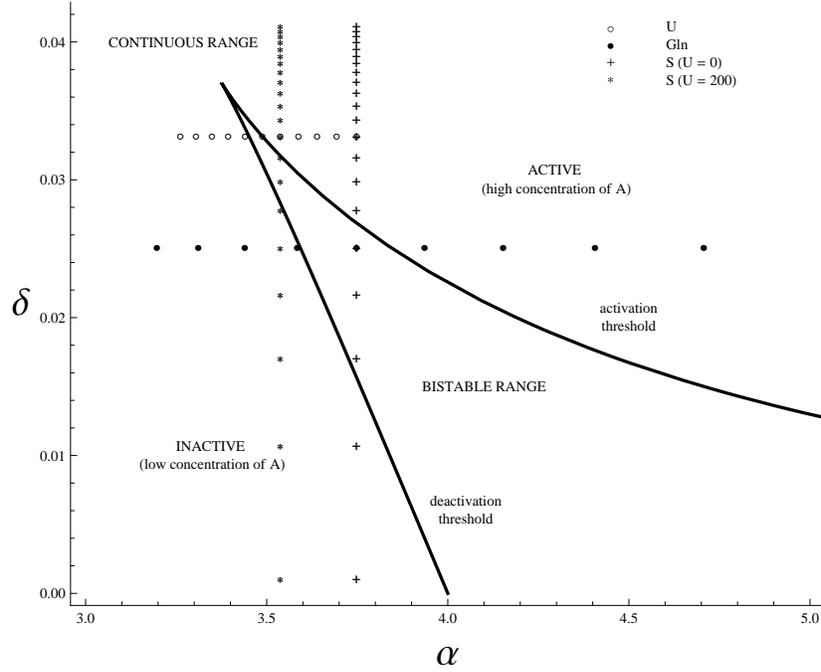
Figure 27: Bifurcation diagram of a single auto-activator in the δ-α plane. The diagram displays solutions of equation (131), calculated using the base values for transcripiton, translation, and decay. Dotted lines show effects of variations in three external parameters: the concentration of cytoplasmic A binding protein U ($\circ$), the effect of *Gln* on mRNA decay ($\bullet$) and the transcription factor S ($+$). The concentration ranges are U: 0-600 molecules/fl, Gln: 6 - $18 \cdot 10^6$ molecules/fl, and S: 0 - 500 molecules/fl. For variation of S the U dependent activator was assumed with a concentration of U of either 0 or 200 molecules/fl. Calculations performed using Mathematica 7.0

Defining:

$$\delta = \frac{V_{aS}}{V_{aA}} \cdot \frac{S}{K_S + S} + \frac{V_b}{V_{aA}}$$

$$\alpha = \frac{kex_m}{kex_m + D_r} \cdot \frac{ktl \cdot (1 + \xi_G)}{D_m + D_{CG} \cdot \xi_G} \cdot$$

$$\frac{kim_p \cdot (V_{aA}/K_A)}{D_{pc} \cdot kex_p + D_p \cdot (D_{pc} + kim_p) + D_{pc} \cdot (D_p + kex_p) \cdot \xi_U}$$

with:

$$\xi_G = \frac{G}{K_{CG} + D_{CG}}, \; \xi_U = \frac{kU_{ass} \cdot U}{kU_{diss} + D_{pc}}$$

the steady states of the system can be derived as:

$$\frac{A}{K_A} = \alpha \cdot \left(\delta + \left(\frac{A}{A + K_A} \right)^2 \right) \tag{131}$$

This means that the system can either possess one stable, or two stable and one unstable steady state in the positive orthant. Figure 27 shows the bifurcation diagram of the system in the α-δ plane and the effects of the variation of the concentrations of U, and S as well as the influence *Gln* dependent mRNA stability.
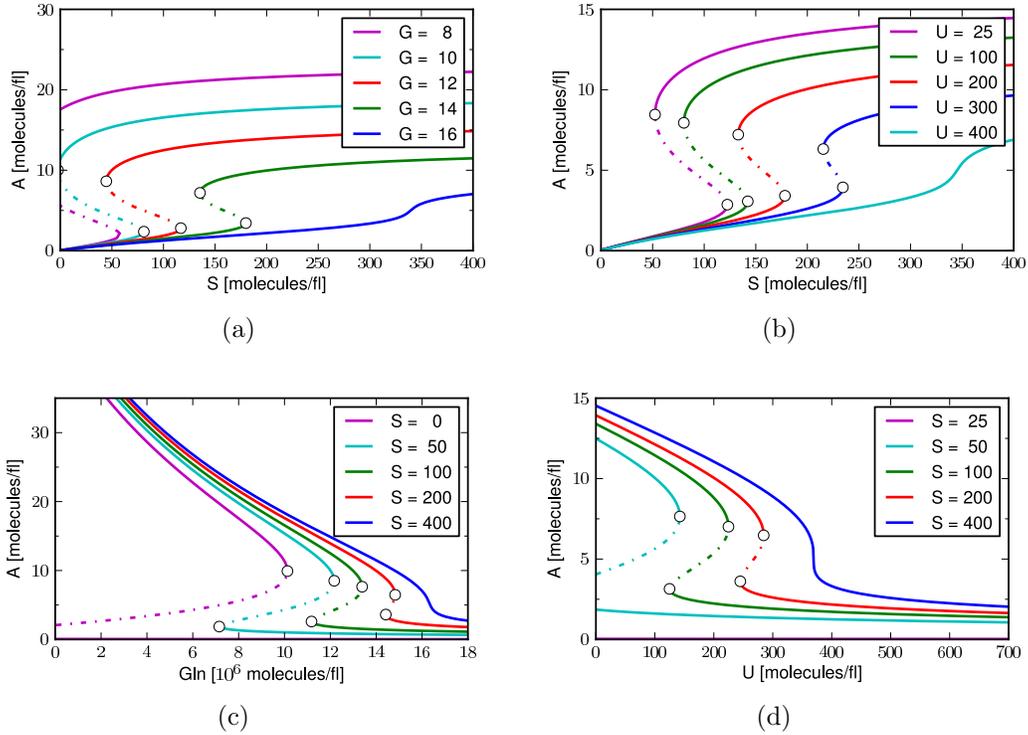
Figure 28: Bifurcation diagrams of expression of A in the model of the single auto-activator in dependence of S, Gln, and U. Solid lines indicate stable, broken unstable steady-states, circles indicate saddle-node bifurcations. For the variation of U (subfig. b and d), *Gln* was set to 12.

The system can show bistability over a range of values of each of the three independent parameters S, Gln and U. With the parameters of the basal model, it can switch reversible and with hysteresis from high to low expression levels of A in dependence of Gln and U (see Fig. 28). For Gln a level of 20 mM $\equiv$ 12 $\cdot 10^6$ molecules/fl was assumed to be high, for U a concentration of 300 molecules/fl is equivalent to conditions of high N availability in the basal NCR model.

This shows, that both the sequestration, as well as the mRNA degradation mechanism on their own are sufficient to achieve reversible switching under the assumed parameter values. For high values of either S, Gln, or U the bistable region collapses to sigmoid behaviour and finally to hyperbolic response curves.

3.5 Effects of Gene Duplication on the Simple Auto-activator
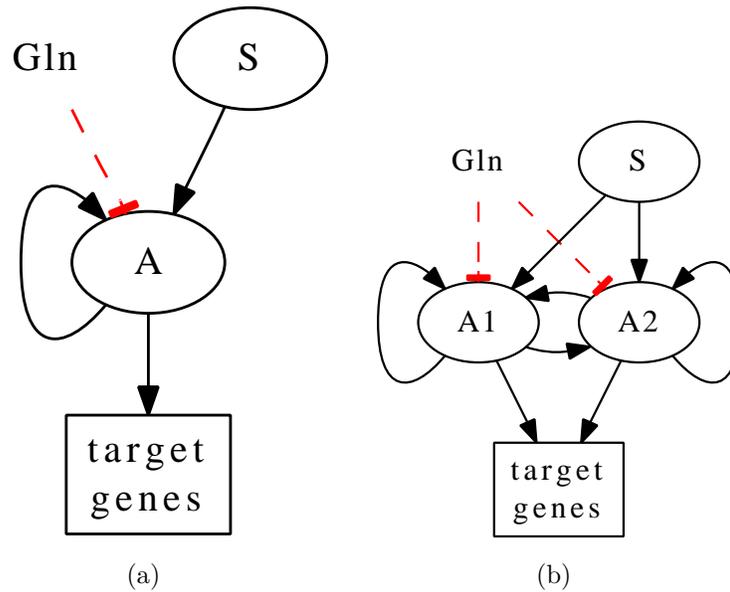
Figure 29: Schematic representation of the interactions in the single auto-activator and the same system after a gene duplication. Solid lines indicate transcriptional regulation, broken lines post-transcriptional. For simplicity only *Gln* is shown as a post-transcriptional inhibitor.

To analyse the effects of gene duplication, a model with two identical copies of gene A, A_1 and A_2, was created, and additional parameters were introduced to vary both the affinity of the transcription factors to the original promoter region, and the affinity of the promoters regions of the genes to the transcription factors (see Fig. 29). While divergence of the produced paralogous genes can also influence other properties, such as mRNA and protein stability, these effects were neglected for a first study.

Interestingly, the model predicts pronounced gene dosage effects on the signal-dependent expression dynamics. For the signal S the switching becomes irreversible, while for Gln, and U the bistable region moves to nearly ten fold higher concentrations (see figure 30).

The change of region of maximal sensitivity to nitrogen availability, as well as the irreversibility switching in response to S, should severely disrupt the normal functioning of this gene-regulatory module. In this case, a gene duplication event could not only allow mutation through functional redundancy [311, 312], but could actually create pressure to relieve an adverse signal-dependency by mutation.
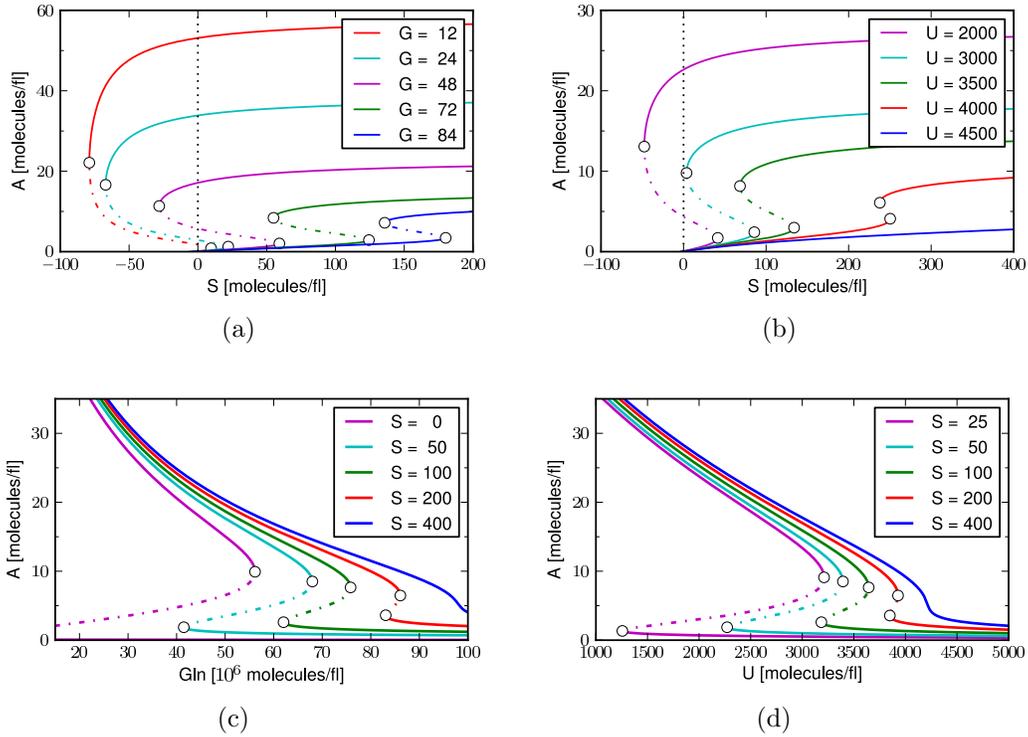
Figure 30: Bifurcation diagrams of expression of A in dependence of S, Gln, and U after a gene-duplication event.

3.5.1 Relieve by Feedback Loop Disruption

As the effects of gene duplication on the behaviour of the model are dramatic, ways to restore switching back to the regions of the single auto-activator were inspected. While a lot of different parameters could be influenced by mutation, such as the protein and mRNA half-lives, or the basal and maximal transcription rates, considering to many factors quickly leads to a combinatorial explosion of parameter combinations to inspect. To limit the different possibilities, only three parts of each gene where varied in isolation as exemplary cases. In the non-coding

region these were the binding sequence for the signal S and the GATA motif, in the coding region, the DNA binding region.

(a)

(b)

(c)

(d)
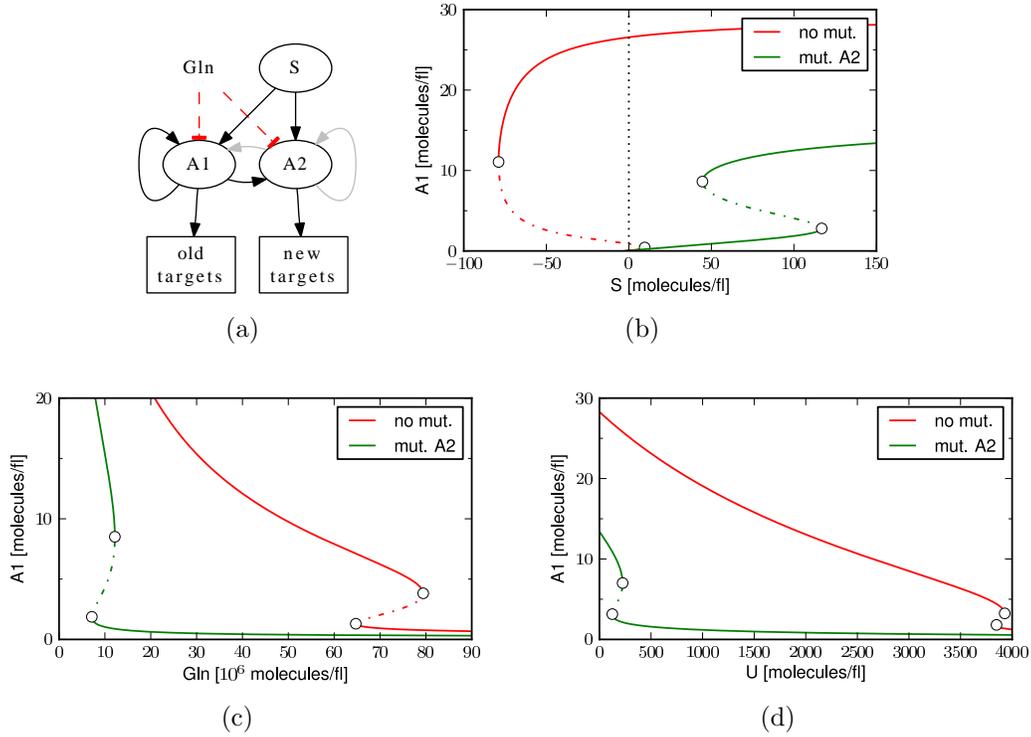
Figure 31: Restoration of the original switching range by mutation and change of the specificity of the `HGATAR` binding domain in the coding region of gene A_2. This abolishes the auto-activatory loop of A_2 and creates a feed-forward loop of S over A_1 on A_2. In subfigure (a) grey lines indicate interactions lost by mutations. In the steady-state response curves, red lines show the curves after gene duplication, the green lines after mutation of the DNA-binding domain of gene A_2. (parameters used in subfig. (b): Gln = 12, U = 0; (c): S = 50, U = 0; (d): Gln = 12, S = 100)

A few mutations of the paralogous genes can suffice to bring switching back into the original regions. The simplest way is mutation of the DNA binding region of one of the proteins, changing its affinity to the `HGATAR` motif.

Figure 31(a) shows the regulatory network resulting from the extreme case of such a mutation leading to complete abolishing of binding of A_2 to the common `HGATAR` motif. Such a mutation destroys the positive feed-back of A_2 on its own expression both directly and via the activation of A_1 expression, resulting in a feed-forward loop. As the binding affinity of A_2 changes, this could also influence the regulation of downstream targets, and could allow, or require A_2 to find new regulatory targets. The switching behaviour of A_1 expression is

118

completely restored to the one of the single auto-activator (figure 31).

(a)   (b)   (c)

(d)   (e)
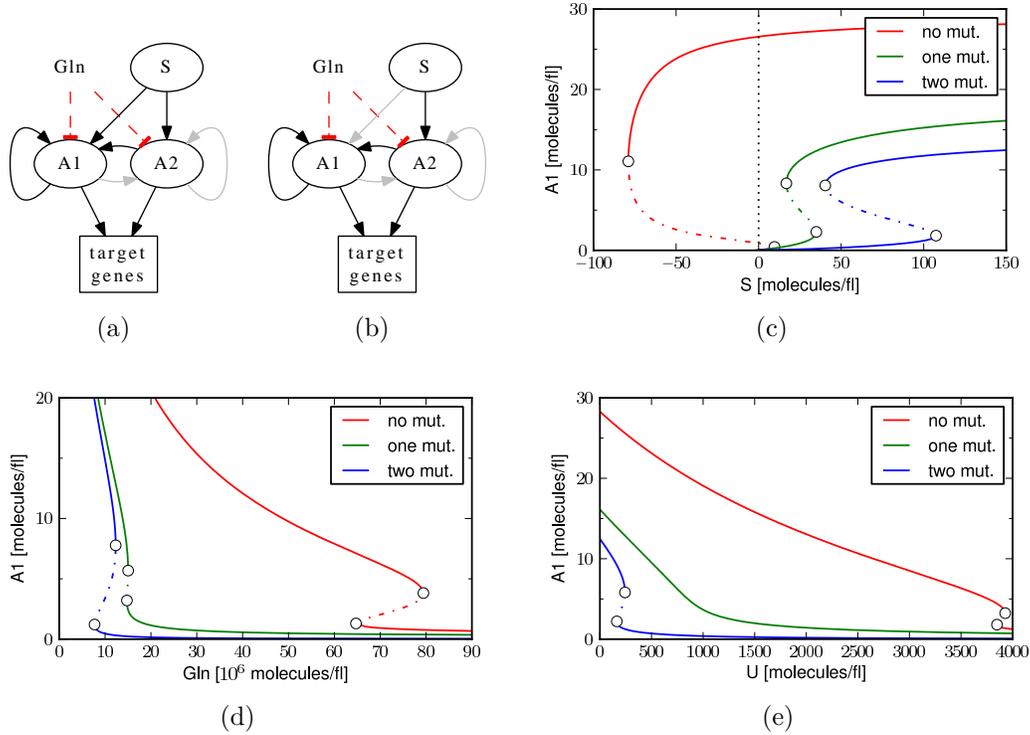
Figure 32: Restoration of the original switching range by cascade formation. In the first step mutation of the regulatory region of gene A_2 leads to change or total loss of the `HGATAR` site (subfig. (a), green curves in (c),(d), and (e)), in the second the binding site for S in the regulatory region of gene A_1 vanishes (subfig. (b), blue curves in (c),(d), and (e)). Parameters as in 31

Alternatively, cross regulation of the two paralogous genes can be diminished by mutation, or complete loss, of the `HGATAR` sites in the upstream regulatory region of one of them. Again, complete loss of this motif leads to a form of feed-forward loop (figure 32(a)). This kind of regulatory motif is similar to the cross-regulation of GLN3 and GAT1 found in *S. cerevisiae* and *A. gossypii* (figures 20(e) and (h)). While this step does not completely restore the switching behaviour to the one before duplication, it moves switching in dependence on S back into the reversible region and on Gln into ranges similar to before (figure 32, green curves).

A further loss of the binding site of the signal S in the other paralogous gene (32(b)) nearly completely restores the switching behaviour to the ranges prior to duplication (figure 32, blue curves). In relation to the signal S this network constitutes a cascade, which is intriguing, as a three tier cascade of GATA factors

has been suggested in endomesoderm development of the nematode *C. elegans*. In this and other nematodes a the two GATA-type transcription factors MED-1 and 2 have been indicated to activate the GATA factors END-1 and 2, which again seem to regulate expression of the, potentially auto-regulatory, GATA factors ELT-2 and 7 [261].

(a)

(b)

(c)

(d)
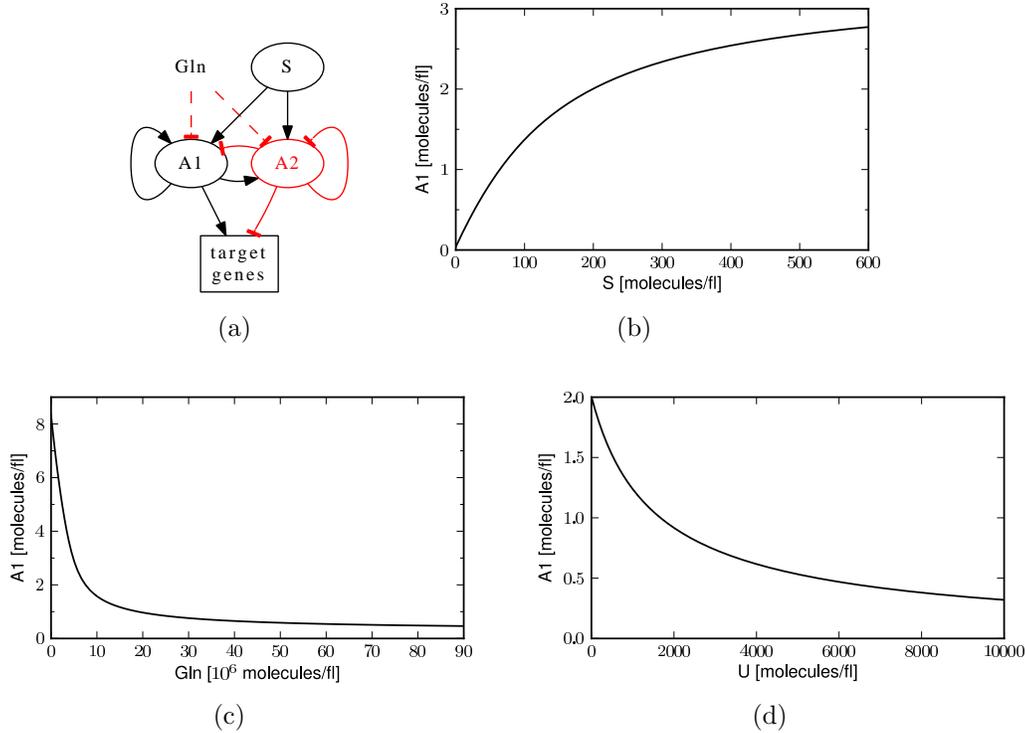
Figure 33: Diagram of the regulatory interactions (a) and response curves after mutation or loss of the TAD in the coding region of gene A_2 turns A_2 from a trans-activator to a competitive repressor.

3.5.2 Loss of the Trans-Activation Domain

Another interesting mutation changing the behaviour of the system is loss of the trans-activating domain (TAD) of one of the paralogous genes. This converts an activating GATA factor into a competitive repressor, similar to Dal80p and Gzf3p in *S. cerevisiae*, both of which have been indicated to have evolved from a GLN3- or GAT1-like, trans-activating ancestor [256].

Simple change of one paralogue to a repressor (figure 33(a)) creates a negative feedback loop and alters the behaviour of the system dramatically. Similar to

the basic NCR model, the system exhibits only gradual, hyperbolic response curves and does not display bistability on variation of S, Gln, or U for the base parameter values (figure 33).

name	value		description
	base	slow. osc.	
ba1	1	2.52	modifier for GATA factor binding to gene A1
Va1_S	0.00125	0.00034	transcription rate for gene A1 bound to S
Va2_S	0.00125	0.000032	transcription rate for gene A2 bound to S
D_pc	0.015	0.0023	A2 protein degradation constant (cytoplasm)

Table 8: Values of parameters changed from the base model to achieve slow oscillations.

As gene regulatory networks with a negative feedback have been shown to exhibit bistable and oscillatory behaviour, Dr. James Lu used his previously developed Mathematica package for inverse eigenvalue analysis [257] to find parameter values leading to stable oscillations. To find the minimal number of changes in parameters, this method uses a sparsity constraint, that is, it employs a penalty function increasing with the number of parameters changed.

Three changes in parameter were found to be necessary to create a system that can exhibit both stable oscillations, and also shows switching behaviour in ranges similar to the system before gene duplication. First the dependence of both genes on the signal S needs to be weakened, less for the activating GATA factor A_1, but to a much higher degree in case of the repressor A_2. Furthermore, the repressor needs to be stabilised in relation to the activator, for example by changing its degradation rate (see table 8).

The resulting system exhibits very slow oscillations with a period of approximately 2000 minutes (figure 34). As shown in the bifurcation diagrams (figures 34(c) and (d)) the system has two curves of saddle-node bifurcation points in the S-Gln and S-U planes vanishing in a cusp point. From one of the saddle-node curves a Hopf-curve emerges at a Bogdanov-Takens bifurcation, and with it stable limit cycle oscillations. This means, that the system can exhibit bistability and stable oscillations in dependence on S, Gln, and U (figure 35). For a small region in the S-U and S-Gln plane falling on the Hopf-curve between the two saddle-node curves, the system shows both bistable switching and stable oscillations together (figure 36(b)), while for a bigger range of combinations it just switches from low expression of A_1 to slow oscillations with a big amplitude
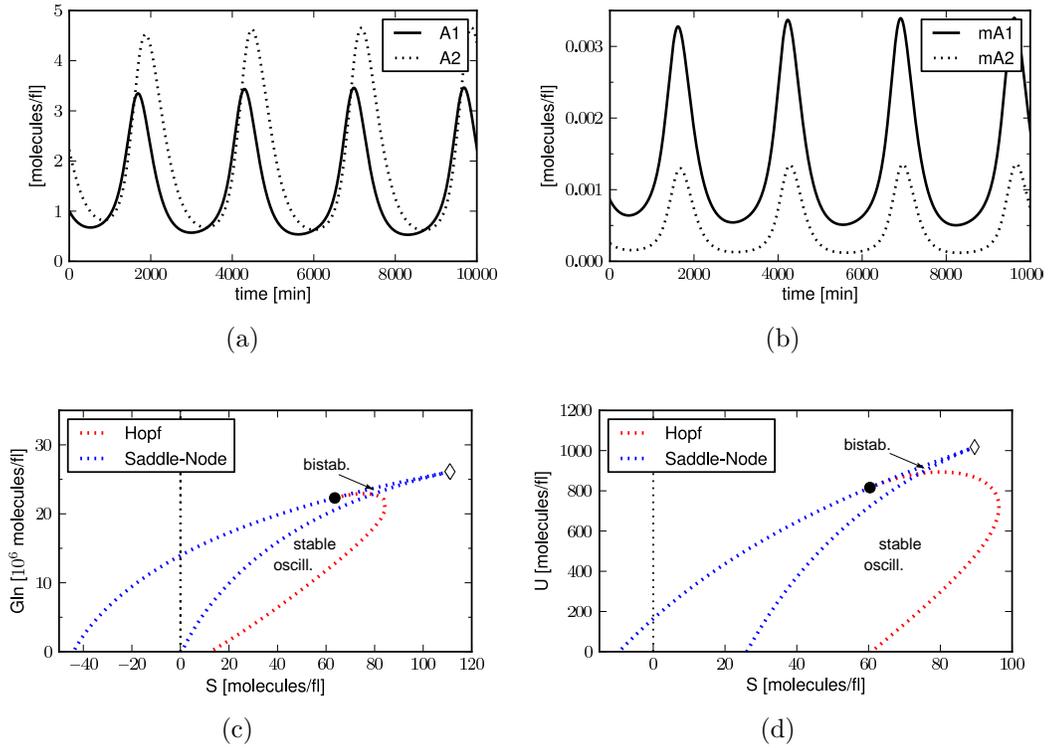
Figure 34: Upper row: time courses of protein (a) and mRNAs (b) of the slowly oscillating system (S = 50, U = 0, and Gln = 12). Lower row: bifurcation diagrams in the S-Gln (with U = 0, (c)) and the S-U (with Gln = 12, (d)) plane. Blue line indicate saddle-node curves, red Hopf-curves. The open diamond stands for a Cusp-point, the black disk for a Bogdanov-Takens bifurcation.
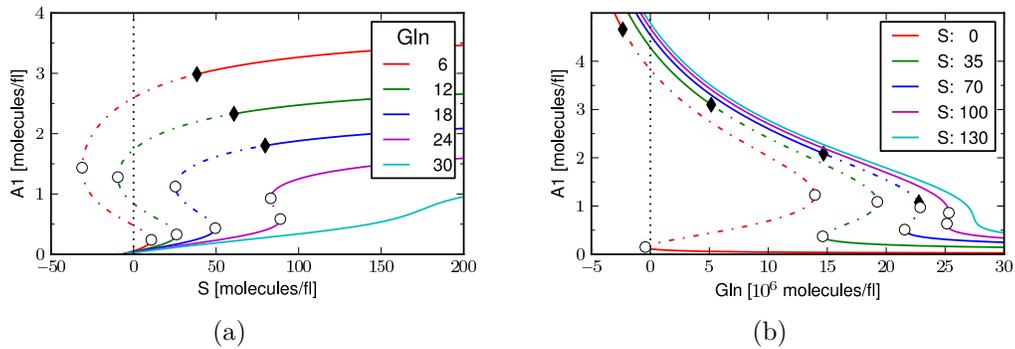
Figure 35: Steady state curves in dependence of S for various values of Gln (a)) and Gln for various values of S (b)) with U = 0. The circles stands for Saddle-Node bifurcations, the black diamonds for Hopf bifurcations. Solid lines connect stable steady states, dashed ones unstable. For all calculations U was set to 0.

(figure 36(a)).

As the oscillation periods of this system lie outside the biologically meaningful
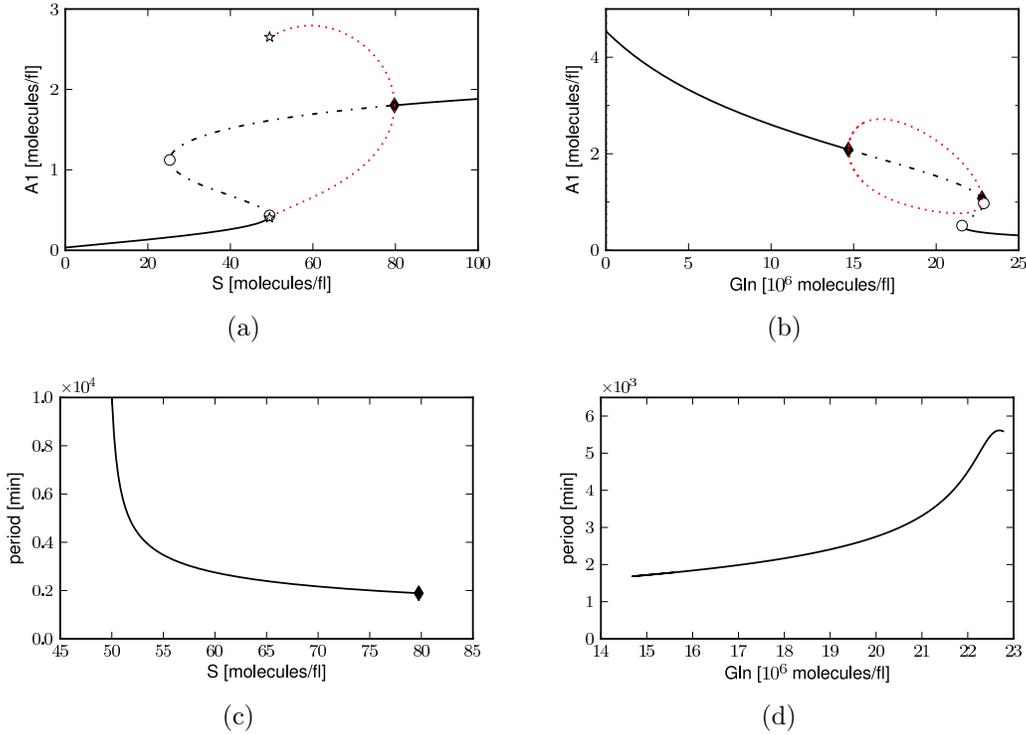
Figure 36: Steady states (upper row) and periodes (lower row) for varying values of S (left, U = 0, Gln = 18) and Gln (right, U = 0, S = 75). For the upper panels the symbols are as in 35. The dotted, red line shows the minimal and maximal concentrations during oscillations, the asterisks stand for Period Doubling bifurcations.

range by at least one magnitude, again inverse bifurcation analysis was used to obtain oscillations with a smaller period. It was found, that faster oscillations require many more parameter changes. The time-courses and bifurcation diagrams for a system with a period of approx 75 minutes are shown in figure 37. For this set of parameters, the system simply changes from stable oscillations to a single stable equilibrium in dependence of any of the three signals, S, Gln, and U.

To achieve these faster oscillations 21 changes of parameters were needed, some for more than a factor 10. The most notable changes are the activator having a decreased stability, and a stronger affinity to both genes. Also, the effect of the activator, the activated transcription rate, is increased.
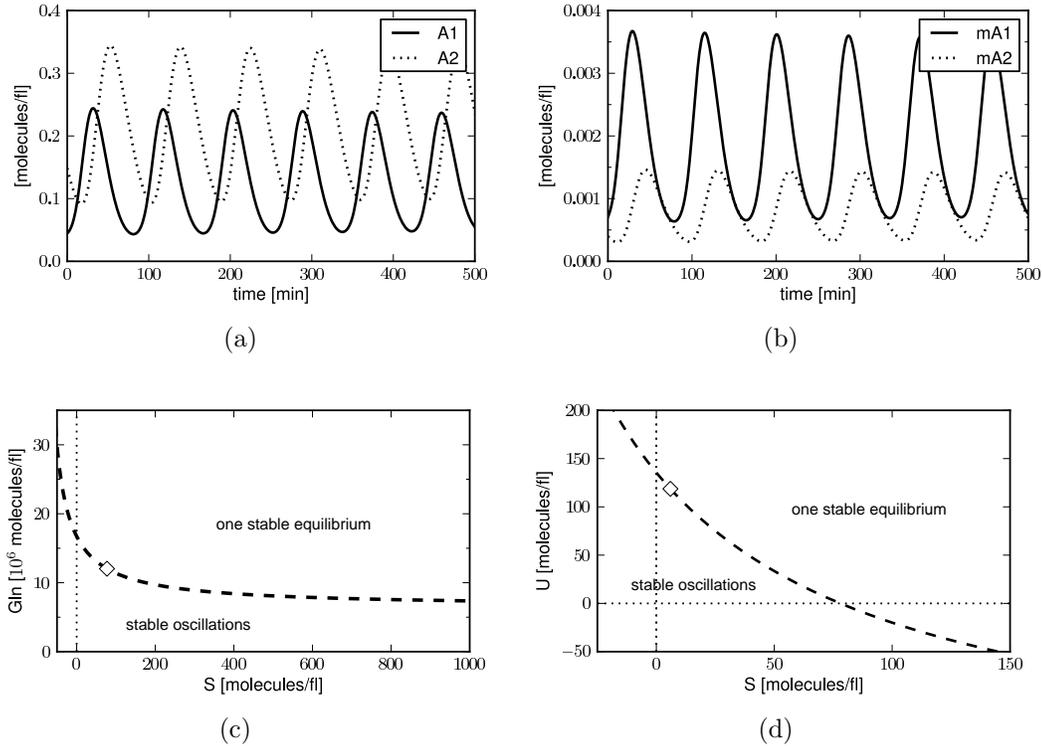
Figure 37: Upper row: time courses of protein (a) and mRNAs (b) of the fast oscillating system (S = 50, U = 0, and Gln = 12). Lower row: bifurcation diagrams in the S-Gln (with U = 0, (c)) and the S-U (with Gln = 12, (d)) plane. The white square indicates the location of a Generalised Hopf point.

3.6   Discussion

A very interesting finding is, that all GFZ3-like genes were found to potentially be controlled by GATA-type transcription factors, with all of them possessing multiple adjacent `HGATAR` sites in their upstream regions, which could indicate transcriptional control by GATA type transcription factors. GLN3-like genes, on the other hand, seem to lack adjacent GATA binding sites in their upstream regions, fitting the observation, that GLN3 is not controlled by GATA-type transcription factors in *S. cerevisiae* [84]. While the GFZ3-like gene was retained after the WGD event before divergence of *S. cerevisae* and *C. glabrata*, no duplicates of GAT1 and GLN3 were found. This loss of duplicates could be due to adverse gene-dosage effects. Further all examined fungi with more than one activating GATA factor, also possess at least one potentially inhibitory GATA factor (see fig 20). Apart from differential regulation, mentioned above, this could point to the function of inhibitory regulation in mitigation of gene dosage effects.
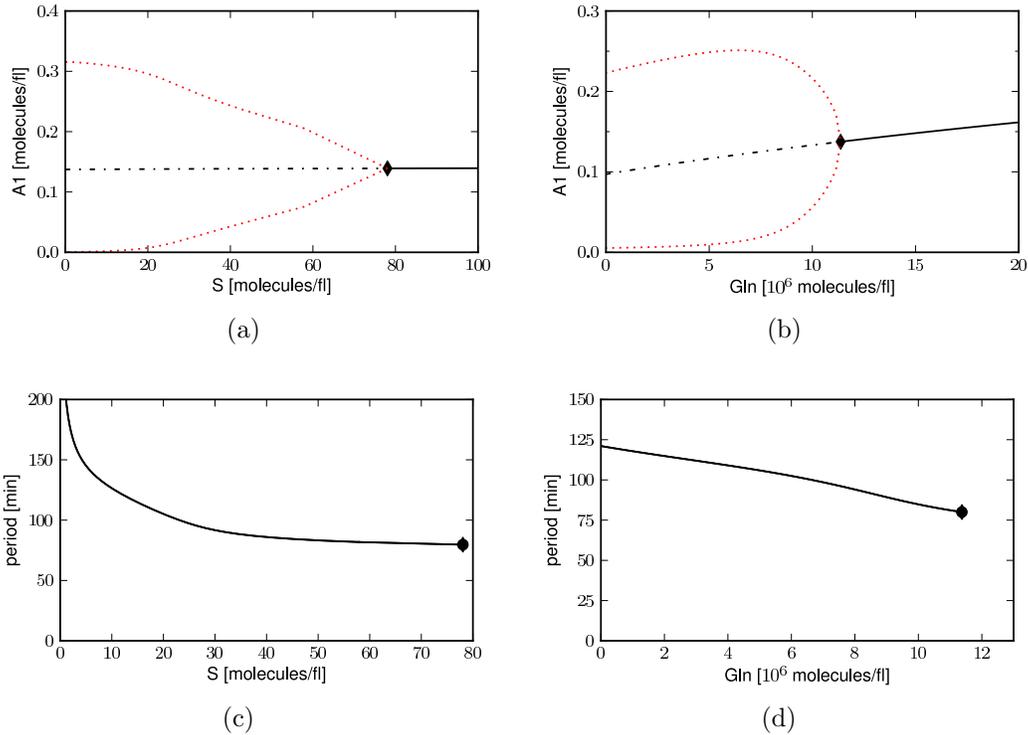
Figure 38: Steady states (upper row) and periodes (lower row) of the faster oscillating system for varying values of S (left, U = 0, Gln = 18) and Gln (right, U = 0, S = 75). The dotted, red line shows the minimal and maximal concentrations during oscillations.

While the methodology for finding and classifying potential GATA-type transcription factors cannot be directly validated, it is somewhat supported by identifying topologies similar to those assumed to exist in nitrogen dependent gene regulation in *S. cerevisae* and filamentous fungi [84, 439]. Further corroboration for the mehtod comes from the finding that GZF3/DAL80-like repressors appear to be an innovation restricted to the *Ascomycetes*, and GLN3-like genes to the *Saccharomycetes* after divergence of *Y. lipolytica*, which fits well with the independently derived results of Wong *et al.* 2008 [439].

For the dynamical modelling, the parameter values taken from the literature together with the rather simple model can qualitatively reproduce time series of mRNA expression, and predict gene deletions in accordance with experimental observations. This fit is astounding, taking into account the diverse sources for parameters and the sparse mechanistic detail incorporated. One reason for the good fit could lie in a structural robustness of the modelled network, which might lead to qualitatively similar results over big ranges of parameter values.

In *S. cerevisae* NCR, DAL80 and GZF3 are assumed to mainly fine-tune the expression of GAT1 and downstream targets [84, 439]. The results of this chapter further suggest that they could be important in creating a gradual, rather than a sigmoid or bistable, response to low nitrogen availability. Such a response could be better suited for differential expression of target genes over a variety of nitrogen sources, as found with NCR sensitive genes [153]. Further, as shown in the simplified model of the auto-activator, inhibition can mitigate effects of varying gene copy numbers, for example during the cell cycle.

While bistable expression of GATA-type transcription factors has not been shown experimentally, it has been suggested to underlie mammalian T_h2 cell differentiation [192]. Apart from cell differentiation, such as the former, or the endomesoderm determination in nematodes, bistability could be beneficial in many contexts, for example in metabolic regulation and cell cycle coupling in yeast. While the differential expression of target genes in dependence of nitrogen sources fits well with a gradual NCR in *S. cerevisiae* [153], the core nitrogen metabolism constitutes a central node between respiration and amino acid metabolism, and is increasingly recognised as a pivotal point of cellular growth regulation [72, 293] with highly conserved target genes in fungi [139]. The potential coupling of GAT1 expression to the cell cycle by Swi5p and Fkh2p [259, 450] in *S. cerevisiae*, could for example activate nitrogen metabolism in a switch-like fashion. As a cell in the growth phase has a higher demand of nitrogen, keeping NCR controlled genes switched on until a higher level of available nitrogen is reached in the cell than required to activate them in the first place, could be beneficial under conditions of insecure and quickly changing nitrogen availability in a natural environment.

For gene duplication events the model predicts profound gene dosage effects. Duplication of an auto-activatory gene, including its regulatory region, could lead to pronounced hypersensitivity to the biological signals which modulate the gene's expression through either activation or deactivation. Several mutations are possible to relieve such hypersensitivity, some leading to regulatory motifs found in yeasts and other organisms such as cascades or feedforward loops. Such transitions to relieve gene-dosage effects after gene and genome duplication could constitute a driving force behind the evolution of auto-regulatory gene networks. A contrary effect due to loss of one allele, and thereby loss of - or changes of - the regions of bistability, could also shed light on several human diseases which have been related to haploinsufficiency of the - potentially auto-regulatory - GATA-

type transcription factors GATA-2 [348], GATA-3 [114], and GATA-4 [322].

The model also predicts, that loss of the trans-activation domain in one paralogue after gene duplication could lead to a simple two-gene oscillatory system, similar in topology to the tunable oscillator of Stricker *et al.* (2008) [397]. The onset of the oscillations and their frequency could be modulated by various signals, for example by transcriptional activation, or, on the post-transcriptional level, by protein sequestration and mRNA stabilisation. Gene regulatory oscillators with long periods have been found to underlie for example circadian clocks [103], and developmental processes [186], although up to date, no oscillatory system involving GATA factors has been identified. One oscillating system possibly involving GATA factors could be underlying yeast respiration, which exhibits oscillations with periods in the range of hours under certain conditions [293]. Given that many compounds involved in nitrogen metabolism, such as glutamine, vary with the oscillations, at least some GATA factors of the NCR network could influence them.

4 MiniCellSim - a Self-Contained *In Silico* Cell Model Based on Macromolecular Interactions

4.1 Introduction

An important question in biology is how the heritable genetic information, the genotype, is transformed into the phenotype, the observable physical, chemical and biological characteristics of an organism. The driving forces of evolution, genetic variation - for example by mutation and recombination - and selection, work on two different layers: variation of the genotype can produce new and distinct phenotypes on which selective pressures can subsequently work. This means that both the genotype and the phenotype have to be integral parts of any model used for studying evolution.

This chapter describes MiniCellSim, a purely computational framework for deriving gene regulatory and metabolic networks from a sequence, which, while using complex genotype-phenotype maps, still remains sufficiently simple to allow for large scale evolutionary studies. The main aim was to build a deterministic hierarchical mapping, which decrypts a dynamical system representing the phenotype from a single string or nucleotide sequence - the genotype. This approach separates the genotype, upon which the genetic variation operators act, from the phenotype which is under selection pressure. The dynamical system itself is a minimal version of a **gen**e regulatory and met**abolic** network[11] represented by a system of ODEs. The source code of MiniCellSim is freely available under http://www.tbi.univie.ac.at/~luen/Diss/MiniCellSim/.

4.1.1 The Genotype-Phenotype Map and Fitness Landscapes

The relationship between a genotype and its physical expression, the phenotype, is a complex one, that involves a forbiddingly large range of processes, interactions and regulations. It is further complicated by environmental and epigenetic factors influencing the phenotype.

Still, as a correlation between genotypes and phenotypes is of fundamental im-

---

[11]Since gene regulation and metabolic control is intimately coupled in cellular dynamics the term *genabolic network* for the functional combination of genetics and metabolism is used.

portance for studying evolutionary processes, many different mathematical mappings, called genotype-phenotype maps, have been developed.

In nature the relation between genotypes and phenotypes is not a simple one to one mapping. One genotype can produce a whole range of phenotypes depending on environmental influences and chance, and, on the other hand, different variants of a gene can produce identical phenotypes. For theoretical considerations, though, it is favourable to assign each genotype a unique phenotype. Additionally, for studying evolutionary dynamics, the reproductive success, or fitness, of a specific genotype needs to be considered, to determine its success under selective environmental conditions. Again, in nature the fitness of an organism is a complex function of the phenotype and the environment. One useful approach to tackle this additional layer in theoretical systems, is to derive an explicit fitness function, which assigns each phenotype a discrete fitness value (see figure 39).
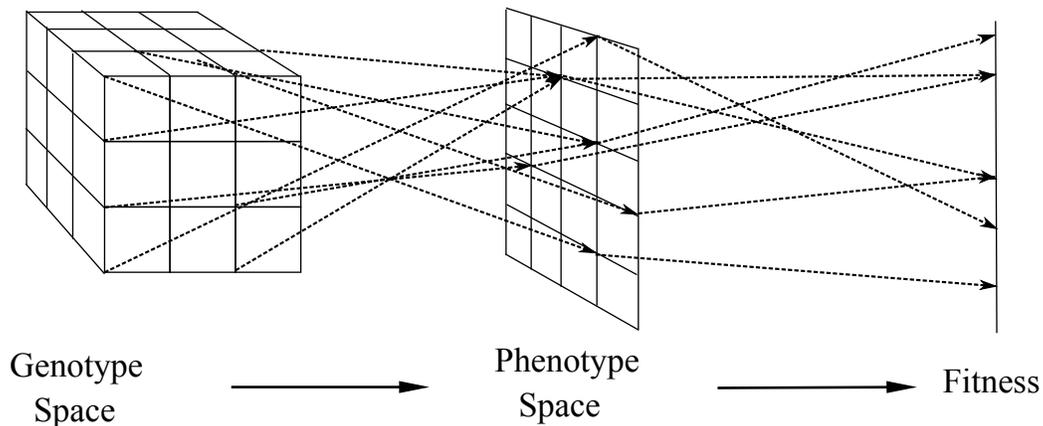
Figure 39: Sketch of a genotype-phenotype map followed by the assignment of fitness values to phenotypes. Both maps can be many-to-one mappings. (adapted from [395] with kind permission from the author)

The genotype-phenotype-fitness mapping facilitates the description of evolutionary processes on "fitness landscapes", a concept first introduced by Wright (1932) [442][12]. In essence it describes a projection from genotype space onto real numbers. Assuming increased fitness and reproductive success is symbolised by higher fitness values, selection pressures on a population in general lead to hill climbing in such a landscape, while variational processes, such as mutation, are shown as random movement in all directions.

---

[12]Wright's original concept of fitness landscapes was not based on the complete genotype space but on recombination space.

It has been proposed that genotype-fitness mappings have to be highly non linear for successful evolutionary optimisation, meaning that small changes in the genotype can lead to large changes in fitness [216]. As small changes of genotypes would only translate to slightly changed fitness values in linear mappings, evolutionary processes could easily get stuck at local optima. Additionally, highly non-linear mappings can also allow organisms to adapt to changing environmental conditions more readily.

Another property of genotype-phenotype mappings that influences evolvability, is the fraction of neutral mutations. Mutations are called neutral, if they do not influence the phenotype, or at least the fitness of an organism. A common example of neutral or nearly neutral mutations in protein coding sequences are synonymous substitutions in codons that do not alter the amino acid encoded by the nucleotide triplet. Other mutations are neutral, in that they change the phenotype, but without affecting the organisms fitness very much, for example by changing amino acid stretches which are non-essential for a protein's function [223]. According to the neutral theory sketched out by Kimura (1983) [222], neutral and deleterious mutations constitute the main part of genetic variation in organisms, while adaptive mutations are much rarer and contribute to a lesser extent to the emergence of new species.

Neutrality results in many-to-one maps with a high degree of redundancy, in which many genotypes give rise to the same phenotype. These neutral genotypes can be connected by point mutations - or are reachable via other single mutational events - and create neutral networks that span wide ranges and different parts of sequence space, thereby allowing populations to explore wide ranges of genome space by random drift without substantial selective disadvantages [202, 343]. This can increase the ability of a population to find new and beneficial genotypes by enabling it to have a wider range of possible genotypes with similar selective advantages and by reducing the risk of getting stuck at local optima [104, 216].

4.1.2 The Function of RNAs

For a long time ribonucleic acids, or RNAs, were largely regarded as mere messengers with a passive role in the information flow from DNA to proteins. This viewpoint has changed in the last three decades with the discovery of RNase P in *E. coli* [166] and the self-splicing introns in *Tetrahymena* [232], showing

that RNAs could also possess catalytic activities. The later discovery of RNA interference [295], and the miRNA pathways in animals [242] and plants [345], established, that RNAs played an active role in regulating gene expression.

Another layer of regulation by RNAs was found in mRNAs containing special segments called riboswitches, which can directly interact with small molecules to modulate gene expression [436]. As riboswitches do not require any additional factors, apart from the interacting RNAs and ligands, and as they are found in a wide range of taxonomic backgrounds with varying functions, it has been suggested they constitute one of the oldest gene regulatory mechanisms identified so far [430].

4.1.3   RNA structure

RNA is a linear hetero-polymer consisting mainly of the four nucleotides adenine, uracil, guanine and cytosine. The 5' and 3' carbons of the ribose units are linked to each other by phospho-diester bonds. DNA differs structurally from RNA mainly in the missing 2' OH group, rendering it chemically more stable and less prone to hydrolysation. Similar to DNA, RNA can form base pairs between cytosine and guanine, adenine and uracil and guanine and uracil, but due to the missing hydroxy group, the preferred form of RNA double-strands are A-type, and not B-type helices as are commonly found in DNA. Moreover, RNAs often act as monomers and can fold back on themselves, leading to complicated structures with stretches of double stranded stems and unpaired loops.

Similar to proteins, the different structures an RNA molecule can assume are mainly determined by its nucleotide sequence. Also, as with proteins and distinct from DNA, the final three dimensional structure of an RNA molecule is of major importance to its functional activity. Due to the specificity of base pairing, RNA possesses the ability to target specific nucleic acid sequences, such as the Shine-Dalgarno sequences in prokaryotic translation [379] and gRNAs involved in RNA editing in kinetoplastid mitochondria [7, 130].

Analogous to proteins the fold of an RNA molecule can be hierarchically classified into primary, secondary and tertiary structure (see figure 40) [60]. The primary structure of an RNA molecule simply consists of its base sequence and it suffices to distinguish chemically different RNA molecules. It is convention

to write it from the 5' to the 3' direction using the first letters of each nucleotide name, A, U, C and G. Between the different bases either Watson-Crick (G-C and A-U) or wobble (G-U) base pairs can form [89, 424]. A list of such base pairs, with the constraint that they do not cross, is called the secondary structure of the RNA. Crossing base-pairs, also known as pseudo-knots [329], are generally not subsumed under secondary structures. The tertiary structure is the three dimensional fold of the RNAmolecule, encompassing pseudo-knots and more complex interactions between bases and longer stretches of sequences. Finally, the quaternary structure of an RNA molecules involves the interactions with other molecules, such as in RNA-protein or multimeric RNA complexes.
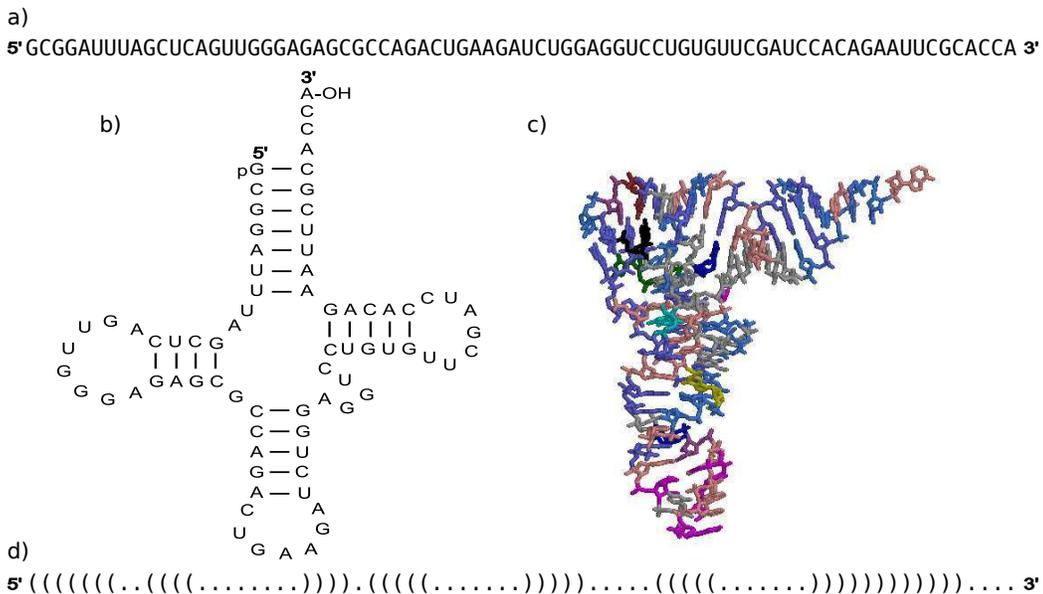
Figure 40: Primary (a), secondary (b and d) and tertiary (c) structure of the tRNA^{Phe} of *S. cerevisae*. (d) shows the secondary structure in dot-bracket notation, in which each dot represents an unbound nucleotide, each opening bracket the 3' and each closing bracket the 5' nucleotide in a base pair. Both the primary and the dot bracket notation are written with the 3' end on the left by convention. (Source: secondary structure: adapted from http://en.wikipedia.org/wiki/File:TRNA-Phe_yeast_en.svg under CC3 Attribution Share like 3.0 license, sequence and tertiary structure PDB ID: 1ehz)

There exist several practical reasons for this distinction between secondary and tertiary structures. First, the Watson-Crick and wobble base pairs which comprise the secondary structure of an RNA molecule, substantially contribute to its final free energy and therefore to the thermodynamical stability of the folded tertiary structure. While the free energy contribution of the individual base-pairs

themselves is rather small, adjacent base pairs are able to stack vertically and stabilize the structure via electrostatic and dispersion attraction [445]. Furthermore, the formation of the final tertiary structure seems to follow the formation of stable secondary structure elements in at least some cases, implying that there exists a kinetic hierarchy in structure formation that is reflected in the classification [60]. Finally, the interaction leading to tertiary structures, especially pseudo-knots and coaxial helix stacking are much harder to predict computationally and to track experimentally. The elements encompassed by secondary structure, on the other hand, can be predicted using efficient dynamic programming algorithms and explored using simpler experimental methods [190, 454]. In support of this view, the predicted RNA secondary structures have been found to be up to 70% faithful to experimentally determined structures on the level of individual bases [100, 271].

An RNA molecule can fold into an ensemble of secondary structures, each of them exhibiting a specific free energy change, ΔG_s, compared to the open chain. The most stable structure is the one with lowest free energy, also called the the minimal free energy (MFE) structure. As the energy differences between the various possible secondary structures are within the ranges of the thermal energies of the molecules at room temperature, RNAs under physiological conditions are not restricted to the optimal structures, but instead various secondary structures coexist with relative frequencies according to their folding free energies. The relative frequency of a sequence x existing in a specific secondary structure s is given by its Boltzmann weight divided by the partition function of the ensemble, Q, which represents the sum of the Boltzmann weights of all possible structures, S, the sequence can assume:

$$P(s|x) = \frac{e^{\frac{-\Delta G_s}{RT}}}{\sum_{t \in S} e^{\frac{-\Delta G_t}{RT}}} \tag{132}$$

For the standard nucleotide alphabet, A, U, C and G, and a sequence of length n, the number of different valid secondary structures S_n is considerably lower than the number of possible sequences, n^4, and according to Schuster $et\ al.$ (1994) [370] is given by:

$$S_n = 1.4848 \cdot n^{-3/2} \cdot 1.8488^n \tag{133}$$

133

This shows a high degree of redundancy, and therefore neutrality, between RNA sequences and secondary structures, or between sequence space and shape space. Computational studies showed furthermore that sequence space is highly connected by randomly distributed neutral networks [365]. Another important property of the RNA sequence-to-structure mapping is, that the most common structures can be attained by closely related sequences. By estimating the average radius of a sphere in sequence space around a random sequence containing the most common structures, it could be shown that this covering radius is much smaller than the overall radius of sequence space [365].

4.1.4   RNA as a Model for Evolution

A self-replicating RNA molecule is one of the simplest imaginable concepts for an evolvable system, as it comprises an inheritable genotype, its nucleotide sequence, and a phenotype, the RNA structures derived from it, in one and the same object. With the help of computational RNA secondary structure prediction this "RNA model" has been used extensively to elucidate evolutionary processes [367, 370].

Intensive studies of the RNA sequence-to-structure map during the last decade have revealed how the properties of this map influence the dynamics of evolutionary processes. However, whilst these studies have been very successful in elucidating the mechanisms governing molecular evolution [126, 366, 368], many concepts associated with biological genotype-phenotype maps, such as signal transduction or developmental processes, have no concrete analogue within the RNA model [125]. In particular, the absence of any form of control, or regulation in the RNA model is a major difference, since regulatory networks play an important role in the unfolding of genotypes to phenotypes in living organisms [27, 95].

Apart from purely computational descriptions, RNA has also successfully been used subject of experimental studies of evolution. Mills *et al.* (1967) [281] created a test tube system using the RNA of the self replicating Qβ phage RNA replicase, the replicase protein and nucleotides. By removing the need for the RNA to encode a functioning replicase, but selecting for fast replication, they were able to follow the emergence of increasingly efficiently replicated RNAs under the loss of replicase function. Later, a similar *in vitro* approach was used to create a more efficient version of a ribozyme [33]. Using serial dilutions and special selection for

a template directed RNA ligase function, Wright and Joyce (1997) [441] created a system to follow the evolution of both the catalytic and the amplification rate of a population of RNA molecules. Most recently, it has even been possible to create a system of cross-replicating RNA molecules, which allows the study of interacting populations of RNA molecules [251].

4.2 Model Description

As a basic requirement to be suitable for studying evolution, a model has to be self-contained in the sense that it does not require input of parameters on the fly. This has been achieved in the RNA model mentioned before by defining rules that provide the frame-work for the computation of the required parameters. The development of the model described in this chapter pursues the same strategy.

The decoding step is done in such a way, that all the relevant parameters needed to compute the time evolution of the dynamical system are calculated from within the model. This enables the individual system to freely explore genotypes by increasing its complexity without imposing limitations from the exterior.

In molecular terms the genotype is thought to be an RNA molecule, which is transcribed in pieces to yield other RNAs, which have regulatory and catalytic functions. It is based on the empirical evidence that natural occurring and engineered RNAs can fulfill a wide scope of different functions [400].

4.2.1 Related Work

The mathematical model presented in this chapter builds on related work in the area of artificial regulatory network (ARN) models. In 1993 Kauffman used random Boolean networks (RBN) to model gene regulatory networks [216]. RBNs show a broad range of dynamical behaviour from cyclic and multiple attractors to chaos. However, the majority of these interesting dynamical features disappear, if the updating rule for the temporal evolution of the network's state is changed from a synchronous to a biologically more realistic asynchronous one. Furthermore, RBNs display only a limited ability to structurally represent genes and genomes.

In 1999 Reil [344] enhanced the RNB approach by introducing the concept of an

artificial genome to overcome this structural weakness. The artificial genome is essentially a biologically inspired representation of genes and their interactions. The model allows manipulation of the topology of the gene regulatory network at the level of the genome, implemented as a string of digits, by a set of genetic variation operators which closely resemble their natural counterparts. This allows the study of questions regarding the evolution of the ARN and its quantised Boolean dynamics from the point of view of the changing genome.

Analogous to Reil, Delleart and Beer [97], Eggenberg [107] and later Bongard [52] embedded an ARN into a hand-coded morphogenetic system to evolve "multicellular" objects capable of performing some predefined tasks. Bongard demonstrated that within this framework, commonly referred to as artificial embryogeny (AE), gene reuse and modularity in terms of regulatory circuits can arise [391].

Banzhaf refrained from using the Boolean paradigm and expressed the dynamics of his ARN model as a system of ODEs [26]. Many dynamical phenomena of natural gene regulatory networks, such as point attractors, damped oscillations and heterochronic control can be reproduced by the ARN model[25]. By introducing an arbitrary "virtual" binding site for a desired output function, networks could be evolved where the activation pattern of the virtual binding site follows a predefined mathematical function [234]. The model described in this chapter differs from prior work with respect to the following points:

- First, the process of competition between molecular species that bind to regulatory regions of genes is modelled explicitly in mass-action-governed elementary reactions. The rational behind this decision lies in the fact that (i) competition for a common resource is one of the core reactions in gene regulatory networks, which requires an accurate mechanistic description and (ii) mechanistic details can have unexpected consequences in terms of dynamic phenomena, especially if coarse-grained approximations e.g. Michaelis-Menten type kinetics or concentration weighted mean values are used [179, 219]. Also the system can more readily interpreted in a stochastic framework, which is especially relevant for gene regulation by low numbers of transcription factors.

- Second, the genome and the gene products are modelled entirely in terms of RNA molecules. At the level of RNA secondary structure, efficient,

136

well-established algorithms exist to compute nearly any desired molecular property. In particular the statistical properties of the sequence to structure map, and its implications for evolutionary processes, have been profoundly enlightened. Therefore operations on RNA molecules, as used in the presented model, possess a certain degree of physical realism which is lacking if binary or real-valued vectors are used.

- Third, molecular interactions, another key feature of gene regulatory networks, are modelled within the framework of RNA secondary structures. This provides us with a physically meaningful, temperature dependent energy function, which is absent in Hamming-distance based approaches operating on bit strings.

- Finally, the current model is equipped with a minimal version of a metabolism and a simple membrane similar to the model described by Kennedy *et al.* (2001) [217].

4.2.2 The Cell

The basic unit of `MiniCellSim` is a putative protocell from the RNA world. It contains a genome, constituting the heritable information, ribozymes catalysing the metabolic reactions and RNA molecules functioning as transcription factors and regulating transcription.

Not every component of this cell is explicitly included in the model. The transcriptional apparatus and all components needed to reproduce the genome were omitted due to the increased complexity of including further classes of ribozymes. While this reduces the degrees of freedom for evolution of the system, assuming a functioning layer of basic processes allows us to concentrate on the regulatory aspects of transcription and increases the probability of obtaining cells that can have a functioning metabolism. Also, it is easier to alter this basic functions, without having to care about a definite underlying mechanism.

4.2.3 Genome, Genes and Gene Products

The genotype of the cell is represented by an RNA string of appropriate length. This genome can be either linear or circular, with both genes and regulatory

regions extending over the string's start and beginning, respectively, in the latter case. The genome is assumed to be single stranded and to be transcribed exclusively in one direction. Genes on the genome are allowed to overlap, but each gene has a strictly defined structure (see figure 41). A gene's coding sequence starts directly after a short sequence pattern which is reminiscent to the Pribnow box of prokaryotic and the TATA-box of eukaryotic promoter regions. Upstream of this sequence pattern up to two regulatory sites of fixed length are located. In the case of linear genomes, up- and downstream of each starting motif, enough space for the regulatory and the coding sequence is required for a gene to be functional.
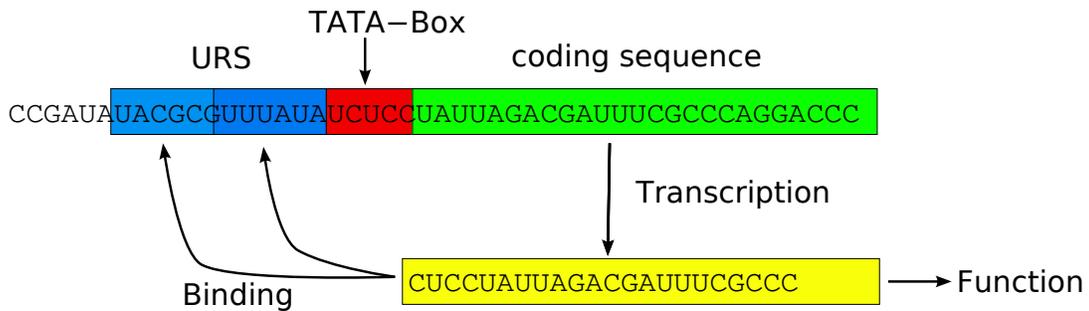
Figure 41: Gene control structure of a typical gene in the proposed model. Upstream of the coding sequence (green), lie the promotor region or TATA-box (red) and an upstream regulating sequence (URR, blue), consisting of two transcription factor binding sites. The function of the gene transcript is determined by folding into secondary structures representing the different classes of functional RNAs. (from [123])

Gene products fall into two major categories: (i) transcription factors and (ii) structural RNAs. While the former constitute the gene regulatory network, the latter fulfil catalytic tasks. The genome is central to the system, as all properties of the cell are derived from it. Mutation of the genome's sequence can alter the whole system.

4.2.4 Classification of the Gene Products

The classification of the gene products determines the decoding of the phenotype of the genome, and therefore is of great influence on the evolvability of the system. Due to the assumed nature of the gene products, a method based on RNA secondary structure prediction was chosen.

The function of a given gene is determined by means of an energy based compar-

ison with given target secondary structures. The gene's RNA sequence is folded into all target structures yielding a series of free energy values. In case of incompatibilities between the sequence and the structure an arbitrary penalty is given for each omitted base pair. The target structure with the lowest free energy for the gene's sequence determines the function. In case of two structures exhibiting the same free energy, the first target in the list is assumed to determine the function, to ensure a unique mapping.

For each functional class, at least one target structure is needed. The first gene classification round is between transcription factors and metabolic genes, the second determines what kind of regulatory effect the transcription factors have and which reactions the metabolic genes catalyse.

The list of target structures has to be provided beforehand, and all targets need to be of the same length as the classified coding sequences. The distribution of functions in the ensemble of gene products can be influenced by choosing target structures of different probabilities of occurrence and folding energy. Target structures with specific probabilities can be retrieved by sampling large amounts of random sequences and looking at the most common structures and folding free energies.

This secondary structure folding approach is supported by the fact that the function of naturally occurring non-coding RNAs is not only determined by their nucleotide sequence alone, but to a big part by their folded structure [400]. Furthermore, the high degree of neutrality and connectivity in RNA sequence-structure mappings are advantageous for evolvability of a system [104].

The statistical properties of the sequence-function mapping were probed in collaboration with Alexander Ullrich. To probe the genotype-function mappings, random neutral walks were performed by Ullrich (see details in Ullrich *et al* (2009) [419]) using different sequence lengths and classification algorithms. In such a walk each step equals a point mutation, that is neutral with respect to the phenotype. In case no new neutral mutation can be found, the walk is abandoned. At each step, the fraction of neutral point mutations and the number of reachable new phenotypes are recorded. The degree of neutrality is a measure for the robustness of the mapping, while the new phenotypes encountered indicate the connectivity of the neutral networks. The sum over all different phenotypes encountered on a neutral walk gives a measure for its potential for innovation,

or evolvability.
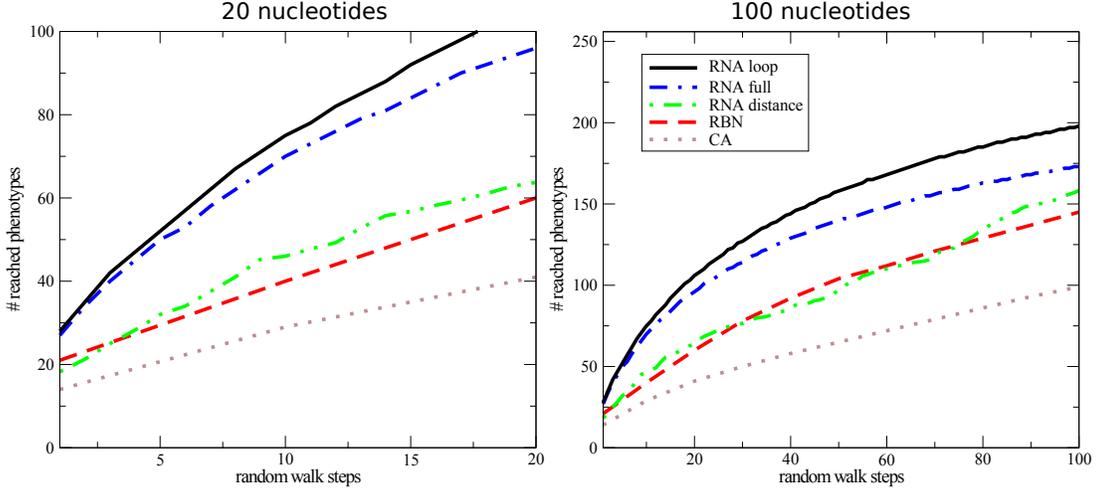
Figure 42: Number of different phenotypes reached on neutral random walks using different mappings for sequences of 20 (left) and 100 (right) nucleotides. For each curve 1000 random neutral walks of length 20 (left) and 100 (right) were performed. RNA loop: mapping using the sequence in the biggest loop of the MFE secondary structure; RNA full: target structure based mapping used in `MiniCellSim`; RNA distance: a target structure based mapping using ensemble distances; RBN: a mapping based on random Boolean networks; CA: cellular automaton mapping. (taken and adapted from [419] with kind permission of the author)

Five sequence-function mappings were compared, three of which were based on RNA folding, one on a random Boolean network (RBN), and one on a cellular automata (CA) (both described in Ebner *et al.* (2001) [104]). Of the three RNA based mappings two were based on comparison to target structures, and one on mapping the sequence of the longest loop of the MFE secondary structure [419]. Amongst the target based mappings, one is described above and used in `MiniCellSim` (RNA full), the other is based on the base pair distance of the whole ensemble of possible secondary structures to the target structure (RNA distance). For this the ensemble distance, $d_e(T, S)$, [165] was changed to the following:

$$d_e(T, S) = \sum_{(i<j | ij \in pt_T)} (1 - p_{ji})^2 + \sum_{(i<j | ij \notin pt_T)} p_{ji}^2 \tag{134}$$

with T: target structure and S = sequence, p_{ij} = bp probability of pair ij in sequence S, pt_T = pair-table of T.

As the ensemble distance based mapping is computationally more expensive than

the simple energy based one, but does exhibit lower connectivity and evolvability, the energy based one was chosen for `MiniCellSim`.

For each mapping 1000 random neutral walks with sequences of a 100 or 20 nucleotides, and 100 or 20 steps, respectively, were performed. The number of all possible phenotypes was restricted to $2^8 = 256$ for all algorithms. In terms of neutral point mutations all mappings performed similarly, with the RBN having the highest fraction (58%), followed by the RNA based mappings (50%), and the CA with 40%. Concerning the degree of innovation, two of the RNA based mappings, the loop and the energy based ones, outperform the others. For the longer sequences, the loop based mapping reaches on average 200 out of 256 phenotypes, the energy based one 175, followed by the ensemble distance (150), the RBN (145), and the CA (100) (fig. 42). The difference is even more pronounced for shorter sequences. This is also reflected in the connectivity of the different mappings. With the loop and energy based mappings a sequence on average has 27 and 26, respectively, unique phenotypes in its one mutation neighbourhood, significantly more than the RBN (21), ensemble distance (19) and CA (14) based ones.

4.2.5 Transcription Factor Binding

Transcription factors are further subdivided into two types, activators and repressors again using the classification method based on target structures described above. Since the structure of the gene regulatory network itself should be a target of evolution, a model based on molecular interactions is required that decides upon two questions: (i) Which transcription factor binds to the distinct URR of the gene, and (ii) to what extend is the URR bound by the different transcription factors.

Both repressors and activators are explicitly modelled to bind to one of the two binding sites in the upstream regulatory region (URR) of the gene (see figure 41), inhibiting or enhancing polymerase recruitment to the gene's promotor. Transcription factor binding to a given regulatory site is assumed to occur via RNA-RNA heteroduplex formation. The binding affinities, ΔG_{bij}, of a transcription factor, TF_i, to the binding sites of a gene, G_j, are calculated using the `RNAcofold` routine of the Vienna RNA package [37, 190]. To remove weak interactions a cutoff binding energy above which interactions are rejected is used.

141

Since computation of binding affinities is straightforward, the topology of the gene regulatory network can be readily derived from the genome sequence.

The free binding energies, ΔG_{bij}, computed this way can be used directly to calculate dissociation constants, K_{Dij}:

$$K_{Dij} = e^{-\Delta G_{bij}/\beta} \text{ with: } \beta = RT \tag{135}$$

Under the assumption that the association rate, k_{asss}, of the TF-URR complex, C_{ij}, is constant and limited by diffusion, a dissociation rate constant, k_{diss}, can be calculated, giving the following rates for complex formation, v_{ass} and dissociation, v_{diss}:

$$\text{TF}_i + \text{G}_j \underset{v_{diss}}{\overset{v_{ass}}{\rightleftharpoons}} \text{C}_{ij} \tag{136}$$
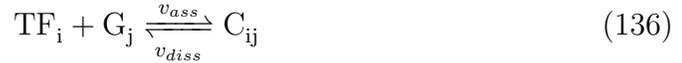
$$v_{ass} = k_{ass} \cdot [TF_i][G_j] \tag{137}$$

$$v_{diss} = K_{Dij} \cdot k_{ass} \cdot [C_{ij}]$$

The definition of the model parameters provides an opportunity to design more complex regulatory mechanisms, such as cooperativity in transcription factor binding. Cooperativity is introduced by allowing stabilising interactions between the transcription factors bound to adjacent sites in a gene's URR (see figure 43).

To derive the parameters for cooperative binding of two transcription factors, TF_A and TF_B, they are first bound the two URR binding sites individually using RNAcofold to derive their binding MFEs, ΔG_A and ΔG_B, and their cofolded structures. The sequences of the two transcription factors are then allowed to hybridise with each other using RNAduplex under the constraint that the base pairs of the MFE structures bound to the URR sites are preserved. A negative free energy, ΔG_{coop} is assumed to lead to an additional stabilisation of the TF-URR complexes and result in cooperative binding, again with a threshold to weed out weak interactions.

Table 9 shows exemplary results of this model of cooperative interactions on a sample of random sequences. For the URR sequences, the influence of AU and GC contents was separately studied. ΔG_{bind} is the sum of the binding energies,
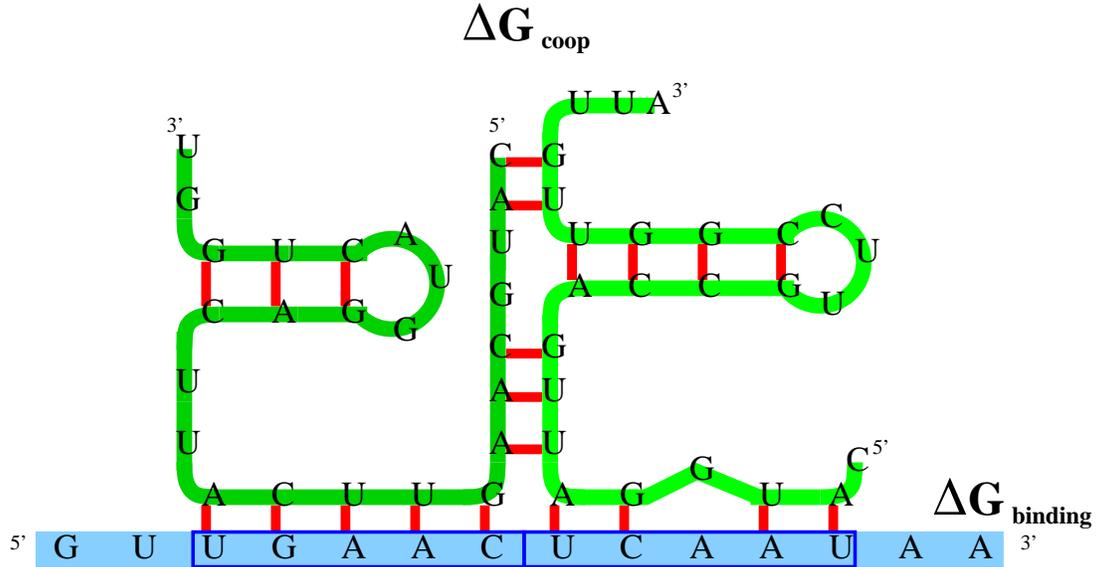
Figure 43: Cooperativity as modeled in the presented framework. If two transcription factors, independently bind the adjacent sites of an URR with an overall binding energy, ΔG_{bind}, they can form interactions with their unbound sequences, leading to an additional stabilising energy ΔG_{coop}. (from [123])

$\Delta G_A + \Delta G_B$, of the two transcription factors. As is to be expected, GC rich URR sequences are able to a much higher degree than AU rich sequences to form stable complexes. Pure AU URR sequences are characterized by slightly higher cooperativity as the G and C nucleotide in the transcription factor sequences are left free to interact with each other. In pure GC sequences binding to the URR and cooperativity energy are of about the same magnitude. The average differences between heterodimers and homodimers of transcription factors (self-cooperativity in table 9) are negligibly small.

For the modelling of the formation of a cooperative complex C_{AB} of the transcription factors TF_A and TF_B on the URR of gene G the following reactions are assumed:

$$\text{TF}_{\text{A}} + \text{G} \overset{K_A}{\rightleftharpoons} \text{C}_{\text{A}-} \qquad (138)$$

$$\text{TF}_{\text{B}} + \text{G} \overset{K_B}{\rightleftharpoons} \text{C}_{-\text{B}}$$

$$\text{TF}_{\text{B}} + \text{C}_{\text{A}-} \overset{K_{BA}}{\rightleftharpoons} \text{C}_{\text{AB}} \overset{K_{AB}}{\rightleftharpoons} \text{TF}_{\text{A}} + \text{C}_{-\text{B}}$$

$$\text{with: } K_A = e^{-\Delta G_A/\beta}; \ K_B = e^{-\Delta G_B/\beta}; \ K_{coop} = e^{-\Delta G_{coop}/\beta}$$

$$K_{BA} = K_B \cdot K_{coop}; \ K_{AB} = K_A \cdot K_{coop}$$
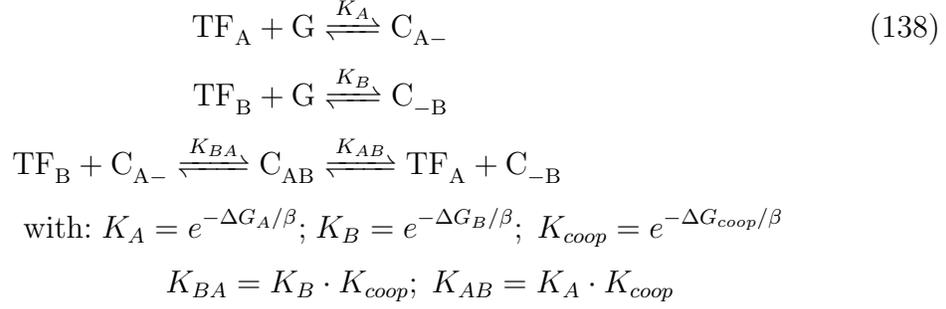
These binding reactions are divided into single steps using the same assumptions as above (see eq. (137)).

4.2.6 Transcriptional Regulation

Transcription of genes is modelled after a three state *regulated recruitment mechanism* observed with some bacterial genes (see [337], pp.13-42 and [336]). The transcriptional activity of a gene is assumed to depend on the state of its URR. Free URRs recruit RNA polymerase at a low, basal rate, URRs bound exclusively to activating transcription factors at a high rate. If at least on least one site of a gene's URR is bound by a repressor, it is considered silenced.

A fixed amount of RNA polymerase molecules, pol, is assumed to be present in each cell, leading to competition between the different promoters. Only free, G, and activator bound, C_{A}, URRs recruit polymerase. Once the polymerase is bound, transcription of the gene starts with the same rate under consumption of activated building blocks, X_{A}. The following reactions govern RNA transcription from each gene:

$$\text{G} + \text{pol} \overset{v_{free}}{\longrightarrow} \text{polG} + n\text{X}_{\text{A}} \overset{k_{TS} \ polG}{\longrightarrow} \text{G} + \text{pol} + \text{RNA} \qquad (139)$$

$$\text{C}_{\text{A}} + \text{pol} \overset{v_{act}}{\longrightarrow} \text{polC}_{\text{A}} + n\text{X}_{\text{A}} \overset{k_{TS} \ polC_A}{\longrightarrow} \text{G} + \text{pol} + \text{RNA}$$

$$\text{with: } v_{free} = k_{BP} \cdot [pol][G]; v_{act} = act \cdot k_{BP} \cdot [pol][C_A];$$
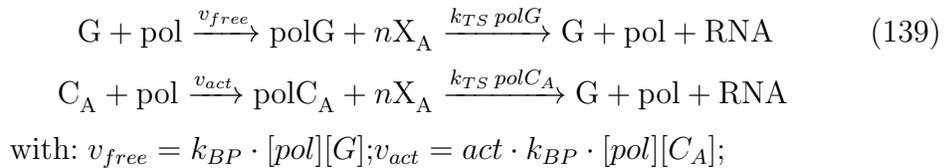
Table 9: Binding energies of random RNA pairs of length 25 to two URR RNAs of length 5. Different base compositions in the URR sequences from pure AU to pure GC were studied. The free energies for cooperative interaction (ΔG_{coop}) are calculated for the conformations of lowest free binding energies (ΔG_{bind}) as indicated in figure 43 and given with the standard deviation. For each URR sequence enough sequence pairs to give 8000 stable hetero-cooperativity complexes where sampled.

Sequence	Sample	Sample size:	Hetero-cooperativity			Self-cooperativity		
% AU	# URR	Sequence pairs	Stable (%)	$-\Delta G_{\text{coop}}$ [kcal·mol^{-1}]	$-\Delta G_{\text{bind}}$ [kcal·mol^{-1}]	Stable (%)	$-\Delta G_{\text{coop}}$ [kcal·mol^{-1}]	$-\Delta G_{\text{bind}}$ [kcal·mol^{-1}]
100	11	1000000	0.1 ± 0.1	3.13 ± 2.26	0.32 ± 0.22	0.2 ± 0.3	4.00 ± 2.77	0.32 ± 0.21
90	56	824301	1.0 ± 1.2	3.02 ± 2.21	0.58 ± 0.55	0.9 ± 1.1	3.71 ± 2.65	0.72 ± 0.68
80	223	364473	8.8 ± 9.2	2.90 ± 2.15	0.87 ± 0.76	7.5 ± 8.5	3.53 ± 2.57	0.92 ± 0.78
70	557	163760	18.4 ± 12.7	2.85 ± 2.12	1.14 ± 0.98	14.2 ± 10.1	3.35 ± 2.46	1.17 ± 0.99
60	1021	62027	31.6 ± 14.3	2.80 ± 2.10	1.45 ± 1.17	23.9 ± 11.7	3.24 ± 2.39	1.47 ± 1.18
50	1265	28598	42.7 ± 13.3	2.78 ± 2.08	1.74 ± 1.35	32.2 ± 11.8	3.20 ± 2.35	1.74 ± 1.32
40	1007	16860	53.8 ± 11.0	2.75 ± 2.07	2.09 ± 1.50	41.6 ± 10.9	3.17 ± 2.33	2.03 ± 1.45
30	574	14305	60.9 ± 9.6	2.75 ± 2.07	2.43 ± 1.66	48.4 ± 10.2	3.21 ± 2.34	2.33 ± 1.59
20	238	13105	65.7 ± 8.6	2.75 ± 2.06	2.69 ± 1.76	54.0 ± 9.6	3.27 ± 2.37	2.55 ± 1.67
10	35	12054	70.9 ± 7.0	2.76 ± 2.07	3.07 ± 1.91	58.9 ± 9.1	3.31 ± 2.39	2.85 ± 1.80
0	8	11860	73.3 ± 8.4	2.85 ± 2.11	3.51 ± 2.07	66.0 ± 4.1	3.67 ± 2.63	3.29 ± 1.98

In eq. (139), n is the length of the gene, X_A are the activated nucleotides, and, k_{TS}, is the transcription rate. The transcription rate, k_{TS}, itself also depends on the concentration of active nucleotides. The analytic expression for the dependence was adopted from the mechanism of RNA replication by the replicase of the phage $Q\beta$[44]:

$$k_{TS} = \frac{[X_A]^2}{cTS_0 \cdot [X_A]^2 + cTS_1 \cdot [X_A] + cTS_2} \tag{140}$$

With the parameters cTS_0, cTS_1, and cTS_2 estimated from Biebricher *et al.* (1983) [44] and Arnold (2003) [14].

4.2.7 The Metabolism

The central reactions of the cellular metabolism are catalysed by the gene products classified as structural RNAs, SRs. Structural RNAs are thought of as ribozymes, and are either capable of catalysing either a chemical reaction that activates a mediator molecule E_I, or a reaction that transforms membrane precursor molecules M_I into membrane building molecules M_A (see figure 44).

The active mediator, E_A, in turn, transfers energy to monomeric building blocks, X_I, converting them into active nucleotides, which can be directly used for transcription. Eventually we end up with RNAs, which may influence their own production directly or in-directly by either transcription regulation or creating more activated mediator, very much in the manner of an auto-catalytic cycle.

Catalysts are assumed to require specific, predefined, structural elements for their activities. To this end a target secondary structure, S_j, is associated with each reaction, j, and the sequence of the structural gene is folded into this structure S_j. The catalytic efficiency, eff, of a ribozyme in the catalysis of a metabolic reaction is derived from the activating energy $E_a = \varepsilon_j - \varepsilon_0$ required for the transition from the MFE structure with an energy ε_0 into the (lowest) suboptimal state, that folds into the structural element, S_j, and constitutes the active form of the ribozyme (figure 45). eff is derived as:

$$eff = e^{-E_A/\beta} = e^{(\varepsilon_0 - \varepsilon_j)/\beta} \tag{141}$$

146

Figure 44: A sketch of the basic reaction network of the presented model. The RNA polymerase is assumed to be available in a fixed amount and recruited to the genes (green) promoter region (red) at a rate determined by the TFs bound to the upstream regulating sequence (blue). The RNA transcription rate depends on the concentration of activated RNA building blocks (X_A) and consumes nX_A per RNA. The RNAs decay to inactivated components (X_I), which are reactivated via consumption of activated energy rich metabolites (E_A). The gene products are categorized into structural (SR) and gene regulatory (TF) RNAs. The SRs catalyze the activation of metabolites (E_I) and the incorporation of membrane building blocks (M_I^{int}) into the membrane (M). The internal pool of M_I^{int} is coupled to the exterior pool (M_I^{ext}) via diffusion through the membrane. All parameters for transcription factor binding to regulatory regions and the catalytic efficiencies of structural proteins are obtained by a mapping process (see text for details) and are therefore targets of evolution. (from [123])

This concept for the evaluation of catalytic efficiency can also be interpreted as a distance measure between structures on an energy scale. It is preferred here over simpler distance measures, such as the Hamming distance based string comparison methods, because it retains the useful and realistic statistical properties of the RNA sequence to secondary structure map as exemplified by the evolutionary fitness landscape [126, 127].

Moreover, the activation concept allows the optimization of the catalytic efficiencies of ribozymes through accumulation of mutations that reduce E_a by stabilizing the suboptimal structure S_j relative to the MFE conformation S_0. Ideally, if a structural gene's MFE structure S_0 is identical to the target structure S_n, $e^{-E_a/\beta} = 1$, and the ribozyme catalyses reactions at maximum velocity.
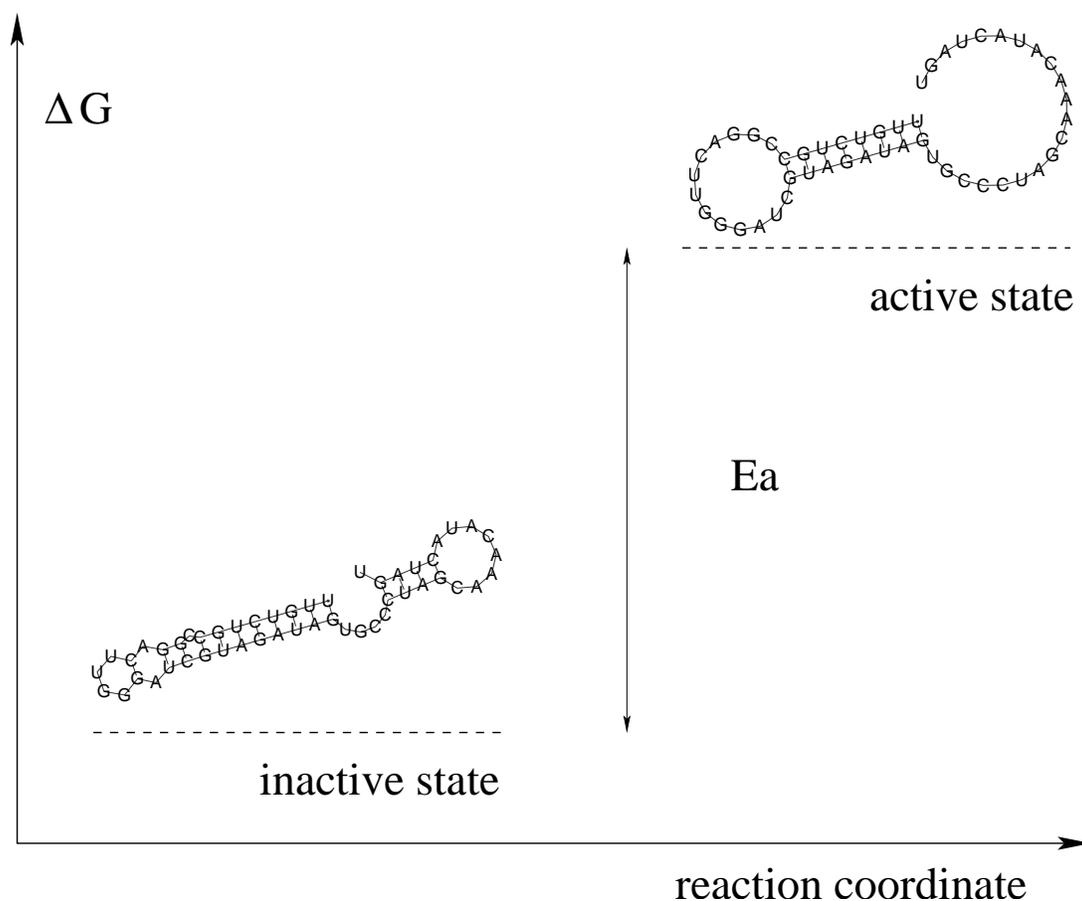
Figure 45: Activation of the ribozyme. The active structure (right) catalyses the metabolic reactions. The free energy needed to form the the secondary structure of the active ribozyme from the minimal free energy secondary structure of the RNA (left) is used to derive the catalytic efficiency of the ribozyme. (from [123])

For reactions catalysed by a ribozyme, SR, simple irreversible Michaelis-Menten kinetics are assumed, in the case of the metabolite, E for example:

$$\mathrm{E_I} \xrightarrow[v_R]{SR} + \mathrm{E_A} \tag{142}$$

$$v_R = eff{\cdot}k_{EA}\frac{[\mathrm{E_I}] \cdot [\mathrm{SR}]}{Km_{EI} + [\mathrm{E_I}]}$$
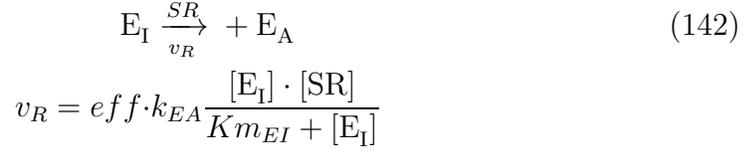
In this k_{EA} is the rate of the ideal ribozyme, and Km_{EI} is the Michaelis constant of the inactive metabolite.

RNA degradation to inactive nucleotides X_I is assumed to be of first order, as is the inactivation of E_A and X_A (see figure 44).

For the cell membrane building blocks, a constant exterior concentration is assumed, and the flow through the membrane is modelled proportional to the concentration gradient times the cell surface, A:

$$\mathrm{M_I^{ext}} \xrightleftharpoons{v_{trans}} \mathrm{M_I^{int}} \tag{143}$$

$$v_{trans} = k_{ex}{\cdot}A \cdot ([\mathrm{M_I^{ext}}] - [\mathrm{M_I^{int}}])$$

The cell surface is determined by the size of the cell membrane M, that grows with introduction of M_I facilitated by ribozymes and decays in a first order fashion. The cell volume, V, is derived via the cell surface and assuming a spherical cell with radius R:

$$R = \sqrt{\frac{A}{4\pi}} V = A \cdot R/3 \tag{144}$$

4.2.8   Creation and Evaluation of the Dynamic System

After the determination of the parameters by the respective mapping, the genabolic network is translated into a set of ODEs. In order to describe the system in a general and easily accessible format, it is implemented in SBML [200]. A full list of reaction rates and parameters is given in appendix D on page 168.

The derived dynamical system can then be evaluated, for example by time course

integration. From the integrated time courses, fitness values can be deduced to drive an evolutionary optimization procedure (see Figure 46). The integrator front end currently used is the SBML-ODE Solver, a versatile integrator for continuous ODE systems [258].

Due to the use of SBML a variety of integrators and analysis software can easily be adopted and a flexible handling is facilitated.
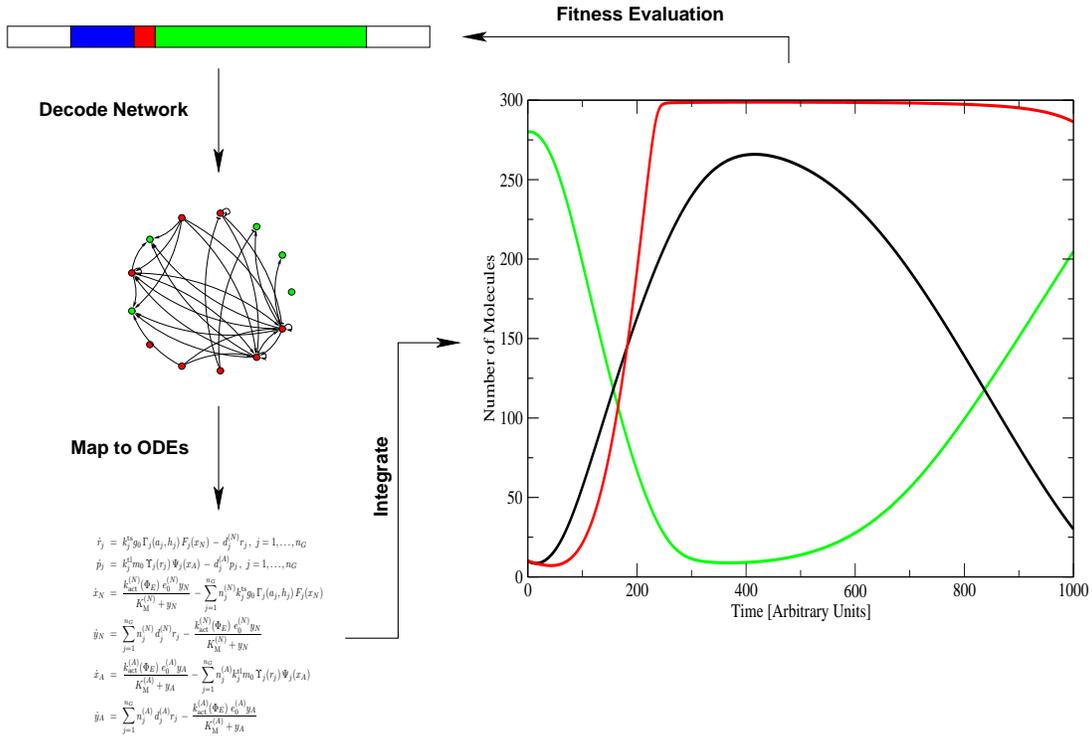
Figure 46: Schematic representation of an evolutionary cycle. The topology of the genabolic network together with the reaction parameters is *decoded* from the genotype. This information is translated into an ODE system which after numeric integration provides the concentration time coarse of the individual chemical species (the phenotype) which in turn modulates via a fitness function the reproductive efficiency of the genome. (from [123])

4.3   Results

As a proof of concept, the following experiment was designed and carried out. The objective was to find out whether a cell with capable of adapting its cell volume to a predefined target volume could be found using a mutational adaptive walk. Figure 47 shows the dynamical behaviour of the final cell of the adaptive

walk. The balance between regulatory and metabolic dynamics indeed adjusts the cell's volume exactly to the target volume.
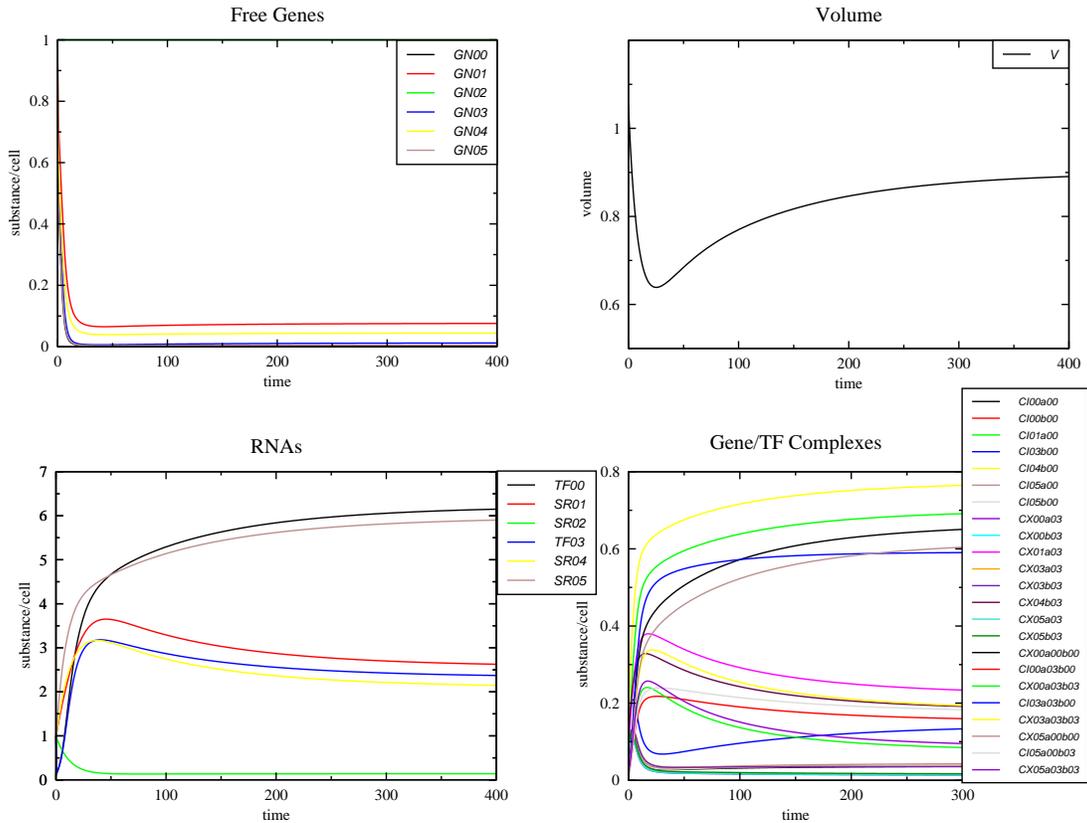
Figure 47: Integrated time course of a cell evolved via an adaptive walk targeting a cell volume of 1. A genome of length 100 and a gene length of 20 nucleotides were chosen. URR length was 5 per site and the promoter sequence motif was CC. Genes, transcription factors and ribozymes are labeled GN, TF and SR. The TF/gene complexes are labeled either CX or CI for activating or inhibiting complexes. The first index denotes the gene, letters a and b the sites the transcription factors bind to. (from [123])

4.4 Discussion

The model described in this chapter can be directly used for evolutionary studies. The encryption of all relevant system information within a string genome allows the description and evolution of genabolic networks in an entirely independent fashion. No external sources of additive information are necessary, the system is self-determined and closed as far as rules and system-sustaining model func-

tions are concerned. In contrast to prior exclusively RNA based auto-catalytic systems, the genotype and the phenotype in the presented model constitute separate objects. This allows an unhindered evolvability of the minimal cell on the way from a random dynamical network to an adapted functional system. There is also a substantial fraction of neutral mutations, which was found to be an important factor for efficient evolutionary optimization [127, 202].

Due to the regulation mechanisms implemented, in particular the direct interaction of the transcription factors in a cooperative manner, the model allows for studies of the evolution of a great variety of regulatory networks. Experiments regarding an optimization of certain qualities or functions, for example high adaptability, high robustness, insensitivity to environmental stress, are conceivable.

In order to study evolutionary phenomena, genome replication could be considered too. One natural selection criterion would be a short time of self-reproduction for the individual system. As the model includes an explicitly modelled membrane as described, for example, in the chemoton systems [136], the growing cell membrane could be taken as an indicator for cell growth and replication. At a certain cell volume and concentration of the cell components the individual cell could be deemed ready for cell division, and the time between divisions be taken as a direct fitness measure. The evolutionary evaluation of different genabolic networks could also be based on a fitness value that results from a combination of replication rate and metabolic efficiency.

4.4.1 Limitations of the Model

One minor limitation of the system is the slow speed of evaluation and computation. While many of the time-critical functions for RNA folding as well as the solver for the dynamical system are fast, the main package is programmed in Perl and is relatively slow. A simple adaptive walk over a 1000 steps with a short genome of 100 nucleotides took about an hour of computation time, but with longer genomes and more genes the cost in computation time increases significantly. Studies on populations of hundreds of individuals would be very time consuming with the current state of the program.

Another more fundamental limitation is the closed concept of the metabolism

and the dependence of the mapping on fixed target structures. While it is possible to create more than one energy-rich metabolite, all potential reactions in a metabolism would have to be encoded in advance and target structures assigned for each reaction, restricting the directions the model can develop considerably.

A much more general model, solving the problem of a rigid and limited metabolism with an open sequence-function mapping, has been developed in our group by Ullrich *et al.* (2008) [418, 419]. Similar to `MiniCellSim`, Ullrich *et al.* use an RNA genome with TATA-box like promoter sequences, and a sequence-function mapping based on RNA secondary structure prediction. However, different from `MiniCellSim`, their mapping does not depend on comparison to prior defined target structures, but uses the sequence of the longest loop in the MFE secondary structure, to generically assign each sequence to a reaction. The pool of all potential reactions is calculated by the graph based artificial chemistry package `ToyChem` [36], which also provides reaction energies and kinetic constants. This way sequences can be mapped onto a much bigger reaction network and create a more realistic cellular metabolism. The chemical reaction network, on which the ribozymes can be mapped, can encompass tens of thousands of metabolites. Due to the energetic parameters provided by the artificial chemistry, the directionality of reactions can be evaluated. For quantifying the fitness of a network, metabolic flux analysis is performed, and the fitness is assumed as the maximum yield over all extreme pathways [124].

The loop based mapping also exhibits a slightly higher innovation rate than the energy based one used in `MiniCellSim` and a high degree of neutrality (see fig. 42 and [419]). Simulations with this framework have shown, that it can produce metabolic networks encompassing more than 500 metabolites with realistic characteristics [418, 420].

4.4.2 Possible Extensions

Several extensions to the `MiniCellSim` model can be envisioned that would increase the framework's utility. The lack of a realistic metabolism could be remedied using the approach of Ullrich *et al.* (2008) [419]. As both models build on a similar RNA genome, both mapping methods can be integrated into one, using the energy based one for a first round of classification, and the loop based one for a second round mapping ribozyme activities to the reaction network. This

way a combination of gene regulatory with realistic metabolic networks could be achieved.

Another missing feature is the direct influence of metabolites and small molecules on gene regulation, which would allow the model to react on changes in the environment more readily, and develop a simple form of signal processing. One idea to include this, based on some riboswitches [436], is to assign a short RNA sequence to each metabolite and see whether the binding of this sequence to transcription factors can make them switch from one classification, say activators, to the other inhibitor. Alternatively, binding of a small RNA sequence could enhance, or inhibit their binding to URR of genes.

A simpler enhancement planned is the inclusion of membrane bound importers. Similar to ribozymes, the transportation rate constants and substrate specificity could be determined by folding into target structures, and transport against a gradient be coupled to consumption of energy-rich metabolites.

Feed-back loops resulting in hysteretic gene switches are one kind of epigenetic phenomena that could occur in `MiniCellSim`. However, `MiniCellSim` could also be altered to include other epigenetic effects, for example by allowing nucleotide modifications akin to DNA methylation as found in eukaryotes [204]. For this additional groups of ribozymes could alter the methylation state of genes' promoters, and thereby the genes' expression rates, in a sequence specific way by either using sequence complementarity or hybridisation.

5 Conclusion

5.1 Thesis Incentive

The aim of this thesis was to use mathematical abstractions of biological regulatory systems to gain a better understanding of their behaviours. Gene regulatory systems are especially suited for this kind of analysis as they are well studied mechanistically, easy to manipulate, and experimentally accessible. Further, the emerging field of synthetic biology shows the direct applicability of mathematical prediction in the design of gene regulatory networks with specific functions [178, 390].

As there was little quantitative data available on the systems studied, and no possibility to measure parameters or verify predictions *in vivo*, quite general abstractions were used. The only exception to this was the model of the GATA network underlying yeast NCR (section 3.3 page 103) where expression data was available for model validation.

For most of the analyses deterministic, continuous models were used, complemented by stochastic simulations for the repressilator-like systems. This combination seems especially promising as the deterministic approach allows the derivation of general behaviours over large ranges of parameter values, while the stochastic simulations can be used to check these behaviours under more realistic assumptions of low molecule numbers and random fluctuations.

5.2 Discussion

The comprehensive analysis of the behaviours of two cyclic repression systems extends previously published work in a few respects. While the global stability of the classical repressilator, *RepLeaky* (section 2.2.1 page 50), has been studied before [24, 108], this work extends the analysis to describe systems of arbitrary numbers of genes and strong repressor binding, as well as giving more detailed dependency of the bifurcations on the different parameters. This allows the identification of key parameters determining certain behaviours, and the ranges in which these behaviours are displayed. Furthermore, a novel system, which is a combination of cyclical repression and auto-activation, *RepAuto* (section 2.2.2,

page 54), is analysed in detail.

Both systems show the either oscillatory behaviour or multistability for odd or even numbers of genes, respectively. Distinct from the classical repressilator, the combination of cyclical repression and auto-activation allows oscillations to appear in the absence of cooperative binding. This means that transcription factors binding as monomers could also be used to create oscillators.

The repressilator with auto-activation can also possess stable heteroclinic cycles as attractors, leading to oscillations with increasing periods. Such dynamics are known as May-Leonard behaviour [191], and have been described in models of competing populations [274, 326, 371]. For discrete systems this can lead to a sort of molecular roulette wheel, as the system could oscillate a few times and then get stuck in one of the corner equilibria [326].

The stochastic modelling showed that some parameter combinations predicted to give oscillations in the deterministic system could prove unstable in real biological systems due to fluctuations and low molecule numbers. Together, stability analysis of the deterministic system and stochastic simulation give a good picture of the main requirements to implement a stable oscillator *in vivo*.

Another point illustrated by the analytical bifurcation analysis of the repressilator and the single auto-activator (section 3.4 page 111) is the importance of rescaling of variables and parameters. Without proper rescaling to reduce the complexity of the problem, it would have been much harder to achieve meaningful results. The independent parameters could be reduced from 7 to 4 in the case of *RepAuto* and from 17 to 3 for the single auto-activator, reducing the dimension of the parameter space to be examined.

The sequence-based approach for deriving potential GATA-type transcription factor regulatory networks in fungi (sect. 3.2 page 98) cannot be directly validated. The correct identification of GATA-type transcription factors and regulatory interactions between them found in *S. cerevisiae* and in some *Aspergillae* [84, 439] supports the methodology. All predicted topologies can be imagined to be derived from a *ur*-GATA-type transcription factor via gene duplication events, followed by loss of trans-activatory domains in protein and `HGATAR` motifs in the promoter regions. It is especially striking that all potentially inhibitory GATA-type transcription factors identified possess paired GATA binding motifs

upstream of their genes. This indicates that they could be controlled by other GATA factors, and maybe even be auto-regulating.

Furthermore, all fungi possessing more than one gene encoding GATA-type transcription factors classified as activating also possess at least one potentially repressing GATA-type transcription factor. The same holds true for all save two fungi possessing potentially auto-activating GATA factors. One reason for this could be the suppression of gene dosage effects, another the transformation of a sigmoid into a gradual response.

The simple model of the central yeast NCR, created with approximate parameter values, validated astonishingly well against experimental data (sect. 3.3 page 103). One possible explanation for this could be the robustness of the underlying network against parameter variation. The negative feedback exerted by GFZ3 and Dal80 has been argued to be involved in rapidly attaining a steady-state, and fine tuning of the nitrogen starvation response [84, 92]. Simulation results of the model suggest, that the negative feedback could be important for achieving a gradual, instead of a highly nonlinear or sigmoid activation in dependent on the nitrogen source.

While the model of NCR regulation as a whole did not exhibit bistable switching dependent on nitrogen availability, the single auto-activator part exhibited bistability over physiologically meaningful ranges of nitrogen concentrations (sect. 3.4 page 111). To keep the model as generic as possible, three signals were incorporated in this simplified GATA-type auto-activator - two acting at the post-transcriptional level via mRNA stability and protein retention in the cytoplasm, and one directly affecting the rate of basal transcription. This model showed reversible, bistable switching over ranges of all three signals, showing that each of the mechanisms alone suffices to create a bistable system. While multistable GATA-type transcription factor expression has not been described in the yeast NCR - and potentially does not exist there because of the negative feedback - bistability has been suggested to underlie T_H2 cell differentiation and could also play a role in endomesoderm development in nematodes [192, 261].

Further, for the parameter ranges chosen, the simple auto-activator also exhibits a high sensitivity to the copy number of genes. In the model a simple gene duplication event would lead to irreversible activation for the signal influencing transcription directly, while, for the post-transcriptional mechanisms, deactiva-

tion shifts to nearly ten times higher signal strengths. This strong gene dosage effect could be an important factor in the divergence of duplicated genes, as the change of both the switching behaviour and the ranges over which it occurs could influence an organism's fitness dramatically.

A few changes in both the non-coding, as well as the coding regions of the duplicated genes, can compensate for this change, and revert switching back to before the duplication. Some of these changes lead to motifs commonly found in gene regulation, such as feedforward loops and cascades [243, 261, 377]. Also some of the topologies for GATA-type gene regulation in fungi predicted in this work could have developed this way. GATA-type transcription factors should be especially suited for such evolution, as they can posses similar DNA sequence specificity, potentially leading to cross-regulation and competitive inhibition, which can be varied by point mutations in the coding region [84, 142, 341].

Another interesting mutation in GATA-type transcription factors is the loss of the trans-activatory region, transforming an activating transcription factor into a repressor. This could have occurred in the evolution of Dal80 and GZF3 from a GAT1-like common ancestor [256]. For the loss of trans-activation in one of the paralogous genes, the model predicts a loss of bistability leading to a gradual in- or decrease of activation in dependence on the three different signals.

Intriguingly, the combination of activation and inhibition can also lead to occurrence of stable oscillations. Inverse eigenvalue analysis revealed that as few as four parameter changes suffice to achieve a system that switches from expression at a single level to stable oscillations dependent on all three signals. While such oscillatory expression has not been described for GATA-type transcription factors up to now, it has been found to be the case for other transcription factors, such as Hes1 involved in somite segmentation [186], the tumour suppressor p53 [28], the mammalian transcriptional activator NF-κB [193], and the Period and Timeless genes in circadian rhythms[244]. Under certain conditions, yeast exhibits respiratory oscillations with periods in the range of hours [293]. As these oscillations also encompass metabolites central to the cellular nitrogen cycle, a connection between them and expression GATA factors involved in NCR is imaginable.

With `MiniCellSim` (see sect. 4 page 128) a framework of a self-contained *in silico* cell is presented, that could be used for studies of the evolution of gene regulatory networks. All information necessary to describe the cell's processes can be

derived from the genome, on which mutation operations can be performed. As the principal mappings are based on RNA folding and binding thermodynamics, they possess some characteristics favourable for evolutionary studies, such as a high fraction of neutral mutations, and a high connectivity between neutral networks. The framework should also be easy to extend, especially to accommodate a more complex metabolism, such as the one used by Ullrich *et al.* (2008) [418].

5.3 Perspective

While oscillating systems of the repressilator type have been analysed exhaustively in a variety of ways, there still exist numerous interesting challenges. One obvious challenge would be an implementation of the *RepAuto* system in a living cell and a subsequent comparison to various other synthetic oscillators. Further, different kinds of cell-cell coupling, for example via a quorum sensing system, such as the LuxR/I system widely used in synthetic biology, and suggested for the classical repressilator [137, 185], could be studied both computationally and *in vivo*.

Another interesting aspect of repressilator like systems is the influence of additional feedback loops and coupling mechanisms which are found in biological systems, such as the recently suggested three element repressilator in the *Arabidopsis thaliana* circadian clock [330]. With the abundance of sequence data available it might be possible to identify more repressilator-like systems in living organisms, and maybe study their evolution directly. An analysis of the compensation of stochastic effects on period robustness via network topology would be especially intriguing, as it also should have a strong impact on the evolution of such oscillators.

The predicted GATA-type transcription factor network topologies are speculative, but all predictions could be quite readily examined experimentally. One possibility for testing would be gene knockouts and over-expression experiments, as well as alterations of GATA-factor binding sequences in the genes' upstream sequences. Another possibility would be the reconstruction of some network topologies in related organisms to test them independently. The strong gene dosage effects predicted would be a particularly interesting experimental find. This could be studied using a reconstructed GATA-type auto-activator in a yeast

strain devoid of GAT1, GFZ3, DAL80, and GZF3, and by looking at reporter gene transcription. Maybe even a synthetic gene-regulatory oscillator could be constructed out of GATA factors this way.

The model of the central NCR in yeast requires some expansion. The mechanisms for Gln3p and Gat1p sequestration could be differentiated and refined in the light of recent publications [141, 142]. Furthermore, the model needs additional quantitative data to allow estimation of the subtle differences in parameters describing binding and expression of the individual GATA factors. One straightforward prediction of the model is the effect of the inhibitor DAL80, which could be examined by growing a DAL80 knockout strain on different nitrogen sources and measuring GAT1 expression.

For `MiniCellSim` numerous extensions are imaginable, as mentioned above. Apart from a more complex metabolism, different ways of interactions between the metabolic state and gene regulation should be easy to implement. Other refinements could be more types of trans-membrane transporters, and modelling of genome duplication of cell division in dependence of the cellular state.

The derivation of gene regulatory networks on its own can be used to identify network topologies capable of performing tasks such as bistable switching, oscillations, and simple arithmetical and Boolean operations using appropriate fitness functions [11, 313]. As all reactions governing gene regulation are implemented as mass action kinetics, the derived models can also be adapted for use in a stochastic framework for better exploration of robustness to random fluctuations.

A   Characteristic Equation of *RepLeaky*

The Jacobian matrix, $\mathbf{J}(x)$, of system (70), at an equilibrium $x = y = \alpha f(x)$ (71), appears as follows (79):

$$\mathbf{J}(x) = \frac{\partial(\dot{x}, \dot{y})}{\partial(x, y)} = \begin{pmatrix} -\beta\, M(x)^{-1} & \beta\, M(x)^{-1} \\ A(x) & -I \end{pmatrix}$$

where:

$$\mathbf{A}(x) = \alpha\, \frac{\partial f(x_{i-1})}{\partial x_j}$$

For a 2×2 block matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ with commuting blocks, $AB = BA$, the determinant can be derived as follows [382]:

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(AD - BC)$$

As $\mathbf{J}(x)$ is a matrix with commuting blocks, the eigenvalues λ_j at an equilibrium can be derived as follows:

$$\begin{aligned}
0 &= |\mathbf{J}(x) - \lambda\, I| \\
&= \begin{vmatrix} -\beta\, M(x)^{-1} - \lambda\, I & \beta\, M(x)^{-1} \\ S(x) & -(1 + \lambda)\, I \end{vmatrix} \\
&= \left| \beta\,(1 + \lambda)\, M(x)^{-1} + \lambda\,(1 + \lambda)\, I - \beta\, S(x)\, M(x)^{-1} \right| \\
&= |T(x)\, M(x)^{-1}|
\end{aligned}$$

where:

$$T(x) = \beta\,(1 + \lambda)\, I + \lambda\,(1 + \lambda)\, M(x) - \beta\, S(x)$$

The characteristic equation for the system (70) at an equilibrium (71) can be written as:

$$|T(x)| = 0$$

B Stochastic model of *RepAuto* and *RepLeaky*

Species Nomenclature	
G_i	gene i
X_i	protein i
Y_i	mRNA i
CR_i	complex of X_{i-1} with the promoter of G_i
CRR_i	complex of two X_{i-1} with G_i
CA_i	complex of X_i with G_i
CAR_i	complex of X_i and X_{i-1} with G_i

B.1 *RepLeaky*

The model for *RepLeaky* was formulated as in Elowitz and Leibler (2000) [108], with slight changes to allow direct incorporation of the parameters α, *beta*, and δ. The basic parameters were chosen identical for both models.

Reactions for Gene i		
reaction	propensity	description
$X_{i-1} + G_i \longrightarrow CR_i$	$2 \cdot k_a \cdot X_{i-1} \cdot G_i$	first repressor binding
$X_{i-1} + CR_i \longrightarrow CRR_i$	$k_a \cdot X_{i-1} \cdot CR_i$	second repressor binding
$CRR_i \longrightarrow CR_i + X_{i-1}$	$kd_1 \cdot CRR_i$	dissociation of CRR_i
$CR_i \longrightarrow G_i + X_{i-1}$	$kd_2 \cdot CR_i$	dissociation of CR_i
$G_i \longrightarrow G_i + Y_i$	$kts \cdot G_i$	transcription of free gene
$CR_{/2i} \longrightarrow CR_{/2i} + Y_i$	$kts_l \cdot (CR_i + CRR_i)$	transcr. of repr. gene
$Y_i \longrightarrow X_i + Y_i$	$k_{tl} \cdot Y_i$	translation
$Y_i \longrightarrow \emptyset$	$kd_m \cdot Y_i$	mRNA degradation
$X_i \longrightarrow \emptyset$	$kd_p \cdot X_i$	protein degradation

Parameter values			
parameter	value	unit	description
α	216.4	–	rescaled transcription rate
β	0.2	–	ratio protein to mRNA decay
δ	10^{-3}	–	ratio leaky to activated transcr.
eff	20	$proteins/mRNA$	translation efficiency
K_A	44.9	$molecules$	num. of repressors for half max. repression
k_a	60	$1/(molec \cdot min)$	association constant
kd_1	13440	$1/min$	first repressor dissociation const.
kd_2	540	$1/min$	second repressor diss. const.
τ_m	2	min	mRNA half life
τ_p	10	min	protein half life
kd_m	$\ln 2/\tau_m$	$1/min$	mRNA decay constant
kd_p	$\beta \cdot kd_m$	$1/min$	protein half life
kts	$\frac{\alpha \cdot \beta \cdot kd_m \cdot K_A}{eff}$	$1/s$	full transcription rate
kts_l	$kts \cdot \delta$	$1/s$	repressed transcription rate

B.2   *RepAuto*

For mutually exclusive binding of activator and repressor, $\kappa = 0$, all association and dissociation reactions leading for the ternary complex CAR$_i$ where omitted.

Reactions for Gene i		
reaction	propensity	description
$X_i + G_i \longrightarrow CA_i$	$k_a \cdot X_i \cdot G_i$	activator binding
$X_{i\text{-}1} + G_i \longrightarrow CR_i$	$k_a \cdot X_{i-1} \cdot G_i$	repressor binding
$X_{i\text{-}1} + CA_i \longrightarrow CAR_i$	$k_a \cdot X_{i-1} \cdot CA_i$	repressor binding CA$_i$
$X_i + CR_i \longrightarrow CAR_i$	$k_a \cdot X_i \cdot CR_i$	activator binding CR$_i$
$CA_i \longrightarrow G_i + X_i$	$kd_a \cdot CA_i$	activator dissociation from CA$_i$
$CR_i \longrightarrow G_i + X_{i\text{-}1}$	$\frac{kd_a}{\rho} \cdot CR_i$	repressor dissociation from CR$_i$
$CAR_i \longrightarrow CR_i + X_i$	$kd_a \cdot CAR_i$	activator dissociation from CAR$_i$
$CAR_i \longrightarrow CA_i + X_{i\text{-}1}$	$\frac{kd_a}{\rho \cdot \kappa} \cdot CAR_i$	repressor dissociation from CAR$_i$
$CA_i \longrightarrow CA_i + Y_i$	$kts \cdot CA_i$	transcription of act. gene
$G_i \longrightarrow G_i + Y_i$	$kts_b \cdot G_i$	basal transcription
$Y_i \longrightarrow X_i + Y_i$	$k_{tl} \cdot Y_i$	translation
$Y_i \longrightarrow \emptyset$	$kd_m \cdot Y_i$	mRNA degradation

X$_i$ $\longrightarrow$ $\emptyset$	$kd_p \cdot X_i$	protein degradation

Typical parameter values			
parameter	value	unit	description
α	113.2	–	rescaled transcription rate
β	0.2	–	ratio protein to mRNA decay
δ	10^{-3}	–	ratio of bas. to act. transcription
ρ	2.0	–	ratio of repressor to activator binding affinities
κ	0	–	degree of cooperativity
eff	20	$proteins/mRNA$	translation efficiency
K_A	44.9	$molecules$	half activation number
k_a	60	$1/(molec \cdot min)$	association constant
kd_a	13440	$1/min$	dissociation prop. const. activator
τ_m	2	min	mRNA half life
τ_p	10	min	protein half life
kd_m	$\ln 2/\tau_m$	$1/min$	mRNA decay constant
kd_p	$\beta \cdot kd_m$	$1/min$	protein decay constant
kts	$\frac{\alpha \cdot \beta \cdot kd_m \cdot K_A}{eff}$	$1/s$	full transcription rate constant
kts_b	$kts \cdot \delta$	$1/s$	basal transcription rate

C   Models of GATA Networks

C.1   Base Model of NCR

General reactions for all genes (X)		
description	reaction	kinetic law
basal transcription of gene X	$\emptyset \longrightarrow rX$	$gene\_x \cdot Vx\_b$
RNA transport to cytoplasm	$rX \longrightarrow mX$	$kex\_m \cdot [rX]$
translation	$\emptyset \longrightarrow Xc$	$ktl \cdot [mX] \cdot c2$
protein translocation to cytoplasm	$Xc \longleftrightarrow X$	$kim\_p \cdot [Xc] - kex\_p \cdot [X]$
RNA degradation (nucleus)	$rX \longrightarrow \emptyset$	$D\_r \cdot [rX] \cdot c1$
RNA degradation (cytoplasm)	$mX \longrightarrow \emptyset$	$[mX] \cdot D\_m \cdot c2$
protein degradation (cytoplasm)	$Xc \longrightarrow \emptyset$	$D\_pc \cdot [Xc] \cdot c2$
protein degradation (nucleus)	$X \longrightarrow \emptyset$	$D\_p \cdot [X] \cdot c1$
Additional cytoplasmatic reactions for activators A		
binding to Ure2p	$Ac + U \rightleftharpoons C$	$([Gln] \cdot kC\_ass \cdot [Ac] \cdot [U] - kC\_diss \cdot [C]) \cdot c2$
complex degradation	$C \longrightarrow \emptyset$	$D\_pc \cdot [C] \cdot c2$
Regulated transcription		
regulated transcription of GAT1	$\emptyset \longrightarrow rA_1$	$\dfrac{gene\_a1 \cdot Va1\_A \cdot ([A_1] \cdot Ka1\_A1 + [A_0] \cdot Ka1\_A0)^2}{(1 + [A_1] \cdot Ka1\_A1 + [A_0] \cdot Ka1\_A0 + [I_0] \cdot Ka1\_I0 + [I_1] \cdot Ka1\_I1)^2}$
regulated transcription of DAL80	$\emptyset \longrightarrow rI_1$	$\dfrac{gene\_i1 \cdot Vi1\_A \cdot ([A_1] \cdot Ki1\_A1 + [A_0] \cdot Ki1\_A0)^2}{(1 + [A_1] \cdot Ki1\_A1 + [A_0] \cdot Ki1\_A0 + [I_0] \cdot Ki1\_I0 + [I_1] \cdot Ki1\_I1)^2}$

Initial conditions		
species	value [molec/fl]	description
$gene_{a0}$	1	GLN3 gene
rA_0	0	GLN3 RNA (nucleus)
mA_0	0	GLN3 mRNA (cytoplasm)
A_0c	0	Gln3p (cytoplasm)
A_0	0	Gln3p (nucleus)

C. Models of GATA Networks

C_0	0	Gln3p-Urep complex
gene_{a1}	1	GAT1 gene
rA_1	0	GAT1 RNA (nucleus)
mA_1	0	GAT1 mRNA (cytoplasm)
A_1c	0	Gat1p (cytoplasm)
A_1	0	Gat1p (nucleus)
C_1	0	Gat1p-Urep complex
gene_{i0}	1	GZF3 gene
rI_0	0	GZF3 RNA (nucleus)
mI_0	0	GZF3 mRNA (cytoplasm)
I_0c	0	Gzf3p (cytoplasm)
I_0	0	Gzf3p (nucleus)
gene_{i1}	1	DAL80 gene
rI_1	0	DAL80 RNA (nucleus)
mI_1	0	DAL80 mRNA (cytoplasm)
I_1c	0	Dal80p (cytoplasm)
I_1	0	Dal80p (nucleus)

Parameters for the base NCR model			
name	value	unit	description
c_1	3	fl	nuclear volume
c_2	24	fl	cytoplasmic volume
U_tot	300	molec/fl	total number of Ure2p molecules
V_b	0.03	min^{-1}	GAT1 and GLN3 basal transcription rate
VI_b	0.0075	min^{-1}	GZF3 and DAL80 basal transcription rate
va_vb	10		maximal to basal transcription rate ratio (GAT1)
vi_vIb	20		maximal to basal transcription rate ratio (DAL80)
ba1	1		modifier for TF affinities to GATA sites (GAT1)
bi1	2		modifier for TF affinities to GATA sites (DAL80)
KDa_A	25	molec/fl	dissociation constant from GATA sites (Gat1p,Gln3p)
KDa_I	25	molec/fl	dissociation constant from GATA sites (Dal80p,Gzf3p)
kex_m	10	fl/min	rate constant for RNA export
ktl	20	min^{-1}	translation rate constant
kim_p	10	fl/min	rate constant for protein import to nucleus
kex_p	10	fl/min	rate constant for protein export to nucleus

D_r	0.023	min⁻¹	rate constant of RNA degradation
D_m	0.046	min⁻¹	rate constant of mRNA degradation
D_pc	0.015	min⁻¹	rate constant of cytoplasmic protein degradation
D_p	0.015	min⁻¹	rate constant of nuclear protein degradation
kC_ass	4	fl/(molec min)	association rate constant of Ure2p complex
kC_diss	100	min⁻¹	dissociation rate constant of Ure2p complex

C.2 Base Model of Auto-Activator

Reactions added to or altered from NCR		
description	reaction	kinetic law
transcriptional activation by S	$\emptyset \longrightarrow rA$	$gene\_a \cdot VA\_S \cdot \frac{[\text{S}]}{Ka_S+[\text{S}]}$
RNA degradation (cytoplasm)	$mA \longrightarrow \emptyset$	$[mA] \cdot c2 \cdot \frac{D\_mA \cdot K\_CG + D\_CG \cdot [\text{Gln}]}{K\_CG+[\text{Gln}]}$
binding to Ure2p	$Ac + U \rightleftharpoons C$	$(kC\_ass \cdot [Ac] \cdot [U] - kC\_diss \cdot [C]) \cdot c2$

Parameters for the base Auto-Activator			
name	value	unit	description
V_b	0.000026	min⁻¹	gene a basal transcription rate
V_aA	0.026	min⁻¹	gene a maximal transcription rate bound to A
V_aS	0.00125	min⁻¹	gene a maximal transcription rate bound to S
K_A	10	molec/fl	dissociation constant of A from GATA sites
K_S	100	molec/fl	dissociation constant from gene a for S
kex_m	10	fl/min	rate constant for RNA export
ktl	20	min⁻¹	translation rate constant
kim_p	10	fl/min	rate constant for protein import to nucleus
kex_p	10	fl/min	rate constant for protein export from nucleus
D_r	0.023	min⁻¹	rate constant of RNA degradation (nucleus)
D_m	0.017	min⁻¹	rate constant of mRNA degradation (low nitrogen)
D_CG	0.1	min⁻¹	rate constant of mRNA degradation (bound to Gln)
K_CG	60	molec/fl	dissociation constant for mRNA-Gln binding
D_pc	0.015	min⁻¹	rate constant of cytoplasmic protein degradation

D_p	0.015	min^{-1}	rate constant of nuclear protein degradation
kU_ass	0.1	fl/(molec min)	association rate constant of A-U complex
kU_diss	300	min^{-1}	dissociation rate constant of A-U complex

D MiniCellSim Reactions & Parameters

Reactions		
reaction	kinetic law	description
$\text{TF}_i + \text{G}_j \rightleftharpoons \text{C}_{ij}$	$k_{ass} \cdot [TF_i][G_j] - K_D \cdot k_{ass} \cdot [C_{ij}]$	transcription factor (TF) binding
$\text{C}_A + \text{pol} \longrightarrow \text{polC}_A$	$act \cdot k_{BP} \cdot [pol][C_A]$	RNA polymerase recruitment to activator bound promoters
$\text{G} + \text{pol} \longrightarrow \text{polG}$	$k_{BP} \cdot [pol][G]$	RNA polymerase recruitment to free promoters
$\text{polG} + l\,\text{X}_A \longrightarrow \text{G} + \text{pol} + \text{RNA}$	$k_{TS} \cdot [polG]$	RNA transcription
$\text{polC} + l\,\text{X}_A \longrightarrow \text{C} + \text{pol} + \text{RNA}$	$k_{TS} \cdot [polC]$	RNA transcription
$\text{RNA} \longrightarrow l\,\text{X}_I$	$k_{qR} \cdot [RNA]$	RNA decay
$\text{X}_A \longrightarrow \text{X}_I$	$k_{qX} \cdot [X_A]$	nucleotide deactivation
$\text{X}_I + \text{E}_A \longrightarrow \text{X}_A + \text{E}_I$	$k_X \cdot [X_I][E_A]$	nucleotide activation
$\text{E}_I \xrightarrow{SR} \text{E}_I$	$k_{EA} \cdot [E_I][SR]$	metabolite activation by SR
$\text{E}_A \longrightarrow \text{E}_I$	$k_{qE} \cdot [E_A]$	metabolite deactivation
$\text{M} \longrightarrow \emptyset$	$k_{qM} \cdot [M]$	membrane decay
$\text{M}_I^{ext} \rightleftharpoons \text{M}_I^{int}$	$k_{ex} \cdot A \cdot ([M_I^{ext}] - [M_I^{int}])$	memb. building block transport
$\text{M}_I^{int} \xrightarrow{SR} \text{M}$	$kaM \cdot eff \frac{[M_I^{int}][SR]}{V\dot{K}_M + [M_I^{int}]}$	membrane growth via SR

Initial conditions		
species	value [molec/fl]	description
pol	20	free polymerase
X_A	10000	activated building blocks
X_I	0	inactive building blocks
E_A	100	activated metabolites
E_I	0	inactive metabolites
M_I^{ext}	1	external membrane building blocks
M_I^{int}	0	cytosolic membrane building blocks
M	50	membrane

Parameter values			
parameter	value	unit	description
l	$-$	$-$	gene length
k_{TS}	$\frac{[X_A]^2}{cTS_0 \cdot [X_A]^2 + cTS_1 \cdot [X_A] + cTS_2}$	$1/s$	transcription rate
k_{PB}	$\frac{[X_A]^2}{cPB_0 \cdot [X_A]^2 + cPB_1 \cdot [X_A] + cPB_2}$	$1/s$	polymerase binding rate
K_{Dij}	$e^{-\Delta G_{ij}/\beta}$	$molecules/fl$	TF dissociation constant
A	$aM \cdot M$	pm^2	membrane area
R	$\sqrt{\frac{A}{4\pi}}$	μm	cell radius
V	$A \cdot R/3$	fl	cell volume
k_{ass}	10	$fl/(molecules \cdot s)$	TF association rate
k_{qR}	0.1	$1/s$	RNA degradation rate constant
k_X	10	$1/s$	X_I activation rate constant
k_{qX}	0.01	$1/s$	X_A deactivation rate constant
k_{EA}	15	$1/s$	E_I activation rate constant
k_{qE}	0.1	$1/s$	E_A deactivation rate constant
K_{ME}	20	$molecules/fl$	MM constant for E_A
k_{ex}	10	$1/s$	M_I transport rate
k_{aM}	1	$1/s$	M growth rate
aM	0.1	pm^2	area per membrane element
K_{MI}	2	$molecules/fl$	MM constant for M growth
k_{qM}	0.05	$1/s$	M decay rate constant
act	100	-	TF dissociation constant
$CTS0$	10	-	k_{TS} constant
$CTS1$	20	$fl/molecules$	k_{TS} constant
$CTS2$	10^5	$(fl/molecules)^2$	k_{TS} constant
$CPB0$	10	-	k_{PB} constant

$CPB1$	10	$fl/molecules$	k_{PB} constant
$CPB2$	10^5	$(fl/molecules)^2$	k_{PB} constant

References

[1] Ackers GK, Johnson AD, and Shea MA. Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci U S A*, 79(4):1129–1133, 1982.

[2] Adair GS. The hemoglobin system. IV. The oxygen dissociation curve of hemoglobin. *J Biol Chem*, 63:529–545, 1925.

[3] Aittokallio T and Schwikowski B. Graph-based methods for analysing networks in cell biology. *Brief Bioinform*, 7(3):243–255, 2006.

[4] Akker SA, Smith PJ, and Chew SL. Nuclear post-transcriptional control of gene expression. *J Mol Endocrinol*, 27(2):123–131, 2001.

[5] Albert R and Othmer HG. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J Theor Biol*, 223(1):1–18, 2003.

[6] Aldana M and Cluzel P. A natural class of robust networks. *Proc Natl Acad Sci U S A*, 100(15):8710–8714, 2003.

[7] Alfonzo JD, Thiemann O, and Simpson L. The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. *Nucleic Acids Res*, 25(19):3751–3759, 1997.

[8] Alon U, Surette MG, Barkai N, and Leibler S. Robustness in bacterial chemotaxis. *Nature*, 397(6715):168–171, 1999.

[9] Altschul SF, Madden TL, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17):3389–3402, 1997.

[10] An Z, Mei B, Yuan WM, and Leong SA. The distal GATA sequences of the sid1 promoter of Ustilago maydis mediate iron repression of siderophore production and interact directly with Urbs1, a GATA family transcription factor. *EMBO J*, 16(7):1742–1750, 1997.

[11] Anastasia Deckard and Herbert M Sauro. Preliminary studies on the in silico evolution of biochemical networks. *Chembiochem*, 5(10):1423–1431, 2004.

[12] Angeli D, Ferrell J, and Sontag E. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc Natl Acad Sci U S A*, 101(7):1822–1827, 2004.

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

[14] Arnold S. *Kinetic Modelling of Gene Expression*. Ph.D. thesis, Univ. Stuttgart, 2003.

[15] Ascher D, Dubois PF, Hinsen K, Hugunin J, and Oliphant T. *Numerical Python*. Lawrence Livermore National Laboratory, Livermore, CA, ucrl-ma-128569 edn., 1999.

[16] Atkins PW. *Physical Chemistry*. Oxford University Press, 4 edn., 1990.

[17] Atkinson MR, Savageau MA, Myers JT, and Ninfa AJ. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell*, 113(5):597–607, 2003.

[18] Ausbrooks R, Buswell S, *et al.* Mathematical Markup Language (MathML) Version 2.0 (Second Edition). Tech. rep., W3C, 2003.

[19] Babu MM, Luscombe NM, Aravind L, Gerstein M, and Teichmann SA. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol*, 14(3):283–291, 2004.

[20] Bagci EZ, Vodovotz Y, Billiar TR, Ermentrout GB, and Bahar I. Bistability in apoptosis: roles of bax, bcl-2, and mitochondrial permeability transition pores. *Biophys J*, 90(5):1546–1559, 2006.

[21] Bähler J and Svetina S. A logical circuit for the regulation of fission yeast growth modes. *J Theor Biol*, 237(2):210–218, 2005.

[22] Bailey-Serres J, Sorenson R, and Juntawong P. Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends Plant Sci*, 14(8):443–453, 2009.

[23] Banga JR. Optimization in computational systems biology. *BMC Syst Biol*, 2:47, 2008.

[24] Banks HT and Mahaffy JM. Stability of cyclic gene models for systems involving repression. *J Theor Biol*, 74(2):323–334, 1978.

[25] Banzhaf W. *Artificial Regulatory Networks and Genetic Programming*, chap. 4. Kluwer Academic Publishers, 2003.

[26] Banzhaf W. On the Dynamics of an Artificial Regulatory Network. In *Advances in Artificial Life, Proceedings of the 7th European Conference (ECAL-2003)*, vol. 2801 of *Lecture Notes in Artificial Intelligence*, 217–227. Springer-Verlag, 2003.

[27] Banzhaf W and Miller J. *The Challenge of Complexity*, chap. 1. Kluwer Academic Publishers, 2004.

[28] Bar Or RL, Maya R, *et al.* Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. *Proc Natl Acad Sci U S A*, 97(21):11250–11255, 2000.

[29] Barabási AL and Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5(2):101–113, 2004.

[30] Barkai N and Leibler S. Robustness in simple biochemical networks. *Nature*, 387(6636):913–917, 1997.

[31] Basu S, Mehreja R, Thiberge S, Chen MT, and Weiss R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A*, 101(17):6355–6360, 2004.

[32] Beard DA, dan Liang S, and Qian H. Energy balance for analysis of complex metabolic networks. *Biophys J*, 83(1):79–86, 2002.

[33] Beaudry AA and Joyce GF. Directed evolution of an RNA enzyme. *Science*, 257(5070):635–641, 1992.

[34] Beck T and Hall MN. The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors. *Nature*, 402(6762):689–692, 1999.

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

[36] Benkö G, Flamm C, and Stadler PF. A Graph-Based Toy Model of Chemistry. *J. Chem. Inf. Comput. Sci.*, 43:1085–1093, 2003.

[37] Bernhart SH, Tafer H, *et al.* Partition Function and Base Pair Probabilities of RNA Heterodimers. *Algo. Mol. Biol.*, 2006. In press.

[38] Bertram PG, Choi JH, *et al.* Tripartite regulation of Gln3p by TOR, Ure2p, and phosphatases. *J Biol Chem*, 275(46):35727–35733, 2000.

[39] Betrán E and Long M. Expansion of genome coding regions by acquisition of new genes. *Genetica*, 115(1):65–80, 2002.

[40] Betz A and Chance B. PHase Relationship Of Glycolytic Intermediates In Yeast Cells With Oscillatory Metabolic Control. *Arch Biochem Biophys*, 109:585–594, 1965.

[41] Bhalla U and Iyengar R. Emergent properties of networks of biological signaling pathways. *Science*, 283(5400):381–387, 1999.

[42] Bhattacharya S, Conolly RB, *et al.* A bistable switch underlying B-cell differentiation and its disruption by the environmental contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol Sci*, 115(1):51–65, 2010.

[43] Bhaumik SR and Green MR. SAGA is an essential in vivo target of the yeast acidic activator Gal4p. *Genes Dev*, 15(15):1935–1945, 2001.

[44] Biebricher CK, Eigen M, and Gardiner WC. Kinetics of RNA replication. *Biochemistry*, 22(10):2544–2559, 1983.

[45] Bier M, Teusink B, Kholodenko BN, and Westerhoff HV. Control analysis of glycolytic oscillations. *Biophys Chem*, 62(1-3):15–24, 1996.

[46] Bintu L, Buchler NE, *et al.* Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*, 15(2):116–124, 2005.

[47] Birchler JA, Riddle NC, Auger DL, and Veitia RA. Dosage balance in gene regulation: biological implications. *Trends Genet*, 21(4):219–226, 2005.

[48] Blackwood EM and Kadonaga JT. Going the distance: a current view of enhancer action. *Science*, 281(5373):60–63, 1998.

[49] Bliss RD, Painter PR, and Marr AG. Role of feedback inhibition in stabilizing the classical operon. *J Theor Biol*, 97(2):177–193, 1982.

[50] Boczko EM, Cooper TG, *et al.* Structure theorems and the dynamics of nitrogen catabolite repression in yeast. *Proc Natl Acad Sci U S A*, 102(16):5647–5652, 2005.

[51] Boer VM, de Winde JH, Pronk JT, and Piper MDW. The genome-wide transcriptional responses of Saccharomyces cerevisiae grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J Biol Chem*, 278(5):3265–3274, 2003.

[52] Bongard J. Evolving modular genetic regulatory networks. In *Proceedings of the IEEE 2002 Congress on Evolutionary Computation (CEC '02)*, vol. 2, 1872–1877. IEEE Press, Honolulu, HI USA, 2002.

[53] Bordbar A, Jamshidi N, and Palsson BO. iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. *BMC Syst Biol*, 5:110, 2011.

[54] Bordbar A, Lewis NE, Schellenberger J, Q Palsson B, and Jamshidi N. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol Syst Biol*, 6:422, 2010.

[55] Borghans JM, Dupont G, and Goldbeter A. Complex intracellular calcium oscillations. A theoretical exploration of possible mechanisms. *Biophys Chem*, 66(1):25–41, 1997.

[56] Bornstein BJ, Keating SM, Jouraku A, and Hucka M. LibSBML: an API library for SBML. *Bioinformatics*, 24(6):880–881, 2008.

[57] Boyer LA, Lee TI, *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–956, 2005.

[58] Braunewell S and Bornholdt S. Superstability of the yeast cell-cycle dynamics: ensuring causality in the presence of biochemical stochasticity. *J Theor Biol*, 245(4):638–643, 2007.

References

[59] Brewer AC, Alexandrovich A, *et al.* GATA factors lie upstream of Nkx 2.5 in the transcriptional regulatory cascade that effects cardiogenesis. *Stem Cells Dev*, 14(4):425–439, 2005.

[60] Brion P and Westhof E. Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct*, 26:113–137, 1997.

[61] Brown CO, Chi X, *et al.* The cardiac determination factor, Nkx2-5, is activated by mutual cofactors GATA-4 and Smad1/4 via a novel upstream enhancer. *J Biol Chem*, 279(11):10659–10669, 2004.

[62] Browning DF and Busby SJ. The regulation of bacterial transcription initiation. *Nat Rev Microbiol*, 2(1):57–65, 2004.

[63] Bruggeman FJ and Westerhoff HV. The nature of systems biology. *Trends Microbiol*, 15(1):45–50, 2007.

[64] Bundschuh R, Hayot F, and Jayaprakash C. The role of dimerization in noise reduction of simple genetic networks. *J Theor Biol*, 220(2):261–269, 2003.

[65] Busby S and Ebright RH. Transcription activation by catabolite activator protein (CAP). *J Mol Biol*, 293(2):199–213, 1999.

[66] Calzone L, Fages F, and Soliman S. BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics*, 22(14):1805–1807, 2006.

[67] Canavier CC, Clark JW, and Byrne JH. Routes to chaos in a model of a bursting neuron. *Biophys J*, 57(6):1245–1251, 1990.

[68] Cardenas ME, Cutler NS, Lorenz MC, Como CJD, and Heitman J. The TOR signaling cascade regulates gene expression in response to nutrients. *Genes Dev*, 13(24):3271–3279, 1999.

[69] Carey M, Lin YS, Green MR, and Ptashne M. A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature*, 345(6273):361–364, 1990.

[70] Carrier TA and Keasling JD. Controlling messenger RNA stability in bacteria: strategies for engineering gene expression. *Biotechnol Prog*, 13(6):699–708, 1997.

[71] Caspi R, Foerster H, *et al.* The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*, 36(Database issue):D623–D631, 2008.

[72] Castrillo J, Zeef L, *et al.* Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol*, 6(2):4, 2007.

[73] Chai Y, Norman T, Kolter R, and Losick R. An epigenetic switch governing daughter cell separation in Bacillus subtilis. *Genes Dev*, 24(8):754–765, 2010.

[74] Chance B, Greenstein DS, Higgins J, and Yang CC. The mechanism of catalase action. II. Electric analog computer studies. *Arch Biochem*, 37(2):322–339, 1952.

[75] Chang A, Scheer M, Grote A, Schomburg I, and Schomburg D. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res*, 2008.

[76] Chant A, Provatopoulou X, Manfield IW, and Kneale GG. Structural and functional characterisation of the DNA binding domain of the Aspergillus nidulans gene regulatory protein AreA. *Biochim Biophys Acta*, 1648(1-2):84–89, 2003.

[77] Chaves M, Sontag ED, and Albert R. Methods of robustness analysis for Boolean models of gene control networks. *Syst Biol (Stevenage)*, 153(4):154–167, 2006.

[78] Chickarmane V, Troein C, Nuber UA, Sauro HM, and Peterson C. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol*, 2(9):e123, 2006.

[79] Chou IC and Voit EO. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math Biosci*, 219(2):57–83, 2009.

[80] Chou KC. Graphic rules in steady and non-steady state enzyme kinetics. *J Biol Chem*, 264(20):12074–12079, 1989.

[81] Christoffels A, Koh EGL, *et al.* Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol*, 21(6):1146–1151, 2004.

[82] Clewley R, Soto-Trevino C, and Nadim F. Dominant Ionic Mechanisms Explored in Spiking and Bursting Using Local Low-Dimensional Reductions of a Biophysically Realistic Model Neuron. *J. Comput. Neurosci.*, 26:75–90, 2009.

[83] Coffman J, Rai R, *et al.* Cross regulation of four GATA factors that control nitrogen catabolic gene expression in Saccharomyces cerevisiae. *Journal of Bacteriology*, 179:3416–3429, 1997.

[84] Cooper TG. Transmitting the signal of excess nitrogen in Saccharomyces cerevisiae from the Tor proteins to the GATA factors: connecting the dots. *FEMS Microbiol Rev*, 26(3):223–238, 2002.

[85] Cornish Bowden A. *Fundamentals of Enzyme Kinetics*. Portland Press, 2004.

[86] Covert MW, Schilling CH, and Palsson B. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol*, 213(1):73–88, 2001.

[87] Covert MW, Xiao N, Chen TJ, and Karr JR. Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. *Bioinformatics*, 24(18):2044–2050, 2008.

[88] Crampin EJ, Schnell S, and McSharry PE. Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Prog Biophys Mol Biol*, 86(1):77–112, 2004.

[89] Crick FH. Codon–anticodon pairing: the wobble hypothesis. *J Mol Biol*, 19(2):548–555, 1966.

[90] Cross FR, Archambault V, Miller M, and Klovstad M. Testing a mathematical model of the yeast cell cycle. *Mol Biol Cell*, 13(1):52–70, 2002.

[91] Cunningham TS, Andhare R, and Cooper TG. Nitrogen catabolite repression of DAL80 expression depends on the relative levels of Gat1p and Ure2p production in Saccharomyces cerevisiae. *J Biol Chem*, 275(19):14408–14414, 2000.

[92] Cunningham TS, Rai R, and Cooper TG. The level of DAL80 expression down-regulates GATA factor-mediated transcription in Saccharomyces cerevisiae. *J Bacteriol*, 182(23):6584–6591, 2000.

[93] Danino T, Mondragón-Palomino O, Tsimring L, and Hasty J. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326–330, 2010.

[94] Dano S, Sorensen PG, and Hynne F. Sustained oscillations in living cells. *Nature*, 402(6759):320–322, 1999.

[95] Davidson EH. *Genomic Regulatory Networks*. Academic Press, 2001.

[96] de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*, 9(1):67–103, 2002.

[97] Dellaert F and Beer RD. A Developmental Model for the Evolution of Complete Autonomous Agents. In P Maes, MJ Mataric, JA Meyer, J Pollack, and SW Wilson, eds., *From Animals to Animates 4: Proceedings of the 4th Conference on Simulation of Adaptive Behavior*. MIT Press, Cape Cod, Massachusetts USA, 1996.

[98] Dhooge A, Govaerts W, and Kuznetsov Y. MATCONT: A MATLAB package for numerical bifurcation analysis of ODEs. *ACM Trans. Math. Software*, 29:141–164, 2003.

[99] Doedel EJ. Auto: A program for the automatic bifurcation analysis of autonomous systems. *Congressus Numerantium*, 30:265–284, 1981.

[100] Doshi KJ, Cannone JJ, Cobaugh CW, and Gutell RR. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics*, 5:105, 2004.

[101] Duarte NC, Becker SA, *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, 104(6):1777–1782, 2007.

References

[102] Dujon B, Sherman D, *et al.* Genome evolution in yeasts. *Nature*, 430(6995):35–44, 2004.

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

[104] Ebner M, Shackleton M, and Shipman R. How Neutral Networks Influence Evolvability. *Complexity*, 7(2):19–33, 2001.

[105] Ederer M and Gilles ED. Thermodynamically feasible kinetic models of reaction networks. *Biophys J*, 92(6):1846–1857, 2007.

[106] Edwards JS and Palsson BO. The *Escherichia coli* MG 1655 *in silico* Metabolic Genotype: Its Definition, Characteristics, and Capabilities. *Proc. Nat. Acad. Sci. USA*, 97:5528–5533, 2000.

[107] Eggenberg P. Evolving Morphologies of Simulated 3D Organisms based on Differential Gene Expression. In P Husbands and I Harvey, eds., *Proceedings of the 4th European Conference on Artificial Life (ECAL '97)*. MIT Press, London, UK Europe, 1997.

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

[109] Elowitz MB, Levine AJ, Siggia ED, and Swain PS. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[110] Endler L, Rodriguez N, *et al.* Designing and encoding models for synthetic biology. *J R Soc Interface*, 6 Suppl 4:S405–S417, 2009.

[111] Entus R, Aufderheide B, and Sauro HM. Design and implementation of three incoherent feed-forward motif based biological concentration sensors. *Syst Synth Biol*, 1(3):119–128, 2007.

[112] Epstein IR and Showalter K. Nonlinear chemical dynamics: Oscillations, patterns, and chaos. *j. Phys. Chem.*, 100(31):13132–13147, 1996.

[113] Ermentrout B. *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*. SIAM, 2002.

[114] Esch HV, Groenen P, *et al.* GATA3 haplo-insufficiency causes human HDR syndrome. *Nature*, 406(6794):419–422, 2000.

[115] Famili I and Palsson BO. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys J*, 85(1):16–26, 2003.

[116] Fauré A, Naldi A, *et al.* Modular logical modelling of the budding yeast cell cycle. *Mol Biosyst*, 5(12):1787–1796, 2009.

[117] Feist AM, Henry CS, *et al.* A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007.

[118] Ferrell JE. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol*, 14(2):140–148, 2002.

[119] Ferrell JE and Machleder EM. The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science*, 280(5365):895–898, 1998.

[120] Fiering S, Northrop JP, *et al.* Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev*, 4(10):1823–1834, 1990.

[121] Fiering S, Whitelaw E, and Martin DI. To be or not to be active: the stochastic nature of enhancer action. *Bioessays*, 22(4):381–387, 2000.

[122] Fitzpatrick DA, Logue ME, Stajich JE, and Butler G. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol Biol*, 6:99, 2006.

[123] Flamm C, Endler L, Müller S, Widder S, and Schuster P. A minimal and self-consistent in silico cell model based on macromolecular interactions. *Philos Trans R Soc Lond B Biol Sci*, 362(1486):1831–1839, 2007.

[124] Flamm C, Ullrich A, *et al.* Evolution of Metabolic Networks: A Computational Framework. *Journal of Systems Chemistry*, 1:4, 2010.

[125] Fontana W. Modelling 'Evo-Devo' with RNA. *BioEssays*, 24:1164–1177, 2002.

[126] Fontana W and Schuster P. Continuity in Evolution: On the Nature of Transitions. *Science*, 280:145–165, 1998.

[127] Fontana W and Schuster P. Shaping Space. The Possible and the Attainable in RNA Genotype-Phenotype Mapping. *J. Theor. Biol.*, 194:491–515, 1998.

[128] Force A, Lynch M, *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151(4):1531–1545, 1999.

[129] Fraser A and Tiwari J. Genetical feedback-repression. II. Cyclic genetic systems. *J Theor Biol*, 47(2):397–412, 1974.

[130] Frech GC, Bakalara N, Simpson L, and Simpson AM. In vitro RNA editing-like activity in a mitochondrial extract from Leishmania tarentolae. *EMBO J*, 14(1):178–187, 1995.

[131] Fritsche-Guenther R, Witzel F, *et al.* Strong negative feedback from Erk to Raf confers robustness to MAPK signalling. *Mol Syst Biol*, 7:489, 2011.

[132] Fukushige T, Goszczynski B, Tian H, and McGhee JD. The evolutionary duplication and probable demise of an endodermal GATA factor in Caenorhabditis elegans. *Genetics*, 165(2):575–588, 2003.

[133] Funahashi A, Morohashi M, Kitano H, and Tanimura N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO*, 1(5):159–162, 2003. ISSN 14785382.

[134] Galagan JE, Calvo SE, *et al.* Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. *Nature*, 438(7071):1105–1115, 2005.

[135] Gama-Castro S, Salgado H, *et al.* RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res*, 39(Database issue):D98–105, 2011.

[136] Gantí T. Organization of chemical reactions into dividing and metabolizing units: The Chemotons. *Biosystems*, 7(1):15–21, 1975.

[137] Garcia-Ojalvo J, Elowitz MB, and Strogatz SH. Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proc Natl Acad Sci U S A*, 101(30):10955–10960, 2004.

[138] Gardner TS, Cantor CR, and Collins JJ. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, 2000.

[139] Gasch AP, Moses AM, *et al.* Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biol*, 2(12):e398, 2004.

[140] Gedeon T, Mischaikow K, Patterson K, and Traldi E. When activators repress and repressors activate: a qualitative analysis of the Shea-Ackers model. *Bull Math Biol*, 70(6):1660–1683, 2008.

[141] Georis I, Feller A, Vierendeels F, and Dubois E. The yeast GATA factor Gat1 occupies a central position in nitrogen catabolite repression-sensitive gene activation. *Mol Cell Biol*, 29(13):3803–3815, 2009.

[142] Georis I, Tate JJ, Cooper TG, and Dubois E. Tor pathway control of the nitrogen-responsive DAL5 gene bifurcates at the level of Gln3 and Gat1 regulation in Saccharomyces cerevisiae. *J Biol Chem*, 283(14):8919–8929, 2008.

[143] Gertz J, Siggia ED, and Cohen BA. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, 457(7226):215–218, 2009.

[144] Ghaemmaghami S, Huh WK, *et al.* Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741, 2003.

[145] Gibson MA and Bruck J. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

[146] Gille C, Bölling C, *et al.* HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol Syst Biol*, 6:411, 2010.

[147] Gillespie DT. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[148] Gillespie DT. Exact Stochastic Simulation of Coupled Chemical Reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

References

[149] Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, 115:1716–1733, 2001.

[150] Gillespie DT. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem*, 58:35–55, 2007.

[151] Gillis WQ, John JS, Bowerman B, and Schneider SQ. Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. *BMC Evol Biol*, 9:207, 2009.

[152] Glossop NR, Lyons LC, and Hardin PE. Interlocked feedback loops within the Drosophila circadian oscillator. *Science*, 286(5440):766–768, 1999.

[153] Godard P, Urrestarazu A, *et al.* Effect of 21 different nitrogen sources on global gene expression in the yeast Saccharomyces cerevisiae. *Mol Cell Biol*, 27(8):3065–3086, 2007.

[154] Goldberg RN, Tewari YB, and Bhat TN. Thermodynamics of enzyme-catalyzed reactions–a database for quantitative biochemistry. *Bioinformatics*, 20(16):2874–2877, 2004.

[155] Goldbeter A, Gonze D, and Pourquié O. Sharp developmental thresholds defined through bistability by antagonistic gradients of retinoic acid and FGF signaling. *Dev Dyn*, 236(6):1495–1508, 2007.

[156] Goldbeter A and Koshland DE. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A*, 78(11):6840–6844, 1981.

[157] Goldbeter A and Koshland DE. Ultrasensitivity in biochemical systems controlled by covalent modification. Interplay between zero-order and multistep effects. *J Biol Chem*, 259(23):14441–14447, 1984.

[158] Gómez D, García I, Scazzocchio C, and Cubero B. Multiple GATA sites: protein binding and physiological relevance for the regulation of the proline transporter gene of Aspergillus nidulans. *Mol Microbiol*, 50(1):277–289, 2003.

[159] Goodwin BC. Oscillatory behavior in enzymatic control processes. *Adv Enzyme Regul*, 3:425–438, 1965.

[160] Grass JA, Boyer ME, *et al.* GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A*, 100(15):8811–8816, 2003.

[161] Grass JA, Jing H, *et al.* Distinct functions of dispersed GATA factor complexes at an endogenous gene locus. *Mol Cell Biol*, 26(19):7056–7067, 2006.

[162] Griffith JS. Mathematics of cellular control processes. I. Negative feedback to one gene. *J Theor Biol*, 20(2):202–208, 1968.

[163] Griffith JS. Mathematics of cellular control processes. II. Positive feedback to one gene. *J Theor Biol*, 20(2):209–216, 1968.

[164] Grima R and Schnell S. A systematic investigation of the rate laws valid in intracellular environments. *Biophys Chem*, 124(1):1–10, 2006.

[165] Gruber AR, Bernhart SH, Hofacker IL, and Washietl S. Strategies for measuring evolutionary conservation of RNA secondary structures. *BMC Bioinformatics*, 9:122, 2008.

[166] Guerrier-Takada C, Gardiner K, Marsh T, Pace N, and Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 Pt 2):849–857, 1983.

[167] Haas H. Molecular genetics of fungal siderophore biosynthesis and uptake: the role of siderophores in iron uptake and storage. *Appl Microbiol Biotechnol*, 62(4):316–330, 2003.

[168] Haas H, Zadra I, Stöffler G, and Angermayr K. The Aspergillus nidulans GATA factor SREA is involved in regulation of siderophore biosynthesis and control of iron uptake. *J Biol Chem*, 274(8):4613–4619, 1999.

[169] Hahn MW, Demuth JP, and Han SG. Accelerated rate of gene gain and loss in primates. *Genetics*, 177(3):1941–1949, 2007.

[170] Hakenberg J, Schmeier S, Kowald A, Klipp E, and Leser U. Finding kinetic parameters using text mining. *OMICS*, 8(2):131–152, 2004.

[171] Hall D and Minton AP. Macromolecular crowding: qualitative and semi-quantitative successes, quantitative challenges. *Biochim Biophys Acta*, 1649(2):127–139, 2003.

[172] Hänggi P. Stochastic resonance in biology. How noise can enhance detection of weak signals and help improve biological information processing. *Chemphyschem*, 3(3):285–290, 2002.

[173] Hardwick JS, Kuruvilla FG, Tong JK, Shamji AF, and Schreiber SL. Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. *Proc Natl Acad Sci U S A*, 96(26):14866–14870, 1999.

[174] He C, Cheng H, and Zhou R. GATA family of transcription factors of vertebrates: phylogenetics and chromosomal synteny. *J Biosci*, 32(7):1273–1280, 2007.

[175] He L and Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*, 5(7):522–531, 2004.

[176] He X, Samee MAH, Blatti C, and Sinha S. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Comput Biol*, 6(9), 2010.

[177] Heijnen JJ. Approximative kinetic formats used in metabolic network modeling. *Biotechnol Bioeng*, 91(5):534–545, 2005.

[178] Heinemann M and Panke S. Synthetic biology–putting engineering into biology. *Bioinformatics*, 22(22):2790–2799, 2006.

[179] Heinrich R and Schuster S. *The Regulation Of Cellular Systems.* Springer, 1 edn., 1996.

[180] Heldwein EE and Brennan RG. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature*, 409(6818):378–382, 2001.

[181] Herrgård MJ, Swainston N, *et al.* A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol*, 26(10):1155–1160, 2008.

[182] Hilioti Z, Sabbagh W, *et al.* Oscillatory phosphorylation of yeast Fus3 MAP kinase controls periodic gene expression and morphogenesis. *Curr Biol*, 18(21):1700–1706, 2008.

[183] Hill AV. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol. (Lond.)*, 40:4–7, 1910.

[184] Hines ML, Morse T, Migliore M, Carnevale NT, and Shepherd GM. ModelDB: A Database to Support Computational Neuroscience. *J Comput Neurosci*, 17(1):7–11, 2004.

[185] Hinze T, Schumann M, Bodenstein C, Heiland I, and Schuster S. Biochemical frequency control by synchronisation of coupled repressilators: an in silico study of modules for circadian clock systems. *Comput Intell Neurosci*, 2011:262189, 2011.

[186] Hirata H, Yoshiura S, *et al.* Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. *Science*, 298(5594):840–843, 2002.

[187] Hittinger CT and Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature*, 449(7163):677–681, 2007.

[188] Hochschild A and Lewis M. The bacteriophage lambda CI protein finds an asymmetric solution. *Curr Opin Struct Biol*, 19(1):79–86, 2009.

[189] Hodgkin AL and Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol*, 117(4):500–544, 1952.

[190] Hofacker IL, Fontana W, *et al.* Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie / Chemical Monthly*, 125(2):167–188, 1994.

[191] Hofbauer J and Sigmund K. *Evolutionary Games and Population Dynamics.* Cambridge University Press, 1998.

[192] Höfer T, Nathansen H, Löhning M, Radbruch A, and Heinrich R. GATA-3 transcriptional imprinting in Th2 lymphocytes: a mathematical model. *Proc Natl Acad Sci U S A*, 99(14):9364–9368, 2002.

References

[193] Hoffmann A, Levchenko A, Scott ML, and Baltimore D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science*, 298(5596):1241–1245, 2002.

[194] Hofmeyr JH and Cornish-Bowden A. The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. *Comput Appl Biosci*, 13(4):377–385, 1997.

[195] Holstege FC, Jennings EG, *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, 95(5):717–728, 1998.

[196] Hoops S, Sahle S, *et al.* COPASI–a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[197] Houart G, Dupont G, and Goldbeter A. Bursting, Chaos and Birhythmicity Originating from Self-modulation of the Inositol 1,4,5-trisphosphate Signal in a Model for Intracellular Ca^{2+} Oscillations. *Bull. of Math. Biol.*, 61:507–530, 1999.

[198] Huang CY and Ferrell JE. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci U S A*, 93(19):10078–10083, 1996.

[199] Huang DY, Kuo YY, and Chang ZF. GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells. *Nucleic Acids Res*, 33(16):5331–5342, 2005.

[200] Hucka M, Finney A, *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[201] Hume DA. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood*, 96(7):2323–2328, 2000.

[202] Huynen MA, Stadler PF, and Fontana W. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93:397–401, 1996.

[203] Izhikevich E. Neural Excitability, Spiking and Bursting. *International Journal of Bifurcation and Chaos*, 10:1171–1266, 2000.

[204] Jaenisch R and Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*, 33 Suppl:245–254, 2003.

[205] Janssens H, Hou S, *et al.* Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. *Nat Genet*, 38(10):1159–1165, 2006.

[206] Jason A Papin, Joerg Stelling, *et al.* Comparison of network-based pathway analysis methods. *Trends Biotechnol*, 22(8):400–405, 2004.

[207] Jeong H, Tombor B, Albert R, Oltvai ZN, and Barabási AL. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[208] Jones E, Oliphant T, Peterson P, *et al.* SciPy: Open source scientific tools for Python, 2001.

[209] Joung JK, Le LU, and Hochschild A. Synergistic activation of transcription by Escherichia coli cAMP receptor protein. *Proc Natl Acad Sci U S A*, 90(7):3083–3087, 1993.

[210] Kadosh D and Struhl K. Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell*, 89(3):365–371, 1997.

[211] Kalir S, Mangan S, and Alon U. A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli. *Mol Syst Biol*, 1:2005.0006, 2005.

[212] Kanehisa M, Araki M, *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res*, 36(Database issue):D480–D484, 2008.

[213] Kaplan S, Bren A, Zaslaver A, Dekel E, and Alon U. Diverse two-dimensional input functions control bacterial sugar genes. *Mol Cell*, 29(6):786–792, 2008.

[214] Karttunen J and Shastri N. Measurement of ligand-induced activation in single viable T cells using the lacZ reporter gene. *Proc Natl Acad Sci U S A*, 88(9):3972–3976, 1991.

[215] Kasahara M. The 2R hypothesis: an update. *Curr Opin Immunol*, 19(5):547–552, 2007.

[216] Kauffman SA. *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, USA, 1993.

[217] Kennedy PJ and Osborn TR. A Model of Gene Expression and Regulation in an Artificial Cellular Organism. *Complex Systems*, 13:33–59, 2001.

[218] Kholodenko BN. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem*, 267(6):1583–1588, 2000.

[219] Kholodenko BN and Westerhoff HV. The macroworld versus the microworld of biochemical regulation and control. *Trends Biochem Sci*, 20(2):52–54, 1995.

[220] Kim D, Kolch W, and Cho KH. Multiple roles of the NF-kappaB signaling pathway regulated by coupled negative feedback circuits. *FASEB J*, 23(9):2796–2802, 2009.

[221] Kim J and Winfree E. Synthetic in vitro transcriptional oscillators. *Mol Syst Biol*, 7:465, 2011.

[222] Kimura M. *The neutral theory of molecular evolution.* Cambridge University Press., 1983.

[223] Kimura M and Ohta T. On some principles governing molecular evolution. *Proc Natl Acad Sci U S A*, 71(7):2848–2852, 1974.

[224] King EL and Altman C. A Schematic Method Of Deriving The Rate Laws For Enzyme-Catalyzed Reactions. *J Phys Chem*, 60(10):1375–1378, 1956.

[225] Kitano H. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.

[226] Klamt S and Gilles ED. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, 2004.

[227] Klamt S, Saez Rodriguez J, Lindquist JA, Simeoni L, and Gilles ED. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7:56, 2006.

[228] Klotz IM. The Application of the Law of Mass Action to Binding by Proteins. Interactions with Calcium. *Arch Biochem*, 9:109–117, 1946.

[229] Klotz IM. Ligand-receptor complexes: origin and development of the concept. *J Biol Chem*, 279(1):1–12, 2004.

[230] Kopelman R. Fractal Reaction Kinetics. *Science*, 241(4873):1620–1626, 1988.

[231] Kringstein AM, Rossi FM, Hofmann A, and Blau HM. Graded transcriptional response to different concentrations of a single transactivator. *Proc Natl Acad Sci U S A*, 95(23):13670–13675, 1998.

[232] Kruger K, Grabowski PJ, *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31(1):147–157, 1982.

[233] Kumar P, Han BC, *et al.* Update of KDBI: Kinetic Data of Bio-molecular Interaction database. *Nucleic Acids Res*, 2008.

[234] Kuo PD, Leier A, and Banzhaf W. Evolving Dynamics in an Artificial Regulatory Network Model. *Lecture Notes in Computer Science*, 3242:571–580, 2004.

[235] Lahav G, Rosenfeld N, *et al.* Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet*, 36(2):147–150, 2004.

[236] Langmead CJ and Jha SK. Symbolic approaches for finding control strategies in Boolean Networks. *J Bioinform Comput Biol*, 7(2):323–338, 2009.

[237] Lavenu-Bombled C, Trainor CD, Makeh I, Romeo PH, and Max-Audit I. Interleukin-13 gene expression is regulated by GATA-3 in T cells: role of a critical association of a GATA and two GATG motifs. *J Biol Chem*, 277(21):18313–18321, 2002.

[238] Lazebnik Y. Can a biologist fix a radio?–Or, what I learned while studying apoptosis. *Cancer Cell*, 2(3):179–182, 2002.

[239] Le Novère N, Hucka M, *et al.* The Systems Biology Graphical Notation. *Nat Biotechnol*, 27(8):735–741, 2009.

[240] Lee J and Goldfarb A. lac repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. *Cell*, 66(4):793–798, 1991.

[241] Lee L, Yin L, Zhu X, and Ao P. Generic Enzymatic Rate Equation Under Living Conditions. *Journal of Biological Systems*, 15(4):495–514, 2007.

[242] Lee RC, Feinbaum RL, and Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854, 1993.

[243] Lee TI, Rinaldi NJ, *et al.* Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science*, 298(5594):799–804, 2002.

[244] Leloup JC and Goldbeter A. 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.

[245] Leloup JC, Gonze D, and Goldbeter A. Limit cycle models for circadian rhythms based on transcriptional regulation in Drosophila and Neurospora. *J Biol Rhythms*, 14(6):433–448, 1999.

[246] Lewis BA and Reinberg D. The mediator coactivator complex: functional and physical roles in transcriptional regulation. *J Cell Sci*, 116(Pt 18):3667–3675, 2003.

[247] Li C, Donizelli M, *et al.* BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst Biol*, 4:92, 2010.

[248] Li F, Long T, Lu Y, Ouyang Q, and Tang C. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A*, 101(14):4781–4786, 2004.

[249] Liebermeister W and Klipp E. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor Biol Med Model*, 3:41, 2006.

[250] Liebermeister W, Uhlendorf J, and Klipp E. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics*, 26(12):1528–1534, 2010.

[251] Lincoln TA and Joyce GF. Self-sustained replication of an RNA enzyme. *Science*, 323(5918):1229–1232, 2009.

[252] Llaneras F and Picó J. Stoichiometric modelling of cell metabolism. *J Biosci Bioeng*, 105(1):1–11, 2008.

[253] Lloyd CM, Halstead MDB, and Nielsen PF. CellML: its future, present and past. *Prog Biophys Mol Biol*, 85(2-3):433–450, 2004.

[254] Lloyd CM, Lawson JR, Hunter PJ, and Nielsen PF. The CellML Model Repository. *Bioinformatics*, 24(18):2122–2123, 2008.

[255] Lomvardas S, Barnea G, *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell*, 126(2):403–413, 2006.

[256] Lowry JA and Atchley WR. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J Mol Evol*, 50(2):103–115, 2000.

[257] Lu J, Engl HW, and Schuster P. Inverse bifurcation analysis: application to simple gene systems. *Algorithms Mol Biol*, 1:11, 2006.

[258] Machné R, Finney A, *et al.* The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics*, 22(11):1406–1407, 2006.

[259] MacIsaac KD, Wang T, *et al.* An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics*, 7:113, 2006.

[260] Madsen MF, Danø S, and Sørensen PG. On the mechanisms of glycolytic oscillations in yeast. *FEBS J*, 272(11):2648–2660, 2005.

[261] Maduro MF. Endomesoderm specification in Caenorhabditis elegans and other nematodes. *Bioessays*, 28(10):1010–1022, 2006.

[262] Maeda YT and Sano M. Regulatory dynamics of synthetic gene networks with positive feedback. *J Mol Biol*, 359(4):1107–1124, 2006.

[263] Magasanik B. The transduction of the nitrogen regulation signal in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, 102(46):16537–16538, 2005.

[264] Mallet-Paret J and Smith HL. The Poincaré-Bendixson Theorem for Monotone Cyclic Feedback Systems. *Journal Of Dynamics and Differential Equations*, 2(2):367–421, 1990.

[265] Mandal N, Su W, Haber R, Adhya S, and Echols H. DNA looping in cellular repression of transcription of the galactose operon. *Genes Dev*, 4(3):410–418, 1990.

[266] Mangan S and Alon U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A*, 100(21):11980–11985, 2003.

[267] Mangan S, Itzkovitz S, Zaslaver A, and Alon U. The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli. *J Mol Biol*, 356(5):1073–1081, 2006.

[268] Mangan S, Zaslaver A, and Alon U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol*, 334(2):197–204, 2003.

[269] Martiel JL and Goldbeter A. A Model Based on Receptor Desensitization for Cyclic AMP Signaling in Dictyostelium Cells. *Biophys J*, 52(5):807–828, 1987.

[270] Mateus D, Gallois JP, Comet JP, and Gall PL. Symbolic modeling of genetic regulatory networks. *J Bioinform Comput Biol*, 5(2B):627–640, 2007.

[271] Mathews DH, Sabina J, Zuker M, and Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, 288(5):911–940, 1999.

[272] Matsumoto G, Aihara K, *et al.* Chaos and phase locking in normal squid axons. *Physics Letters A*, 123(4):162 – 166, 1987. ISSN 0375-9601.

[273] Matthews L, Gopinath G, *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res*, 2008.

[274] May RM and Leonard WJ. Nonlinear Aspects of Competition Between Three Species. *SIAM Journal on Applied Mathematics*, 29(2):243–253, 1975.

[275] McAdams HH and Arkin A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A*, 94(3):814–819, 1997.

[276] McAdams HH and Arkin A. Simulation of prokaryotic genetic circuits. *Annu Rev Biophys Biomol Struct*, 27:199–224, 1998.

[277] McLeod SM, Aiyar SE, Gourse RL, and Johnson RC. The C-terminal domains of the RNA polymerase alpha subunits: contact site with Fis and localization during co-activation with CRP at the Escherichia coli proP P2 promoter. *J Mol Biol*, 316(3):517–529, 2002.

[278] Mendes P and Kell D. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869–883, 1998.

[279] Mestl T, Plahte E, and Omholt SW. A mathematical framework for describing and analysing gene regulatory networks. *J Theor Biol*, 176(2):291–300, 1995.

[280] Mi H, Guo N, Kejariwal A, and Thomas PD. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res*, 35(Database issue):D247–D252, 2007.

[281] Mills DR, Peterson RL, and Spiegelman S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci U S A*, 58(1):217–224, 1967.

[282] Milo R, Shen Orr S, *et al.* Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[283] Minton AP. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol Cell Biochem*, 55(2):119–140, 1983.

[284] Minton AP. How can biochemical reactions within cells differ from those in test tubes? *J Cell Sci*, 119(Pt 14):2863–2869, 2006.

[285] Moles CG, Mendes P, and Banga JR. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*, 13(11):2467–2474, 2003.

[286] Monod J, Wyman J, and Changeaux JP. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.*, 12:88–118, 1965.

[287] Morgenstern B. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics*, 15(3):211–218, 1999.

[288] Morozov IY, Galbis-Martinez M, Jones MG, and Caddick MX. Characterization of nitrogen metabolite signalling in Aspergillus via the regulated degradation of areA mRNA. *Mol Microbiol*, 42(1):269–277, 2001.

[289] Morozov IY, Martinez MG, Jones MG, and Caddick MX. A defined sequence within the 3' UTR of the areA transcript is sufficient to mediate nitrogen metabolite signalling via accelerated deadenylation. *Mol Microbiol*, 37(5):1248–1257, 2000.

[290] Müller S, Hofbauer J, *et al.* A generalized model of the repressilator. *J Math Biol*, 53(6):905–937, 2006.

[291] Müller-Hill B. Some repressors of bacterial transcription. *Curr Opin Microbiol*, 1(2):145–151, 1998.

[292] Murphy KM, Ouyang W, Ranganath S, and Murphy TL. Bi-stable transcriptional circuitry and GATA-3 auto-activation in Th2 commitment. *Cold Spring Harb Symp Quant Biol*, 64:585–588, 1999.

[293] Murray DB, Beckmann M, and Kitano H. Regulation of yeast oscillatory dynamics. *Proc Natl Acad Sci U S A*, 104(7):2241–2246, 2007.

[294] Myers CR, Gutenkunst RN, and Sethna JP. Python unleashed on systems biology. *Computing in Science and Engineering*, 9(34), 2007.

[295] Napoli C, Lemieux C, and Jorgensen R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell*, 2(4):279–289, 1990.

[296] Nelson DE, Ihekwaba AEC, *et al.* Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science*, 306(5696):704–708, 2004.

[297] Newberry KJ and Brennan RG. The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus. *J Biol Chem*, 279(19):20356–20362, 2004.

[298] Ng A, Bursteinas B, Gao Q, Mollison E, and Zvelebil M. Resources for integrative systems biology: from data through databases to networks and dynamic system models. *Brief Bioinform*, 7(4):318–330, 2006.

[299] Nielsen K, Sorensen PG, Hynne F, and Busse HG. Sustained oscillations in glycolysis: an experimental and theoretical study of chaotic and complex periodic behavior and of quenching of simple oscillations. *Biopphys. Chem.*, 72:49–62, 1998.

[300] Nikolov DB and Burley SK. RNA polymerase II transcription initiation: a structural view. *Proc Natl Acad Sci U S A*, 94(1):15–22, 1997.

[301] Nitzan Rosenfeld, Jonathan W Young, Uri Alon, Peter S Swain, and Michael B Elowitz. Gene regulation at the single-cell level. *Science*, 307(5717):1962–1965, 2005.

[302] Nitzan Rosenfeld, Michael B Elowitz, and Uri Alon. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol*, 323(5):785–793, 2002.

[303] Nookaew I, Jewett MC, *et al.* The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol*, 2:71, 2008.

[304] Nordström K and Uhlin BE. Runaway-replication plasmids as tools to produce large quantities of proteins from cloned genes in bacteria. *Biotechnology (N Y)*, 10(6):661–666, 1992.

[305] Novak B and Tyson JJ. Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. *J Cell Sci*, 106 (Pt 4):1153–1168, 1993.

[306] Novak B and Tyson JJ. Modeling the control of DNA replication in fission yeast. *Proc Natl Acad Sci U S A*, 94(17):9147–9152, 1997.

[307] Novák B and Tyson JJ. Design principles of biochemical oscillators. *Nat Rev Mol Cell Biol*, 9(12):981–991, 2008.

[308] Nudler E and Mironov AS. The riboswitch control of bacterial metabolism. *Trends Biochem Sci*, 29(1):11–17, 2004.

[309] Odom DT, Zizlsperger N, *et al.* Control of pancreas and liver gene expression by HNF transcription factors. *Science*, 303(5662):1378–1381, 2004.

[310] Ohno S. *Evolution by gene duplication.* Springer-Verlag., 1970.

[311] Ohta T. Time for acquiring a new gene by duplication. *Proc Natl Acad Sci U S A*, 85(10):3509–3512, 1988.

[312] Ohta T. Mechanisms of molecular evolution. *Philos Trans R Soc Lond B Biol Sci*, 355(1403):1623–1626, 2000.

[313] Paladugu SR, Chickarmane V, *et al.* In silico evolution of functional modules in biochemical networks. *Syst Biol (Stevenage)*, 153(4):223–235, 2006.

[314] Palani S and Sarkar CA. Positive receptor feedback during lineage commitment can generate ultrasensitivity to ligand and confer robustness to a bistable switch. *Biophys J*, 95(4):1575–1589, 2008.

[315] Palani S and Sarkar CA. Integrating extrinsic and intrinsic cues into a minimal model of lineage commitment for hematopoietic progenitors. *PLoS Comput Biol*, 5(9):e1000518, 2009.

[316] Palmer AC and Shearwin KE. Guidance for data collection and computational modelling of regulatory networks. *Methods Mol Biol*, 541:337–354, 2009.

[317] Palsson BO. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, 1 edn., 2006. ISBN 0521859034.

[318] Papp B, Pál C, and Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. *Nature*, 424(6945):194–197, 2003.

[319] Paszek P, Lipniacki T, *et al.* Stochastic effects of multiple regulators on expression profiles in eukaryotes. *J Theor Biol*, 233(3):423–433, 2005.

[320] Patient RK and McGhee JD. The GATA family (vertebrates and invertebrates). *Curr Opin Genet Dev*, 12(4):416–422, 2002.

[321] Paulsson J, Berg OG, and Ehrenberg M. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci U S A*, 97(13):7148–7153, 2000.

[322] Pehlivan T, Pober BR, *et al.* GATA4 haploinsufficiency in patients with interstitial deletion of chromosome region 8p23.1 and congenital heart disease. *Am J Med Genet*, 83(3):201–206, 1999.

[323] Pelletier B, Trott A, Morano KA, and Labbé S. Functional characterization of the iron-regulatory transcription factor Fep1 from Schizosaccharomyces pombe. *J Biol Chem*, 280(26):25146–25161, 2005.

[324] Pérez-Rueda E and Collado-Vides J. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Res*, 28(8):1838–1847, 2000.

[325] Peterson CL and Tamkun JW. The SWI-SNF complex: a chromatin remodeling machine? *Trends Biochem Sci*, 20(4):143–146, 1995.

[326] Phillipson PE, Schuster P, and Johnston RG. An Analytic Study of the May-Leonard Equations. *SIAM Journal on Applied Mathematics*, 45:541–554, 1985.

[327] Plahte E and Kjoglum S. Analysis and generic properties of gene regulatory networks with graded response functions. *Physica D*, 201:150–176, 2005.

[328] Platt A, Langdon T, *et al.* Nitrogen metabolite signalling involves the C-terminus and the GATA domain of the Aspergillus transcription factor AREA and the 3' untranslated region of its mRNA. *EMBO J*, 15(11):2791–2801, 1996.

[329] Pleij CW, Rietveld K, and Bosch L. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res*, 13(5):1717–1731, 1985.

[330] Pokhilko A, nas Fernández AP, *et al.* The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops. *Mol Syst Biol*, 8:574, 2012.

[331] Polynikis A, Hogan SJ, and di Bernardo M. Comparing different ODE modelling approaches for gene regulatory networks. *J Theor Biol*, 261(4):511–530, 2009.

[332] Poolman MG, Miguet L, Sweetlove LJ, and Fell DA. A genome-scale metabolic model of Arabidopsis and some of its properties. *Plant Physiol*, 151(3):1570–1581, 2009.

[333] Potapov AP, Voss N, Sasse N, and Wingender E. Topology of mammalian transcription networks. *Genome Inform*, 16(2):270–278, 2005.

[334] Prokudina EI, RYu V, and Tchuraev RN. A new method for the analysis of the dynamics of the molecular genetic control systems. II. Application of the method of generalized threshold models in the investigation of concrete genetic systems. *J Theor Biol*, 151(1):89–110, 1991.

[335] Ptashne M. *A Genetic Switch: Phage λ and Higher Organisms*. Cell,Cambridge, Massachusetts, 1992.

[336] Ptashne M. Regulated Recruitment and Cooperativity in the Design of Biological Regulatory Systems. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 361(1807):1223–1234, 2003. ISSN 1364503X.

[337] Ptashne M and Gann A. *Genes & Signals*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2002.

[338] Radde N, Gebert J, and Forst CV. Systematic component selection for gene-network refinement. *Bioinformatics*, 22(21):2674–2680, 2006.

[339] Rai R, Daugherty JR, Cunningham TS, and Cooper TG. Overlapping positive and negative GATA factor binding sites mediate inducible DAL7 gene expression in Saccharomyces cerevisiae. *J Biol Chem*, 274(39):28026–28034, 1999.

[340] Raman K and Chandra N. Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform*, 10(4):435–449, 2009.

[341] Ravagnani A, Gorfinkiel L, *et al.* Subtle hydrophobic interactions between the seventh residue of the zinc finger loop and the first base of an HGATAR sequence determine promoter-specific recognition by the Aspergillus nidulans GATA factor AreA. *EMBO J*, 16(13):3974–3986, 1997.

[342] Ravasz E, Somera AL, Mongru DA, Oltvai ZN, and Barabási AL. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, 2002.

[343] Reidys C, Forst C, and Schuster P. Replication and Mutation on Neutral Networks. *Bull.Math.Biol.*, 63:57–94, 2001.

[344] Reil T. Dynamics of gene expression in an artificial genome-implications for biological and artificial ontogeny. *Advances in Artificial Life–Proceedings of the 5th European Conference on Artificial Life (ECAL '99)*, 1674:457–466, 1999.

[345] Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, and Bartel DP. MicroRNAs in plants. *Genes Dev*, 16(13):1616–1626, 2002.

[346] Richard P, Teusink B, Westerhoff HV, and van Dam K. Around the growth phase transition S. cerevisiae's make-up favours sustained oscillations of intracellular metabolites. *FEBS Lett*, 318(1):80–82, 1993.

[347] Richet E, Vidal-Ingigliardi D, and Raibaud O. A new mechanism for coactivation of transcription initiation: repositioning of an activator triggered by the binding of a second activator. *Cell*, 66(6):1185–1195, 1991.

[348] Rodrigues NP, Janzen V, *et al.* Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*, 106(2):477–484, 2005.

[349] Rohwer JM and Botha FC. Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data. *Biochem J*, 358(Pt 2):437–445, 2001.

[350] Rojas A, Val SD, *et al.* Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development*, 132(15):3405–3417, 2005.

[351] Saez-Rodriguez J, Simeoni L, *et al.* A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol*, 3(8):e163, 2007.

[352] Saier MH and Ramseier TM. The catabolite repressor/activator (Cra) protein of enteric bacteria. *J Bacteriol*, 178(12):3411–3417, 1996.

[353] Sauro HM, Hucka M, *et al.* Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS*, 7(4):355–372, 2003.

[354] Savageau MA. Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation. *J Theor Biol*, 26(2):215–226, 1970.

[355] Savageau MA. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature*, 252(5484):546–549, 1974.

[356] Savageau MA. Michaelis-Menten mechanism reconsidered: implications of fractal kinetics. *J Theor Biol*, 176(1):115–124, 1995.

[357] Savageau MA. Demand theory of gene regulation. I. Quantitative development of the theory. *Genetics*, 149(4):1665–1676, 1998.

[358] Savageau MA. Demand theory of gene regulation. II. Quantitative application to the lactose and maltose operons of Escherichia coli. *Genetics*, 149(4):1677–1691, 1998.

[359] Scherens B, Feller A, Vierendeels F, Messenguy F, and Dubois E. Identification of direct and indirect targets of the Gln3 and Gat1 activators by transcriptional profiling in response to nitrogen availability in the short and long term. *FEMS Yeast Res*, 6(5):777–791, 2006.

[360] Schmidt T, Friehs K, and Flaschel E. Rapid determination of plasmid copy number. *J Biotechnol*, 49(1-3):219–229, 1996.

[361] Schnell S and Turner TE. Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog Biophys Mol Biol*, 85(2-3):235–260, 2004.

[362] Schoeberl B, Eichler-Jonsson C, Gilles ED, and Müller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol*, 20(4):370–375, 2002.

[363] Schrider DR and Hahn MW. Gene copy-number polymorphism in nature. *Proc Biol Sci*, 277(1698):3213–3221, 2010.

[364] Schröder I, Darie S, and Gunsalus RP. Activation of the Escherichia coli nitrate reductase (narGHJI) operon by NarL and Fnr requires integration host factor. *J Biol Chem*, 268(2):771–774, 1993.

[365] Schuster P. Landscapes and molecular evolution. *Phys. D*, 107(2-4):351–365, 1997. ISSN 0167-2789.

[366] Schuster P. Evolution *in silico* and *in vitro*: the RNA model. *Biol Chem*, 382(9):1301–1314, 2001.

[367] Schuster P. A Testable Genotype-Phenotype Map: Modeling Evolution of RNA Molecules. In M Lässig and A Valleriani, eds., *Biological Evolution and Statistical Physics*, 56–83. Springer-Verlag, Berlin, 2002.

[368] Schuster P. Molecular Insight into the Evolution of Phenotypes. In JP Crutchfield and P Schuster, eds., *Evolutionary Dynamics – Exploring the Interplay of Accident, Selection, Neutrality, and Function*, 163–215. Oxford University Press, New York, 2003.

[369] Schuster P. *Stochastic chemical kinetics (Lecture Notes)*, 2011.

[370] Schuster P, Fontana W, Stadler PF, and Hofacker IL. From Sequences to Shapes and back: A case study in RNA secondary Structure. *Proc. Roy. Soc. (London) B*, 255:279–284, 1994.

[371] Schuster P, Sigmund K, and Wolff R. On ω-Limits for Competition Between Three Species. *SIAM Journal on Applied Mathematics*, 37:49–54, 1979.

[372] Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, and Gaul U. Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature*, 451(7178):535–540, 2008.

[373] Segel IH. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. John Wiley & sons: New York, 1993.

[374] Segre D, Vitkup D, and Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A*, 99(23):15112–15117, 2002.

[375] Sha W, Moore J, *et al.* Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. *Proc Natl Acad Sci U S A*, 100(3):975–980, 2003.

References

[376] Shea MA and Ackers GK. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J Mol Biol*, 181(2):211–230, 1985.

[377] Shen-Orr SS, Milo R, Mangan S, and Alon U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet*, 31(1):64–68, 2002.

[378] Shewmaker F, Mull L, Nakayashiki T, Masison DC, and Wickner RB. Ure2p function is enhanced by its prion domain in Saccharomyces cerevisiae. *Genetics*, 176(3):1557–1565, 2007.

[379] Shine J and Dalgarno L. Determinant of cistron specificity in bacterial ribosomes. *Nature*, 254(5495):34–38, 1975.

[380] Shlomi T, Berkman O, and Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A*, 102(21):7695–7700, 2005.

[381] Siciliano V, Menolascina F, *et al.* Construction and modelling of an inducible positive feedback loop stably integrated in a Mammalian cell-line. *PLoS Comput Biol*, 7(6):e1002074, 2011.

[382] Silvester JR. Determinants of Block Matrices. *The Mathematical Gazette*, 84:460–467, 2000.

[383] Sivakumaran S, Hariharaputran S, Mishra J, and Bhalla US. The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics*, 19(3):408–415, 2003.

[384] Smith HL. Oscillations and Multiple Steady States in a Cyclic Gene Model with Repression. *J. Math. Biol.*, 25:169–190, 1987.

[385] Smith HL. *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*. Mathematical Surveys and Monographs, Vol. 41. American Mathematical Society, Providence, RI, 1995.

[386] Smith LP, Bergmann FT, Chandran D, and Sauro HM. Antimony: a modular model definition language. *Bioinformatics*, 25(18):2452–2454, 2009.

[387] Smits WK, Hoa TT, Hamoen LW, Kuipers OP, and Dubnau D. Antirepression as a second mechanism of transcriptional activation by a minor groove binding protein. *Mol Microbiol*, 64(2):368–381, 2007.

[388] Sneppen K, Dodd IB, *et al.* A mathematical model for transcriptional interference by RNA polymerase traffic in Escherichia coli. *J Mol Biol*, 346(2):399–409, 2005.

[389] Snoep JL and Olivier BG. Java Web Simulation (JWS); a web based database of kinetic models. *Mol Biol Rep*, 29(1-2):259–263, 2002.

[390] Sprinzak D and Elowitz MB. Reconstruction of genetic circuits. *Nature*, 438(7067):443–448, 2005.

[391] Stanley KO and Miikkulainen R. A Taxonomy for Artificial Embryogeny. *Artificial Life*, 9(2):93–130, 2003.

[392] Starich MR, Wikström M, Arst HN, Clore GM, and Gronenborn AM. The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. *J Mol Biol*, 277(3):605–620, 1998.

[393] Stelling J, Klamt S, Bettenbrock K, Schuster S, and Gilles ED. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, 2002.

[394] Stelling J, Sauer U, Szallasi Z, Doyle FJ, and Doyle J. Robustness of cellular functions. *Cell*, 118(6):675–685, 2004.

[395] Stephan-Otto Attolini C. *From Molecular Systems to Simple Cells: a Study of the Genotype-Phenotype Map*. Ph.D. thesis, Univ. of Vienna, 2005.

[396] Sterner DE, Grant PA, *et al.* Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. *Mol Cell Biol*, 19(1):86–98, 1999.

[397] Stricker J, Cookson S, *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, 2008.

[398] Strogatz SH. *Nonlinear Dynamics and Chaos. With Applications to Physics, Biology, Chemistry, and Engineering*. Westview at Perseus Books Group, 1994.

[399] Swat M, Kel A, and Herzel H. Bifurcation analysis of the regulatory modules of the mammalian G1/S transition. *Bioinformatics*, 20(10):1506–1511, 2004.

[400] Talini G, Gallori E, and Maurel MC. Natural and unnatural ribozymes: back to the primordial RNA world. *Res Microbiol*, 160(7):457–465, 2009.

[401] Taylor JS and Raes J. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet*, 38:615–643, 2004.

[402] Tchuraev RN. A new method for the analysis of the dynamics of the molecular genetic control systems. I. Description of the method of generalized threshold models. *J Theor Biol*, 151(1):71–87, 1991.

[403] Thieffry D and Thomas R. Dynamical behaviour of biological regulatory networks–II. Immunity control in bacteriophage lambda. *Bull Math Biol*, 57(2):277–297, 1995.

[404] Thieffry D and Thomas R. Qualitative analysis of gene networks. *Pac Symp Biocomput*, 77–88, 1998.

[405] Thomas R. Boolean formalisation of genetic control circuits. *J. Theor. Biol.*, 42:565–583, 1973.

[406] Thomas R and D'Ari R. *Biological Feedback*. CRC Press, Boca Raton, FL, 1990.

[407] Thomas R and Kaufman M. Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior. *Chaos*, 11(1):170–179, 2001.

[408] Thomas R and Kaufman M. Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos*, 11(1):180–195, 2001.

[409] Thomas R, Thieffry D, and Kaufman M. Dynamical behaviour of biological regulatory networks–I. Biological role of feedback loops and practical use of

the concept of the loop-characteristic state. *Bull Math Biol*, 57(2):247–276, 1995.

[410] Thompson JD, Higgins DG, and Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–4680, 1994.

[411] Timney BL, Tetenbaum-Novatt J, *et al.* Simple kinetic relationships and nonspecific competition govern nuclear import rates in vivo. *J Cell Biol*, 175(4):579–593, 2006.

[412] Tsai SF, Strauss E, and Orkin SH. Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes Dev*, 5(6):919–931, 1991.

[413] Tuttle LM, Salis H, Tomshine J, and Kaznessis YN. Model-driven designs of an oscillating gene network. *Biophys J*, 89(6):3873–3883, 2005.

[414] Tyson JJ. Modeling the cell division cycle: cdc2 and cyclin interactions. *Proc Natl Acad Sci U S A*, 88(16):7328–7332, 1991.

[415] Tyson JJ, Chen KC, and Novak B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol*, 15(2):221–231, 2003.

[416] Tyson JJ and Novak B. Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions. *J Theor Biol*, 210(2):249–263, 2001.

[417] Tyson JJ and Othmer HG. The dynamics of feedback control circuits in biochemical pathways. *Progress in Theoretical Biology*, 5:1–62, 1978.

[418] Ullrich A and Flamm C. Functional Evolution of Ribozyme-Catalyzed Metabolisms in a Graph-Based Toy-Universe. In *CMSB*, 28–43. 2008.

[419] Ullrich A and Flamm C. A Sequence-to-Function Map for Ribozyme-catalyzed Metabolisms. *Advances in Artificial Life*, 5778:481, 2009.

[420] Ullrich A, Rohrschneider M, Scheuermann G, Stadler PF, and Flamm C. In silico evolution of early metabolism. *Artif Life*, 17(2):87–108, 2011.

[421] Valentin-Hansen P, Sogaard-Andersen L, and Pedersen H. A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol Microbiol*, 20(3):461–466, 1996.

[422] van Hijum SAFT, Medema MH, and Kuipers OP. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol Mol Biol Rev*, 73(3):481–509, Table of Contents, 2009.

[423] van Kampen NG. The expansion of the master equation. *Adv. Chem. Phys.*, 34:245–309, 1976.

[424] Varani G and McClain WH. The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep*, 1(1):18–23, 2000.

[425] Varma A and Palsson BO. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Nat Biotech*, 12(10):994–998, 1994.

[426] Veitia RA. A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. *Biol Rev Camb Philos Soc*, 78(1):149–170, 2003.

[427] Vilar JMG, Kueh HY, Barkai N, and Leibler S. Mechanisms of noise-resistance in genetic oscillators. *Proc Natl Acad Sci U S A*, 99(9):5988–5992, 2002.

[428] Visser D and Heijnen JJ. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metab Eng*, 5(3):164–176, 2003.

[429] Vitreschak AG, Rodionov DA, Mironov AA, and Gelfand MS. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res*, 30(14):3141–3151, 2002.

[430] Vitreschak AG, Rodionov DA, Mironov AA, and Gelfand MS. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet*, 20(1):44–50, 2004.

[431] Voit EO and Almeida J. Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics*, 20(11):1670–1681, 2004.

[432] Wade JT, Belyaeva TA, Hyde EI, and Busby SJ. A simple mechanism for co-dependence on two activators at an Escherichia coli promoter. *EMBO J*, 20(24):7160–7167, 2001.

[433] Wall ME, Hlavacek WS, and Savageau MA. Design principles for regulator gene expression in a repressible gene circuit. *J Mol Biol*, 332(4):861–876, 2003.

[434] Weissbach H and Brot N. Regulation of methionine synthesis in Escherichia coli. *Mol Microbiol*, 5(7):1593–1597, 1991.

[435] Wierling C, Herwig R, and Lehrach H. Resources, standards and tools for systems biology. *Brief Funct Genomic Proteomic*, 6(3):240–251, 2007.

[436] Winkler WC and Breaker RR. Genetic control by metabolite-binding riboswitches. *Chembiochem*, 4(10):1024–1032, 2003.

[437] Wittig U, Golebiewski M, *et al.* SABIO-RK: Integration and curation of reaction kinetics data. In U Leser, F Neumann, and B Eckman, eds., *Data Integration In The Life Sciences, Proceedings*, vol. 4075 of *Lecture Notes In Computer Science*, 94–103. Springer-Verlag Berlin, Berlin, Germany, 2006. ISBN 3-540-36593-1. ISSN 0302-9743.

[438] Wittmann DM, Krumsiek J, *et al.* Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC Syst Biol*, 3:98, 2009.

[439] Wong KH, Hynes MJ, and Davis MA. Recent advances in nitrogen regulation: a comparison between Saccharomyces cerevisiae and filamentous fungi. *Eukaryot Cell*, 7(6):917–925, 2008.

[440] Wong KH, Hynes MJ, Todd RB, and Davis MA. Deletion and overexpression of the Aspergillus nidulans GATA factor AreB reveals unexpected pleiotropy. *Microbiology*, 155(Pt 12):3868–3880, 2009.

[441] Wright MC and Joyce GF. Continuous in vitro evolution of catalytic function. *Science*, 276(5312):614–617, 1997.

[442] Wright S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the Sixth International Congress on Genetics*, 355–366. 1932.

[443] Wu X, Bayle JH, Olson D, and Levine AJ. The p53-mdm-2 autoregulatory feedback loop. *Genes Dev*, 7(7A):1126–1132, 1993.

[444] Yagil G and Yagil E. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys J*, 11(1):11–27, 1971.

[445] Yakovchuk P, Protozanova E, and Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res*, 34(2):564–574, 2006.

[446] Yamaguchi M, Biswas SK, Suzuki Y, Furukawa H, and Takeo K. Three-dimensional reconstruction of a pathogenic yeast Exophiala dermatitidis cell by freeze-substitution and serial sectioning electron microscopy. *FEMS Microbiol Lett*, 219(1):17–21, 2003.

[447] Yanofsky C. Transcription attenuation: once viewed as a novel regulatory strategy. *J Bacteriol*, 182(1):1–8, 2000.

[448] Yao G, Lee TJ, Mori S, Nevins JR, and You L. A bistable Rb-E2F switch underlies the restriction point. *Nat Cell Biol*, 10(4):476–482, 2008.

[449] Yi TM, Huang Y, Simon MI, and Doyle J. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A*, 97(9):4649–4653, 2000.

[450] Yu T and Li KC. Inference of transcriptional regulatory network by two-stage constrained space factor analysis. *Bioinformatics*, 21(21):4033–4038, 2005.

[451] Zaslaver A, Mayo AE, *et al.* Just-in-time transcription program in metabolic pathways. *Nat Genet*, 36(5):486–491, 2004.

[452] Zawel L and Reinberg D. Initiation of transcription by RNA polymerase II: a multi-step process. *Prog Nucleic Acid Res Mol Biol*, 44:67–108, 1993.

[453] Zinzen RP, Senger K, Levine M, and Papatsenko D. Computational models for neurogenic gene expression in the Drosophila embryo. *Curr Biol*, 16(13):1358–1365, 2006.

[454] Zuker M and Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, 9(1):133–148, 1981.

[455] Zurawski G, Elseviers D, Stauffer GV, and Yanofsky C. Translational control of transcription termination at the attenuator of the Escherichia coli tryptophan operon. *Proc Natl Acad Sci U S A*, 75(12):5988–5992, 1978.

Lukas Endler

Contact details

address: Saileräckergasse 49-55/2/26
1190 Wien, Austria

telephone: +43-69919667406 (mobile)
+43-1-9224745 (home)

e-mail: lukas.endler@gmx.at

date and place of birth: June, 14th 1974, Vienna, Austria

citizenship: Austrian

Education

04/2004–present   PhD studies, Inst. for Theor. Chemistry, Univ. of Vienna, supervisor: Prof. Peter Schuster
Dynamical Models of Biological Networks.

10/2002   Mag. rer. nat. in Chemistry, Univ. of Vienna, supervisor: Prof. Gustav Ammerer
G2 Specific Transcription in the Yeast Saccharomyces cerevisiae.

10/1992–04/2002   studies of biochemistry at the University of Vienna

1992   Matura (high school graduation), BGXIX

Professional Experience

05/2008–06/2011   Senior scientific database curator for BioModels database
Computational Neurobiology, European Bioinformatics Institute, Cambridge, UK

03/2004–04/2008   Graduate student at the Inst. for theor. Chemistry, Univ. Vienna, project WWTF-MA05: Inverse Methods in Biology and Chemistry

10/2002–12/2003   Research assistant at the Department of Medical and Chem-

ical Laboratory Diagnostics, University of Vienna Medical School

07/1998–09/1998   Graduate research scholar at the Hormone Receptor Laboratory, Prof. J. Wittliff, University of Louisville, Kentucky, USA

Military Service

10/2001–05/2002   Clinical laboratory at the Vienna Military Hospital

Language Skills: German (native), English (fluent), French (basic)

Computer Skills

Operating systems: UNIX (Linux, MacOsX), Windows XP

Programming: Perl, Python, shell scripting, R, Mathematica

Text processing: OpenOffice, LATEX

Publications

[1] L. Endler, MI. Stefan, SJ. Edelstein, N. Le Novère. Using chemical kinetics to model neuronal signalling pathways. in *Computational Systems Neurobiology*, N. Le Novère (Ed.), Springer 2012.

[2] R. Henkel, L. Endler, A. Peters, N. Le Novère, and D. Waltemath. Ranked retrieval of Computational Biology models. *BMC Bioinformatics.* 11:423, 2010.

[3] C. Li, M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, L. Li, E. He, A. Henry, MI. Stefan, JL. Snoep, M. Hucka, N. Le Novère, and C. Laibe. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst Biol.*, 4:92, 2010.

[4] V. Chelliah, L. Endler, N. Juty, C. Laibe, C. Li, N. Rodriguez, and N. Le Novère. Data Integration and Semantic Enrichment of Systems Biology Models and Simulations. *Data Integration in the Life Sciences*, 5647:515, 2009.

[5] L. Endler, N. Rodriguez, N. Juty, V. Chelliah, C. Laibe, C. Li, and N. Le Novère. Designing and encoding models for synthetic biology. *J R Soc Interface*, 6 Suppl 4:S405S417, Aug 2009.

[6] A. Dräger, H. Planatscher, D. Motsou Wouamba, A. Schröder, Michael Hucka, L. Endler, M. Golebiewski, W. Müller, and A. Zell. SBML2LATEX: conversion of SBML files into human-readable reports. *Bioinformatics*, 25(11):1455 1456, Jun 2009.

[7] C. Flamm, L. Endler, S. Müller, S. Widder, and P. Schuster. A minimal and self-consistent in silico cell model based on macromolecular interactions. *Philos Trans R Soc Lond B Biol Sci*, 362(1486):18311839, Oct 2007.

[8] H. Abrahamian, G. Endler, M. Exner, H. Mauler, M. Raith, L. Endler, H. Rumpold, M. Gerdov, C. Mannhalter, R. Prager, K. Irsigler, and O. F. Wagner. Association of low-grade inflammation with nephropathy in type 2 diabetic patients: role of elevated CRP-levels and 2 different gene-polymorphisms of proinflammatory cytokines. *Exp Clin Endocrinol Diabetes*, 115(1):3841, Jan 2007.

[9] S. Müller, J. Hofbauer, L. Endler, C. Flamm, S. Widder, and P. Schuster. A generalized model of the repressilator. *J Math Biol*, 53(6):905937, Dec 2006.

[10] M. Funk, G. Endler, M. Exner, R. Marculescu, L. Endler, H. Abrahamian, H. Mauler, A. Grimm, M. Raith, C. Mannhalter, R. Prager, K. Irsigler, and OF. Wagner. PAI-1 4G/5G insertion/deletion promoter polymorphism and microvascular complications in type 2 diabetes mellitus. *Wiener Klinische Wochenschrift*, 117(19-20):707710, Oct 2005.

[11] G. Endler, M. Exner, M. Raith, R. Marculescu, C. Mannhalter, L. Endler, J. Wojta, K. Huber, and OF. Wagner. The E-selectin S128R polymorphism is not a risk factor for coronary artery disease in patients with diabetes mellitus type 2.*Thromb Res*, 112(1-2):4750, 2003.

[12] M. Koranda, A. Schleiffer, L. Endler, and G. Ammerer. Forkhead-like transcription factors recruit Ndd1 to the chromatin of G2/M-specific promoters. *Nature*, 406(6791):9498, Jul 2000.