

**The Dynamics of
the XlnR Regulon
of *Aspergillus niger*:
a Systems Biology Approach**

Jimmy Omony

Thesis committee

Promotor

Prof.dr.ir. G. van Straten
Professor Emeritus of Systems and Control
Wageningen University

Co-promotors

Dr.ir. A.J.B. van Boxtel
Associate professor, Systems and Control Group
Wageningen University

Dr.ir. L.H. de Graaff
Associate professor, Laboratory of Systems and Synthetic Biology
Wageningen University

Other members

Prof.dr.ir. J. Molenaar
Wageningen University

Prof.dr. R. Breitling
University of Glasgow, UK and University of Groningen

Prof.dr. B. Teusink
VU University Amsterdam

Dr.ir. J.A. Roubos
DSM Biotechnology Center, Delft

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

**The Dynamics of
the XlnR Regulon
of *Aspergillus niger*:
a Systems Biology Approach**

Jimmy Omony

Thesis

submitted in fulfilment of the requirement of the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof.dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on 19th December 2012
at 11 : 00 a.m in the Aula.

Jimmy Omony

The Dynamics of the XlnR Regulon of *Aspergillus niger*: a Systems Biology Approach, 158 pages.

PhD thesis, Wageningen University, Wageningen, The Netherlands (2012).
With references, with summaries in English and Dutch.

ISBN: 978-94-6173-417-4

Dedications

I am very grateful for the support of my lovely wife Mariëlle Omony-Miedema during my PhD research. I suppose it has been difficult putting up with my divided attention between studies and family time. The good news is, no situation is permanent. Long story short, now it is time for celebrations! To our son Lucas Joshua Oloo Omony, we are very proud of you. You are still too young to grasp the content of this thesis but with time we are confident you will come to understand what it is all about.

"If you don't make mistakes, you're not working on hard enough problems. And that's a big mistake." **Frank Wilczek, American physicist (1951 -)**

Contents

1	General background	1
1.1	Introduction	2
1.2	The <i>Aspergillus niger</i> fungi	2
1.2.1	The XlnR regulon of <i>A. niger</i>	3
1.2.2	Carbon catabolite repression by CreA	4
1.3	On genetic network reconstruction	4
1.4	General challenges in network reconstruction	4
1.5	Problem statement and research goals	7
1.6	Thesis outline	8
2	Modeling and analysis of the dynamic behavior of the XlnR regulon in <i>Aspergillus niger</i>	11
2.1	Background	13
2.2	Methods	14
2.2.1	Regulation mechanism for the XlnR regulon	14
2.2.2	Transcription model	15
2.2.3	Translation model	16
2.2.4	System stability	17
2.2.5	Feedback in the network	18
2.3	Results	22
2.3.1	System specification	22
2.3.2	Stability and response analysis - without feedback	22
2.3.3	Feedback in the network	24
2.4	Discussion	28
2.5	Conclusions	31
3	Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in <i>Aspergillus niger</i>	33
3.1	Introduction	35
3.2	Methods	36
3.2.1	Target gene transcription in the XlnR regulon	36
3.2.2	Parameter sensitivity functions	38

Contents

3.2.3	Experimental evaluation and parameter estimation	39
3.2.4	Data generation and evaluation procedure	41
3.2.5	Experimental strategy - double D-xylose pulse	42
3.2.6	Evaluation of parameter sensitivities	43
3.2.7	Software used	43
3.3	Results	43
3.3.1	Low D-xylose pulse experiments	43
3.3.2	Experiments with varying pulse strength	47
3.3.3	Single versus double D-xylose pulses	48
3.3.4	Number of samples and inter-sampling time	49
3.3.5	Comparing criteria for experimental design	50
3.4	Discussion and conclusions	51
4	D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in <i>A. niger</i>	55
4.1	Introduction	57
4.2	Materials and methods	60
4.2.1	Strains and growth conditions	60
4.2.2	RNA extraction and reverse transcription (RT)	60
4.2.3	qPCR analysis	61
4.2.4	Statistical analyses of qPCR data	61
4.2.5	HPLC analysis	62
4.3	Results	62
4.3.1	Response of xylan backbone-degrading enzyme expression to different D-xylose concentrations	62
4.3.2	Expression of accessory enzyme-encoding genes is mostly favored using a high D-xylose concentration	68
4.3.3	Potential of D-xylose to induce expression of cellulase-encoding genes	69
4.3.4	High D-xylose concentrations are preferable for inducing the gene expression of enzymes in the pentose metabolic pathway	69
4.3.5	Absence of a functional CreA positively influences expression of genes encoding xylan-degrading enzymes independent of D-xylose concentration	70
4.3.6	The <i>xlnR</i> gene is constitutively expressed independent of CreA and D-xylose concentration	70
4.4	Discussion	72
5	Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of <i>Aspergillus niger</i>	75
5.1	Background	77
5.2	Materials and methods	77
5.2.1	Description of experiments	77
5.2.2	Data analysis criteria	78

5.2.3	D-xylose uptake in time	79
5.2.4	The CreA effect ratio (CER) statistic	80
5.3	Results	80
5.3.1	Comparing transcription dynamics	80
5.3.2	CreA effect ratio (CER)	83
5.4	Discussion and conclusion	85
6	On the transcription dynamics in the XlnR regulon of <i>A. niger</i> to D-xylose triggers	89
6.1	Introduction	91
6.2	Materials and methods	92
6.2.1	Experimental procedure	92
6.2.2	List of genes used from the XlnR regulon	93
6.3	Results	93
6.3.1	Transcription measurements	93
6.4	Mathematical modeling and discussion	96
6.4.1	Procedure for model fit to experimental data	99
6.4.2	Model evaluation	100
6.5	Conclusion	103
6.6	Appendix	105
6.6.1	Nomenclature - variables and parameters	105
7	Retrospectives, perspectives and contributions	109
7.1	Retrospectives and perspectives	110
7.1.1	Introduction	110
7.1.2	Modeling and simulation of the XlnR regulon dynamics	110
7.1.3	On strategic design of time course experiments	111
7.1.4	On the qualitative and quantitative effects of CreA on transcription	112
7.1.5	Modeling and validation of transcription dynamics	113
7.2	My take on biological network reconstruction	113
	Appendices	116
A	State-of-the-art in network reconstruction	117
A.1	Introduction	117
A.2	The (Probabilistic) Bayesian Network formalism	117
A.3	Regression-based methods for network identification	118
A.4	The Boolean Network formalism	118
A.5	The Ordinary Differential Equation (ODE) formalism	119
A.5.1	Variants of the ODE formalism	119
A.5.2	Mathematical representation	120

Contents

B Modeling cycle, abbreviations, acronyms and definitions	121
B.1 The modeling cycle in network reconstruction	121
B.2 List of abbreviations and acronyms	122
B.3 Some useful definitions	123
Bibliography	125
Summary	141
Samenvatting	145
Acknowledgements	149
About the author - English	151
Over de auteur - Nederlands	152
Completed training activities	157

Chapter 1

General background

1.1 Introduction

The link between systems theory and its applications to biology is an item of great research interest today. For long biologists have been trying to find solutions to challenging questions. The answers to some scientifically interesting questions, for instance in biological signaling networks (e.g. the tumor suppressor protein p53 which regulates the activity of hundreds of genes involved in cell growth and death [1, 2, 3, 4]); in *Arabidopsis* [5, 6, 7]; and in the *E. coli* bacteria [8, 9, 10] and in the species of yeast *S. cerevisiae* [11, 12], lies in the inference of biological networks. Advancements in the use of mathematical and statistical methods has been much to the relief of the scientists working on biological networks.

The items of interest in network inference include, e.g. i) network topology and dynamics, ii) stability analysis after network perturbations, and iii) real-life applicability of the results. Some studies on biological networks focus purely on theoretical approaches to network analysis, and others focus on the applications of proposed network inference method. Often these methods are either validated on experimental data from a specific model organism, or validated using synthetic data. The common factor here is that all these studies are motivated by some real life application, as is the case in this thesis with the xylanolytic activator (XlnR) regulon of the fungus *A. niger*.

The examination of a biological network can be achieved through modeling. Modeling is a guide within the reasoning processing for a given system. There are two main reasons for modeling biological networks namely: i) the need to gain insight into the network regulation mechanisms, and ii) the examination of the resulting model attributes. Modeling and identification of the dynamic interaction between the genes is required to advance the utilization of *A. niger*. Knowledge acquisition of the time behavior of the subnetwork XlnR offers significant perspectives for more efficient utilization of this fungus.

1.2 The *Aspergillus niger* fungi

The filamentous fungi *A. niger* has many different applications, for example the production of organic acids such as citric acid (E330) and gluconic acid (E574). Numerous enzymes for commercial use are produced from this organism as well. Industrial applications for *A. niger* include pulp and paper, textile, detergents, food and beverages, and agriculture [13, 14, 15, 16, 17]. The vast industrial uses of *A. niger* have drawn attention to the need to understand the carbohydrate modifying network in *A. niger* - particularly with the goal to improve enzyme substrate utilization and production.

The XlnR regulon of *A. niger* was used as the organism of interest in this thesis. *A. niger* produces a wide spectrum of polysaccharide-hydrolytic enzymes that are responsible for plant biomass degradation. The *xlnR* gene controls the activity of arabinolytic enzymes. Regulation of cellulases- and hemicellulase-encoding genes in *A. niger* and *Hypocrea jecorina* (*Trichoderma reesei*) was studied

by Stricker et al. [18]. The XlnR protein is the transcriptional regulator of the xylanase encoding gene, and controls the transcription of approximately 20 to 40 genes encoding hemicellulases and cellulases.

1.2.1 The XlnR regulon of *A. niger*

Gene	Enzyme function	Accession no.
<i>xlnR</i>	Transcriptional activator of hemi-cellulases and cellulases	AJ001909
<i>xlnB</i>	Endo-xylanase B	D38071
<i>xlnC</i>	Endo-xylanase C	C5J411
<i>xlnD</i>	β -Xylosidase	Z84377
<i>axeA</i>	Acetyl xylanesterase	A22880
<i>axhA</i>	Arabinoxylan hydrolase	Z780811
<i>abfB</i>	Arabinofuranosidase B	X74777
<i>aguA</i>	α -Glucuronidase	Y15405
<i>faeA</i>	Ferulic acid esterase A	Y09330
<i>estA</i>	Esterase A	AY456379
<i>lacA</i>	β -Galactosidase	AJ564428
<i>aglB</i>	α -Galactosidase	Y18586
<i>cbhA</i>	Cellobiohydrolase A	AF156268
<i>cbhB</i>	Cellobiohydrolase B	AF156269
<i>eglA</i>	Endo-glucanase A	AJ224451
<i>eglB</i>	Endo-glucanase B	AJ224452
<i>eglC</i>	Endo-glucanase C	AY040839
<i>bglA</i>	β -Glucosidase	AB003470
<i>xyrA</i>	D-xylose reductase	AF219625
<i>xkiA</i>	D-xylose kinase	AJ305311
<i>xdhA</i>	D-xylose dehydrogenase	AJ854041
<i>ladA</i>	L-Arabitol dehydrogenase	AJ854040
<i>talB</i>	Transaldolase B	EG11556

Table 1.1: **Genes in the XlnR regulon.** An overview of genes in the XlnR regulon, their corresponding enzyme function and accession numbers.

The XlnR regulon consists of target genes as shown in Table 1.1. The target genes are categorized into cellulases and hemicellulase encoding genes. Some of the enzymes produced by these genes degrade plant cell walls - particularly the degradation of lignocellulose. The transcriptional activator for the XlnR regulon is the xylanolytic activator XlnR. Apart from the resulting role of the XlnR protein, the activity state of the XlnR regulon is also influenced by the presence of the carbon catabolite repressor protein (CreA). The target genes are expressed under certain conditions, e.g. with a low supply of carbon nutrient in the form of D-xylose while a high D-xylose concentration reduces target gene transcription.

1.2.2 Carbon catabolite repression by CreA

The repressing effect of CreA has been demonstrated in the *A. niger* species [19]. The carbon catabolite repressing effects of CreA has been reported for *A. nidulans* [20, 21] following a D-xylose induction. The CreA protein contains a stretch of 42 amino acids which is identical in *A. niger* and *A. nidulans* [22]. Previously, the effect of the CreA protein was only qualitatively described. Unfortunately, qualitative descriptions are insufficient for effective dynamics modeling. In this work, experiments were performed to quantify the effect of CreA on transcription in the XlnR regulon. The regulation mechanisms of the XlnR regulon are studied in the subsequent chapters of this thesis.

1.3 On genetic network reconstruction

The history of network reconstruction dates back to the 1960s following the work of Erdős and Rényi [23], who developed the mathematical theory for networks where nodes are randomly connected by edges. Today, unraveling complex biological networks is key to understanding disease traits and cancer research, and it has numerous industrial and agricultural applications. In this thesis, mathematical modeling is used to study the xylanolytic activator, XlnR regulon of *A. niger*. The pioneering work of reverse engineering¹ biological networks from experimental gene-*omic* data sets can be dated as far back as the 1990s [26, 27, 28]. Since then there have been many studies on biological networks. The inference of biological networks dates back a couple of decades by Kauffman [29]. Over the years various models and inference methods have been proposed and successfully evaluated on real experimental data. This consolidated effort between researchers lead to the field of Systems Biology².

Kitano [31] summarized the required tasks for understanding biological systems: i) system structure identification (analysis of network topology), ii) system behavior analysis (analysis of network dynamics), iii) system design, and iv) system control. This thesis focuses on aspect, ii). Prior to the research described in this thesis nothing had been done in modeling the dynamics of the XlnR regulon. This work seeks to find and evaluate models that describe transcription profiles for genes in the XlnR regulon using *in silico* and real-time qPCR data sets.

1.4 General challenges in network reconstruction

In literature, thousands of papers have been published that either: i) present a review of existing methods, or ii) present and/or validate methods on specific

¹Reverse engineering genetic networks is defined as: i) the process of identifying gene interactions from experimental data through computational analysis [24], ii) the process of analyzing a subject system to identify the system's components and their inter-relationships, and to create representations of the system in another form at higher levels of abstraction [25].

²See the article by Breitling [30] for an overview of the various definitions of Systems Biology.

1.4. General challenges in network reconstruction

biological networks. All these methods have pros and cons [32, 33] many of which are summarized in this subsection with additional information in Appendix A. To put this into perspective, Figure 1.1 shows the major steps in network reconstruction.

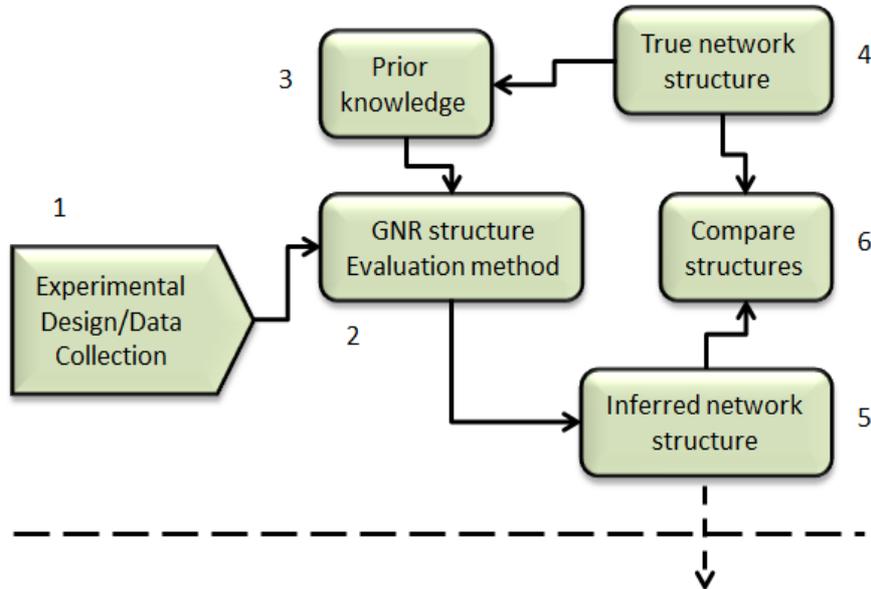


Figure 1.1: **Key steps in genetic network reconstruction.** This scheme represents the major constituents of the study of biological networks. The "Prior knowledge" constitutes literature information from scientific publications, biological databases, expert knowledge on the subject. The performance of a genetic network reconstruction method is judged by how well the "Inferred network" structure matches the "True network" structure. Starting with a given data set and "Prior knowledge" of a network, often the goal is to infer knowledge about the "true" network structure.

The reverse engineering process starts from planning experiments and data collection to test specific hypotheses and research objectives (Figure 1.1, step 1). This is a very important step, though often overlooked. A well designed experiment ensures informative data acquisition, which is necessary for good network identification (step 2, Figure 1.1). The main problem with network reconstruction is that often the "true network structure" depicted in Figure 1.1 (in step 4) is unknown. In these cases one is forced to rely on an experts' knowledge inferred from manifestations from the observation space as a starting point for the network reconstruction. Prior knowledge is also helpful for planning of experiments. For *in silico* studies, the reverse engineering process is based on simulations by setting up a hypothetical network structure to mimic the "true" network structure, as well as by setting up a control experiment.

For well-studied biological networks, consolidated information on regulation mechanisms can be found in literature and biological databases. Much is known on the regulation mechanisms in the XlnR regulon. Acquisition of such information requires testing many models including various hypotheses using data sets from carefully designed experiments - a process that is associated to

Chapter 1. General background

step 2 in Figure 1.1. These processes are time-consuming, financially expensive and involves extensive research. Hypothesis testing and model validation enables us to find the best fitting models to a given data set. More information on the details involved in network reconstruction is given in the modeling cycle (Appendix B).

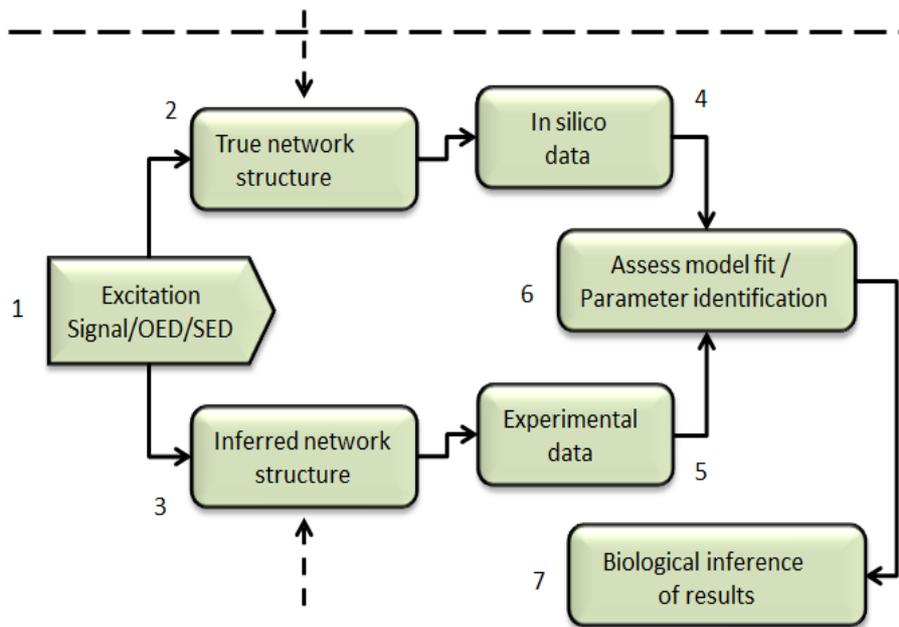


Figure 1.2: **Scheme for network identification and experimental design.** Here OED refers to Optimal Experimental Design. The term OED can be equivalently replaced by Strategic Experimental Design (SED). The scheme shows the association between the components considered for studying biological networks.

The scheme in Figure 1.2 shows the steps (1 to 7) involved in network identification and experimental design. The goal of identification is to compare how well the true network structure approximates the inferred network structure (Figure 1.2, step 6). Often, approximations are sufficient for studying network dynamics if the discrepancy between the “true” and approximated network structures is low. A comparison of these discrepancies is done by assessing the existence of corresponding edges³ between the nodes⁴ in the “true” and “identified” network structures. Here, the nodes represent the genes and the edges represent the regulatory signal from one gene to the other. The last step in the reverse engineering process is the “Biological inference of results” (Figure 1.2, step 7). This step involves drawing biologically meaningful inferences from the model output and identified system parameters.

³The **edges** in a network may represent the regulation of transcription, regulation of translation, signal transduction, catabolic relationships like enzymes to substrates and the transformation of a chemical reaction like substrates to products.

⁴A **node** is a component of a graph which represents an object with a specific relationship to other objects (in this case other nodes) in a graph.

1.5 Problem statement and research goals

The challenge we set out to address at the start of this project was to develop mathematical tools for the inference of biological networks and test them on the XlnR regulon. It is challenging to obtain good network inference results from short time course experiments. The ability to measure gene expression in time offers the possibility of monitoring molecular processes, and presents an opportunity for model based identification. In this thesis, an attempt is made to improve the design strategies for performing time course experiments, particularly with the goal of getting more information from a minimum number of data samples. The problem of how best to excite a network to yield a desired response was also of interest.

Real-time qPCR experiments were performed using genes from the XlnR regulon (see Table 1.1 for gene list, and Figure 1.3 for the experimental schemes). The experiments were performed with the CreA wild type and mutant strain of the *A. niger* fungi.

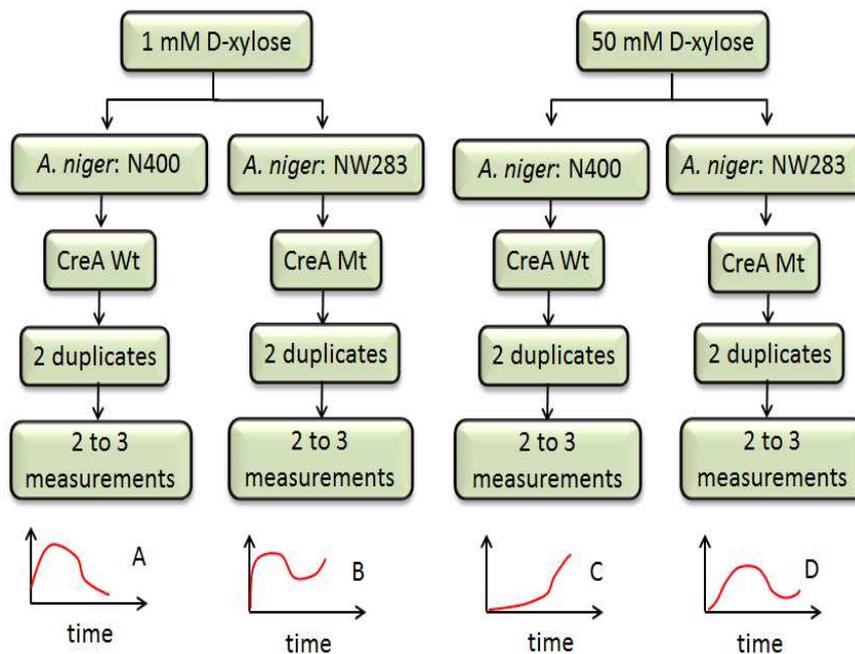


Figure 1.3: The experimental scheme and strains used to study the time course transcription in XlnR regulon. From this scheme various research questions are investigated with respect to the observed data in the transcription dynamics (plots A-D), with the time and transcription values on the horizontal and vertical axis, respectively. Within each duplicate, a number of replicate measurements were made and the data values averaged at each time point. The terms Wt and Mt refer to the wild type and mutant strains, respectively.

The main objective of this work was to develop methodologies for quantitative inference of biological networks with respect to: i) improvements of genetic network identification by perturbation studies, and ii) efficient identification of quantitative relationships from limited data sets obtained from transcriptomics.

Chapter 1. General background

The research questions are:

1. Postulate a model structure that describes the XlnR regulon dynamics. What is the influence of a positive and negative feedback loops, time delay, and promoter activity on the network dynamics?
2. What are the best strategies for perturbing a genetic network to improve the identification results?
3. In what ways can we get meaningful information from transcriptome data with large sampling intervals?
4. Are there any differences within and between the responses of the different groups of enzymes over time? What are the biological reasons behind these variations?
5. To what extent does transcription vary between the wild type (Wt) and mutant (Mt) strains after induction with D-xylose?
6. For a given strain of *A. niger*, are there deviations in transcription following induction with low and high D-xylose concentration?
7. What is the quantitative effect of CreA on transcription in the XlnR regulon?
8. What are the most plausible regulation mechanisms for the XlnR regulon? Using experimental data, identify and validate a model structure that best describes the transcription dynamics of the XlnR regulon.

1.6 Thesis outline

Chapter 2. Here, studies of network dynamics were performed in which the potential effects of positive and negative feedback, response stability, transcription and protein expression are analyzed. Conditions for the existence of oscillatory behavior are established and their biological implications explained. Through systems dynamics analysis, the benefits of using nonlinear kinetic models, e.g. Hill functions in modeling the XlnR regulon is evident. Conditions on parameters for the existence of oscillatory behavior in gene responses and protein abundance were analyzed. Research question 1 is addressed in this chapter.

Chapter 3. Here, parameter estimation was performed using *in silico* time course experiments. The responses of the profiled genes to various stimulus strength, e.g. 1 mM and 50 mM D-xylose, were studied. In these experiments strategic input designs were studied to find the best pulsing strategies for the most information gain in the XlnR network identification. In the presence of a carbon source (sorbitol), induction of the XlnR regulon results in consumption the D-xylose. The dynamic model for the

XlnR regulon was used as a basis for the experimental design. Parameter sensitivity analysis, the Fisher Information Matrix (FIM) and other design attributes such as varying the data sampling scheme and network excitation signal strength were also crucial in determination of the most optimal experimental design strategy. The focus was mainly put on using strategic experimental designs to improve the information content in transcription data sets with respect to the network identification. This chapter addresses research questions 2 and 3.

Chapter 4. The biological explanations to the observed experimental transcription profiles are provided in this chapter. It contains both details of the wet-lab qPCR experimental process and a comparative analysis of transcription of the target genes in the XlnR regulon. Here, qualitative roles of CreA were studied. An assessment of how the absence of functional CreA positively influences expression of genes encoding xylan-degrading enzymes is made. By using low and high D-xylose induction concentrations, gene expression of enzymes in the pentose metabolic pathway were determined. The potential for D-xylose to induce cellulase coding genes was investigated. The work in this chapter answers research question 4.

Chapter 5. In this chapter, the quantitative effect of CreA on target gene regulation from the XlnR regulon was studied. Two strains of CreA (the wild-type and mutant strains) were used. The time course experiments involved triggering transcription with a low (1 mM) and high (50 mM) D-xylose concentration. The target genes were assessed for differential transcription. The research questions were formulated to enable examination of the extent to which the mutated strains affected transcription. By doing the above we provide solutions to the research questions 5 to 7.

Chapter 6. This chapter deals with a posterior modeling of the regulation mechanisms and classification of the transcription profiles for the genes in the XlnR regulon. The models used in Chapters 2 and 3 were modified and evaluated on transcription experimental data. Extra dynamics for CreA and other Unknown Transcription Factors (UTFs) were used to improve the overall transcription model. The influence of the positive feedback loops on target gene transcription was assessed. The modeling and analysis help in the examination of appropriate model structures for explaining the transcription patterns. In this chapter we provide answers to the research question 8.

Chapter 7. Here, a retrospective view on the research questions and achievements of this thesis are provided. Perspectives into the challenges arising from the work in this thesis are discussed, including items for future studies, and recommendations are also discussed. This chapter also contains my opinion on the inference of biological networks as a whole. An objective assessment and advise on the do's and don't when it comes to network inference are also provided.

**Modeling and analysis of the
dynamic behavior of the XlnR
regulon in *Aspergillus niger***

Modeling and analysis of the dynamic behavior of the XlnR regulon in *Aspergillus niger*

Abstract.

Background: In this paper the dynamics of the transcription-translation system for XlnR regulon in *Aspergillus niger* is modeled. The model is based on Hill regulation functions and uses ordinary differential equations. The network response to a trigger of D-xylose is considered and stability analysis is performed. The activating, repressive feedback, and the combined effect of the two feedbacks on the network behavior are analyzed.

Results: Simulation and systems analysis showed significant influence of activating and repressing feedback on metabolite expression profiles. The dynamics of the D-xylose input function has an important effect on the profiles of the individual metabolite concentrations. Variation of the time delay in the feedback loop has no significant effect on the pattern of the response. The stability and existence of oscillatory behavior depends on which proteins are involved in the feedback loop.

Conclusions: The dynamics in the regulation properties of the network are dictated mainly by the transcription and translation degradation rate parameters, and by the D-xylose consumption profile. This holds true with and without feedback in the network. Feedback was found to significantly influence the expression dynamics of genes and proteins. Feedback increases the metabolite abundance, changes the steady state values, alters the time trajectories and affects the response oscillatory behavior and stability conditions. The modeling approach provides insight into network behavioral dynamics particularly for small-sized networks. The analysis of the network dynamics has provided useful information for experimental design for future *in vitro* experimental work.

Keywords: Perturbation, *Aspergillus niger*, XlnR regulon, dynamic response, stability, feedback.

This chapter is based on:

J. Omony, L.H. de Graaff, G. van Straten and A.J.B. van Boxtel. Modeling and analysis of the dynamic behavior of the XlnR regulon in *Aspergillus niger*. *BMC Systems Biology*, 5(Suppl 1): S14, 2011.

2.1 Background

The filamentous fungus *Aspergillus niger* is an important organism in the production of enzymes and precursors for the food and chemical industries. Industrial citric acid production by *A. niger* represents one of the most efficient, highest yield bio-processes in use by industry. The xylanolytic activator gene *xlnR* is a main controlling gene in the XlnR regulon of *A. niger*.

The XlnR regulon is activated by D-xylose in the culturing media [34]. The current description of this system is based on static interpretation of the system. Experiments [35] showed, however, that the expression of genes in the XlnR regulon is dynamic. Therefore, to advance the application of *A. niger* by better understanding of the XlnR regulon, the dynamics of the regulon needs to be quantified. For this purpose time course experiments are scheduled. However, planning of the experiments is improved by quantifying the behavior by a simulation and analysis study prior to the experiments.

For the XlnR regulon, literature information on the network structure was used as a basis for the simulations. To our knowledge, currently very little has been done on modeling the dynamics of the XlnR regulon and also on time course profiling of the genes that constitute the XlnR regulon in *A. niger*.

The challenge with genetic network modeling lies with determining a specific equation formalism to represent the network structure. One of the suggested strategies of modeling using differential equations is to fix the form of the equation [36, 37]. Prior knowledge on the network structure is essential to develop a quantitative model [38]. The descriptive information on the XlnR regulon [34] enables us to hypothesize models for the interaction between the different network components.

Generally, in the study of biological networks, positive feedback (PFB), negative feedback (NFB) [39, 40, 41, 42], feedforward loops and time delay [43] have been shown to be influential. NFB loops cause oscillatory behavior if the signal propagation around the feedback loop is low and PFBs can lead to a bistability behavior [44]. Overall, feedback plays an important part in biological networks by allowing the cell to adjust to the repertoire of functional proteins to current needs. Other examples of biological systems in which the effect of feedback and time delay were extensively studied can be found in the developmental regulator *Hes1*, which inhibits its own transcription [45, 46] or in the NFB loops in the p53 response [47, 48].

Bliss et al. [49] investigated conditions on parameters that ensure stability of the unique steady states in an operon using differential equations modeling. These authors then chose parameters that allowed the model to describe the tryptophan operon of *E. coli*. Their investigations focus on network stability in steady state. They showed that with parameters corresponding to a mutant with reduced repression, stability conditions were violated leading to oscillations. The analysis of the condition on parameters that ensure steady state stability also lead to insight into direct repression of a gene by its own product [49].

Chapter 2. Modeling and analysis of the dynamic behavior of the XlnR regulon in *Aspergillus niger*

In computational systems biology, numerous studies have been done on genetic network reconstruction using time course data but little attention has been given to understanding the network dynamics. It is crucial to understand or at worst have a fuzzy idea of a biological network dynamics if one is to gain deeper insight into the biological network dynamics and functionality.

This paper concerns the analysis of the network dynamic behavior, the effect of feedback loops and the conditions under which oscillatory responses in metabolite expression may be exhibited are investigated. Modeling of the XlnR regulon is explored by using nonlinear differential equations and Hill functions for the transcription and linear reaction kinetics for the translation process. To ensure that detailed aspects of the transcription-translation model formalism are captured, some assumptions are incorporated in the modeling. *In silico* perturbation experiments were performed by triggering the genetic network at steady state. The advantages of using ordinary differential equations (ODEs) are vast since they are capable of modeling degradation effects and causal effects in a network [50].

Applications of dynamical systems in modeling transcription regulatory networks can be found in [33, 51, 52, 53]. Many of these studies used the continuous-time domain to model gene expression as biochemical processes using in ODEs. The modeling and analysis aims to identify which factors determine the dynamics to aid and guide future time course experimental studies. In our work, we highlight the need to understand the dynamics of biological networks with the hypothesis that modeling and experimentation should go hand in hand.

2.2 Methods

2.2.1 Regulation mechanism for the XlnR regulon

In the model organism *Aspergillus niger*, transcription of genes encoding xylanolytic and cellulolytic enzymes take place [34]. Activation enables the degradation of the cellulose and hemicellulose from plant cell walls. XlnR is a zinc binuclear cluster protein consisting of about 875 amino acids. It is suspected that XlnR binds as a monomer [34]. The *xlnR* gene is induced in the presence of D-xylose in the culturing media and repressed in the presence of the carbon catabolic repressor, CreA [19].

Gene regulation can take place at different stages of the central dogma of molecular biology (DNA \rightarrow RNA \rightarrow Protein). These stages include among others transcription, translation and post-translational modifications (PTMs) of the associated protein. In Figure 2.1 a scheme of the activities in the XlnR regulon is given. The *xlnR* gene is induced by D-xylose. At induction the *xlnR* gene produces messenger RNA (mRNA) which is translated in proteins. These proteins then activate the target genes (TG). For the XlnR regulon, the number of target genes are estimated to be in the order of 20 to 40. In Figure 2.1 all target genes are represented by TG. After transcription and translation of the target genes,

the target proteins (TP) are obtained. Protein from PTMs can be involved in the regulation of the *xlnR* gene through a feedback loop. At each step in transcription and translation mRNA and proteins can be degraded and/or used for other processes (D1–D4).

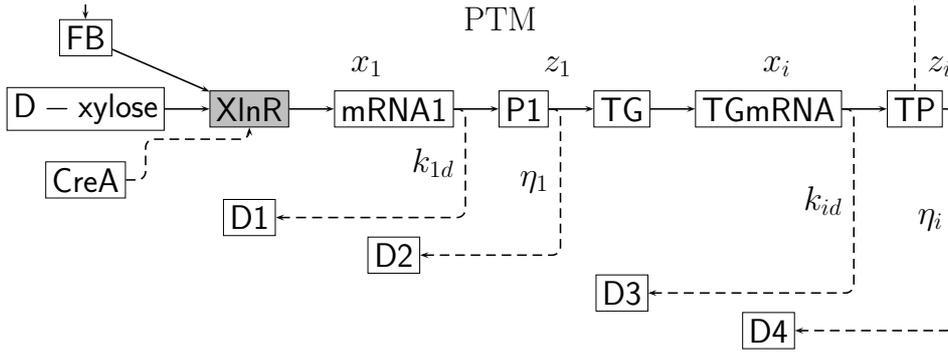


Figure 2.1: **The XlnR regulon scheme.** The XlnR regulon induced by D-xylose in the presence or absence of CreA. The representations P1 and TP are the proteins from the *xlnR* gene and target genes, respectively. The terms mRNA1 and TGmRNA represent the transcription products from the *xlnR* gene and target genes, respectively. FB represents the feedback protein in which post-translational modifications might take place.

2.2.2 Transcription model

Commonly, hyperbolic functions and the sigmoid class of functions are used to represent the kinetics of gene regulation [54]. Such functions mimic the non-linearity in gene regulation, by assuming that a critical amount of protein has to be accumulated before a gene can be considered regulated or repressed. The most common form of function used for modeling gene transcription is the Hill function [55, 56].

Let the vector $\mathbf{z} = [z_1, \dots, z_n]^T$ represent the concentrations of the translated proteins corresponding to the genes $1, \dots, n$; where n is the number of genes involved. Throughout this work, the notation, e.g. z_i is used to represent the time dependent variable $z_i(t)$ (where $t \in [0, \infty)$). Then the activating and repressing functions are given by

$$\Psi(z_i, \theta_i) = \begin{cases} \psi^+(z_i, \theta_i) = \frac{z_i^h}{\theta_i^h + z_i^h} & \text{Activator.} \\ \psi^-(z_i, \theta_i) = \frac{\theta_i^h}{\theta_i^h + z_i^h} & \text{Repressor.} \end{cases} \quad (2.1)$$

where $\psi^-(z_i, \theta_i) = 1 - \psi^+(z_i, \theta_i)$, θ_i is the gene specific half-saturation parameter and the positive number h represents the Hill coefficient. The regulation mechanism for each target gene i is captured by the function $\Psi(z_i, \theta_i)$ in (2.1).

According to Hasper et al. [57] there is evidence that although most zinc bin-

Chapter 2. Modeling and analysis of the dynamic behavior of the XlnR regulon in *Aspergillus niger*

uclear cluster proteins bind as a dimer, it seems that XlnR binds as a monomer - therefore, a Hill coefficient with $h = 1$ is used. Given the availability of structural prior knowledge and that the master regulator activates the target genes, the nonlinear system that models the transcription process is given by

$$\Sigma_{\text{nls}} = \begin{cases} \dot{x}_1 = \rho_1 - k_{1d}x_1 + b_1u_1 \\ \dot{x}_2 = \rho_2 + k_{2s}\frac{k_{21}z_1}{1 + k_{21}z_1} - k_{2d}x_2 \\ \vdots \\ \dot{x}_n = \rho_n + k_{ns}\frac{k_{n1}z_1}{1 + k_{n1}z_1} - k_{nd}x_n \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \quad (2.2)$$

where $k_{i1} = 1/\theta_i$ for $h = 1$, x_i - mRNA concentration from gene i , ρ_i is the basal (or leaky) transcription rate for gene i and is associated with very low levels of mRNA, k_{i1} - effective affinity constant for gene 1 activating gene i ($i = 2, \dots, n$), k_{is} - maximum synthesis parameter for gene i , k_{id} - first order degradation rate (or consumption rate) for gene i , \mathbf{x}_0 - vector of initial mRNA concentration, z_i - concentration of translated protein from gene i , $\mathbf{b} = [b_1, \dots, b_n]^T$ is the input matrix and $\mathbf{u} = [u_1, \dots, u_n]^T$ is the input vector (gene triggering compounds). The model formulation with no feedback later on aids the assessment of the marginal contribution of a feedback loop in the network dynamics.

2.2.3 Translation model

A system of linear differential equations to model the protein abundance (translation process) is then considered. The linear model representations (2.3) are used to capture the dynamics of the translation process with both the production and degradation terms being linear.

$$\left. \begin{aligned} \dot{z}_1 &= r_1x_1 - \eta_1z_1 \\ \dot{z}_2 &= r_2x_2 - \eta_2z_2 \\ &\vdots \\ \dot{z}_n &= r_nx_n - \eta_nz_n \\ z_i(0) &= z_{i0} \end{aligned} \right\} \quad (2.3)$$

where r_i - specific translation rate for gene i , η_i - degradation rate for protein i . The z_i 's represent the target proteins in the scheme in Figure 2.1. At steady state the response rate and degradation rate balance, i.e. $\dot{x}_1 \approx \dot{x}_2 \approx \dots \approx \dot{x}_n \approx 0$. By setting the transcription rates $\dot{x}_i = 0$ for all i in (2.2), it follows that

$$\left. \begin{aligned} \tilde{x}_1 &= \frac{1}{k_{1d}}(\rho_1 + b_1\tilde{u}_1) \\ \tilde{x}_i &= \frac{1}{k_{id}}\left(\rho_i + k_{is}\frac{k_{i1}\tilde{z}_1}{1 + k_{i1}\tilde{z}_1}\right) \quad i \geq 2 \end{aligned} \right\} \quad (2.4)$$

The steady state values in (2.4) are based on the assumption that, for a small time window the change in concentration of the input stimulus and metabolite concentrations remain nearly unchanged. The model specifications for the transcription and translation process describe the rates of change of concentration of the genes and proteins. Overall, the system of $2n$ coupled differential equations in (2.2) and (2.3) describe the network dynamics.

2.2.4 System stability

The interesting case to analyze is the systems behavior in the absence of the inhibitor, CreA. Let us denote the equilibrium concentrations of mRNA and protein quantities by the vectors $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_n]^T$ and $\tilde{\mathbf{z}} = [\tilde{z}_1, \dots, \tilde{z}_n]^T$ respectively. Using (2.3) the steady states lead to the relationships $\tilde{z}_i = r_i \tilde{x}_i / \eta_i$ for all i . The stability of each steady state (from (2.2) and (2.3)) can be analyzed using Hopf Bifurcation, an analytic approach that has been widely used in investigating stability conditions in gene expression networks [58, 59, 60, 61].

Let $F : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$ be a set of smooth functions (with $F = (F_1, \dots, F_{2n})$) that capture the XlnR regulon system dynamics. In this case we have $F_1 = \dot{x}_1, \dots, F_n = \dot{x}_n$, and $F_{n+1} = \dot{z}_1, \dots, F_{2n} = \dot{z}_n$ in (2.2) and (2.3) respectively. By definition, the Jacobian matrix is given by

$$\mathbb{J}_n(\cdot) := \left(\begin{array}{ccc} \partial F_1 / \partial x_1 & \dots & \partial F_1 / \partial z_n \\ \vdots & \ddots & \vdots \\ \partial F_{2n} / \partial x_1 & \dots & \partial F_{2n} / \partial z_n \end{array} \right) \Big|_{[\tilde{\mathbf{x}} \ \tilde{\mathbf{z}}]} \quad (2.5)$$

This Jacobian matrix is used to assess the regulon stability and to identify which parameters dictate the transcript abundance. First, consider a case of three genes and three proteins, $n = 3$. The Jacobian is given by

$$\mathbb{J}_3(\cdot) = \frac{\partial F}{\partial [\mathbf{x} \ \mathbf{z}]} \Big|_{[\tilde{\mathbf{x}} \ \tilde{\mathbf{z}}]} \quad (2.6)$$

where $\mathbf{x} = [x_1, x_2, x_3]^T$ and $\mathbf{z} = [z_1, z_2, z_3]^T$. The corresponding steady states of the vectors $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{z}}$ can be computed, accordingly. Using expressions (2.2) and (2.3) in (2.5) we obtain

$$\mathbb{J}_3(\cdot) = \left(\begin{array}{cccccc} -k_{1d} & 0 & 0 & 0 & 0 & 0 \\ 0 & -k_{2d} & 0 & \varphi_{24}(\tilde{z}_1, \cdot) & 0 & 0 \\ 0 & 0 & -k_{3d} & \varphi_{34}(\tilde{z}_1, \cdot) & 0 & 0 \\ r_1 & 0 & 0 & -\eta_1 & 0 & 0 \\ 0 & r_2 & 0 & 0 & -\eta_2 & 0 \\ 0 & 0 & r_3 & 0 & 0 & -\eta_3 \end{array} \right) \quad (2.7)$$

Chapter 2. Modeling and analysis of the dynamic behavior of the XlnR regulon in *Aspergillus niger*

where

$$\varphi_{i4}(\tilde{z}_1, \cdot) = \frac{k_{is}k_{i1}}{1 + k_{i1}\tilde{z}_1} \left(1 - \frac{1}{1 + k_{i1}\tilde{z}_1}\right) \quad (2.8)$$

for $i = 2, 3$. A similar generalized expression for $\varphi_{i,n+1}(\tilde{z}_1, \cdot)$ can be obtained given a regulon with a known number of transcripts n . The characteristic polynomial obtained from (2.7) is given by

$$\mathcal{P}(\cdot) = \prod_{i=1}^3 (\lambda + k_{id})(\lambda + \eta_i) \quad (2.9)$$

In the case of this regulon, the derived characteristic polynomial turns out to be the same as the determinant, i.e. $\mathcal{P}(\cdot) = |\mathbb{J}_n(\cdot)|$. Using the expression (2.9), conditions that ensure global stability can be established on the parameters.

The formulations of the Jacobian matrix and the eigenvalue spectra can be extended to an n -dimensional network system. The generalization for the eigenvalue spectra using a similar analysis as above leads to the expression

$$\mathcal{P}(\cdot) = \prod_{i=1}^n (\lambda + k_{id})(\lambda + \eta_i) \quad (2.10)$$

for $n \in \mathbb{Z}^+$, a positive integer. In the network without feedback, it turns out that the trace of the Jacobian matrix is equal to the sum of all the eigenvalues (i.e. $\text{Trace}(\mathbb{J}_n) = \sum_{i=1}^n \lambda_i$ holds true). The above analysis shows that for the open loop system there is no possibility for oscillatory behavior to occur, and that the time constant only depends on the degradation coefficients.

According to Aro et al. [62], van Peij et al. [34] and Hasper et al. [63], the *A. niger* genes *eglA*, *eglB*, *eglC*, *cbhA*, *cbhB*, *xlnB*, *xlnC* and *xlnD* contain binding sequences (GGCTAAA) to the XlnR protein as well as binding sequences to CreA, a repressor protein acting in the presence of monomeric sugars (i.e., glucose) as an auto-regulating mechanism. This property ensures that most target genes have similar expression dynamics in time.

2.2.5 Feedback in the network

Numerous transcription systems are known to include genes that regulate their own expression values [64]. In our analysis, to model the effect of feedback we hypothesize that the TPs and PTMs in the feedback loop in the scheme in Figure 2.1 only act on the *xlnR* gene. Therefore, only the equation that captures the dynamics for the first gene (x_1) has to be modified accordingly. The adapted equation is given by

$$\begin{aligned} \dot{x}_1(t) = & \rho_1 - k_{1d}x_1(t) + \left[b_1u_1(t) \right. \\ & \left. + k_{ls} \left(\frac{1}{1 + k_{RL} \sum_{j \in \mathbf{S}_1} z_j(t - \tau)} \frac{k_{AL} \sum_{l \in \mathbf{S}_2} z_l(t - \tau)}{1 + k_{AL} \sum_{l \in \mathbf{S}_2} z_l(t - \tau)} \right) \right] H \end{aligned} \quad (2.11)$$

where

$$H = \frac{1}{1 + k_A C_A(t)} \quad (2.12)$$

is the repressor Hill function and C_A - quantitative activity state for CreA, k_A - inverse of the Hill constant of CreA. The term $\tau > 0$ represents a time delay in the feedback loop. The sets $\mathbf{S}_1 = \{j \mid j = 1, \dots, m\}$ and $\mathbf{S}_2 = \{l \mid l = m + 1, \dots, n - 1\}$ where $\mathbf{S}_1 \cup \mathbf{S}_2 = \{1, \dots, n - 1\}$ i.e. collection of all the target proteins in the regulon. All the supposed repressing and activating proteins are lumped in the sets \mathbf{S}_1 and \mathbf{S}_2 , respectively. The effect of the D-xylose and the feedback loop is modeled as additive. Equation (2.11) also specifies the build up of proteins and repression or activation of the *xlnR* gene through the feedback loop. Through the sequence of PTMs the protein availability in the feedback loop is delayed. All the other components representative of the target genes in the network models (2.2) and (2.3) remain unchanged.

2.2.5.1 *xlnR* gene promoter activity under feedback

Let us define the promoter activities by the expressions (2.13) and (2.14). The promoter activity corresponding to the case when an activating protein is involved in the feedback loop is represented by the term Γ_A and that for the case of a repressing feedback effect denoted by Γ_R .

$$\Gamma_A = \frac{k_{AL} \sum_{l \in \mathbf{S}_2} z_l(t - \tau)}{1 + k_{AL} \sum_{l \in \mathbf{S}_2} z_l(t - \tau)} \quad (2.13)$$

$$\Gamma_R = \frac{1}{1 + k_{RL} \sum_{j \in \mathbf{S}_1} z_j(t - \tau)} \quad (2.14)$$

The extracts from the denominator functions are given by (2.15) and (2.16), respectively.

$$P_A = k_{AL} \sum_{l \in \mathbf{S}_2} z_l(t - \tau) \quad (2.15)$$

$$P_R = k_{RL} \sum_{j \in \mathbf{S}_1} z_j(t - \tau) \quad (2.16)$$

These terms are used in the calculations for the activating and repressing promoter activity for the XlnR regulon. For the sake of illustrations, two target genes were considered (i.e. values of $j = 1$ and $l = 2$) in the simulation with

Chapter 2. Modeling and analysis of the dynamic behavior of the *XlnR* regulon in *Aspergillus niger*

one as an activator and the other as a repressor (the values $k_{RL} = k_{AL} = 1$ were used). We consider the sets \mathbf{S}_1 and \mathbf{S}_2 of unit elements which index the proteins that are responsible for regulating the *xlnR* gene through a series of mechanisms in the PTM channel.

2.2.5.2 Existence of oscillatory behavior

The eigenvalue spectra from the derived Jacobian matrix can be used for this analysis. The presence of at least a pair of eigenvalues with complex parts implies the existence of oscillatory behavior. The PTMs in the feedback loop may produce oscillatory behavior depending on the individual attributes of the target genes and the consequent proteins in the feedback loop.

We observed that in the absence of a feedback loop, the system dynamics is dictated by the degradation parameters. Active degradation of proteins or mRNA is a major part of many metabolic and stress response systems [44]. This may not necessarily hold true for the system with a feedback loop because of the structural change in the Jacobian matrix. The metabolites involved in this oscillatory dynamics are presumably determined by the individual biochemical and mechanistic attributes of the individual molecules. Therefore, no single hard rule for classifying which metabolites are responsible for the overall oscillatory behavior of the expression profiles of the genes and proteins can be stipulated. This might however be possible for some specific network pathways for which extensive information is available.

Consider the Jacobian matrix corresponding to three genes and three proteins (2.17). We now analyze the effect of having a protein in the feedback loop. These proteins are considered to regulate the *xlnR* gene and their possible positions in the Jacobian matrix are indicated by "×" as shown in (2.17).

$$\mathbb{J}_3(\cdot) = \begin{pmatrix} -k_{1d} & 0 & 0 & \times & \times & \times \\ 0 & -k_{2d} & 0 & \varphi_{24}(\tilde{z}_1, \cdot) & 0 & 0 \\ 0 & 0 & -k_{3d} & \varphi_{34}(\tilde{z}_1, \cdot) & 0 & 0 \\ r_1 & 0 & 0 & -\eta_1 & 0 & 0 \\ 0 & r_2 & 0 & 0 & -\eta_2 & 0 \\ 0 & 0 & r_3 & 0 & 0 & -\eta_3 \end{pmatrix} \quad (2.17)$$

Using the adapted model (2.11), the computed entry in the $(1, n + i)$ -th cell ($i = j, l$ for all values of j and l) of the Jacobian matrix (2.5) is given by (2.18) and/or (2.19) depending on which proteins in the feedback loop are involved in the regulation of the *xlnR* gene. By taking partial derivatives of the function F_1 with respect to the variable of interest within each of the sets \mathbf{S}_1 and \mathbf{S}_2 , we then have the more compact expression

$$\frac{\partial F_1}{\partial z_j} \Big|_{j \in \mathbf{S}_1} = - \frac{k_{ls}}{(1 + k_{RL} \sum_{j \in \mathbf{S}_1} z_j(t - \tau))^2} \frac{k_{RL} k_{AL} \sum_{l \in \mathbf{S}_2} z_l(t - \tau)}{(1 + k_{AL} \sum_{l \in \mathbf{S}_2} z_l(t - \tau))} < 0 \quad (2.18)$$

This term corresponds to the repressing proteins, and the terms given by

$$\frac{\partial F_1}{\partial z_l} \Big|_{l \in \mathbf{S}_2} = \frac{k_{ls}}{(1 + k_{RL} \sum_{j \in \mathbf{S}_1} z_j(t - \tau))} \frac{k_{AL}}{(1 + k_{AL} \sum_{l \in \mathbf{S}_2} z_l(t - \tau))^2} > 0 \quad (2.19)$$

for the activating proteins. The parameters k_{AL} and k_{RL} represent the lumped affinity constants for the activating and repressing proteins, respectively. Suppose that the XlnR protein ($i = 1$) is the only metabolite responsible for auto-regulation in the feedback loop. By representing the corresponding entry at the $(1, 4)$ -th position in the matrix (2.5) by a nonzero parameter $\omega_1 \in \mathfrak{R} \setminus \{0\}$, a parameter that intrinsically represents the auto-regulation effect of the *xlnR* gene. The computed eigenvalue spectra from (2.7) using the expression $|\mathbb{J}_n - \lambda \mathbf{I}_n| = 0$ is given by

$$\lambda(\mathbb{J}_3(\cdot)) = \{-\eta_2, -\eta_3, -k_{3d}, -k_{2d}, \lambda_5(\cdot), \lambda_6(\cdot)\} \quad (2.20)$$

where

$$\lambda_5(\cdot) = -\frac{1}{2}(\eta_1 + k_{1d} - \sqrt{(\eta_1^2 - 2\eta_1 k_{1d} + k_{1d}^2 + 4\omega_1 r_1)}) \quad (2.21)$$

$$\lambda_6(\cdot) = -\frac{1}{2}(\eta_1 + k_{1d} + \sqrt{(\eta_1^2 - 2\eta_1 k_{1d} + k_{1d}^2 + 4\omega_1 r_1)}) \quad (2.22)$$

are the conjugate roots. The eigenvalues $\lambda_5(\cdot)$ and $\lambda_6(\cdot)$ may take on values from the real space, \mathfrak{R} or the complex space, \mathbb{C} . From (2.21) and (2.22) we observe that oscillation can only be obtained if the condition $\omega_1 < -(\eta_1 - k_{1d})^2/4r_1$ for $r_1 > 0$ is satisfied. This condition on ω_1 signifies a contribution from a NFB loop in the XlnR regulon network. Notice that the expression $(\eta_1 - k_{1d})^2 > 0$ for all values of η_1 and k_{1d} . This finding adds to consolidate the findings by Tiana et al. [48] from a theoretical analysis of three eukaryotic genetic regulatory network in which they attributed the existence of oscillation to a common design of a NFB with underlying time delay. Considering the expressions for the eigenvalues in (2.21) and (2.22), we observe that for a PFB effect of the XlnR protein there is no possibility for oscillatory behavior. This result does not necessarily hold true for the proteins in the feedback loop corresponding to the cells at positions $(1, n + i)$ for $i = 2, \dots, n$.

The presence or absence of oscillatory behavior is insufficient for drawing conclusions about stability in system responses. Stability, using (2.21) and (2.22) exists if the conditions $\mathcal{R}e(\lambda_5(\cdot)) < 0$ and $\mathcal{R}e(\lambda_6(\cdot)) < 0$ are simultaneously fulfilled. Hence, the inequality

$$\mathcal{R}e\left(-\frac{1}{2}(\eta_1 + k_{1d} \pm \sqrt{(\eta_1^2 - 2\eta_1 k_{1d} + k_{1d}^2 + 4\omega_1 r_1)})\right) < 0 \quad (2.23)$$

Solving the inequality (2.23) for ω_1 leads to the condition $\omega_1 < \eta_1 k_{1d}/r_1$. A similar analysis for the existence of oscillatory behavior and stability dynamics can be done for the other proteins in the feedback loop, for example at position

(1, 5) and (1, 6) or combinations in the matrix (2.17). However, although such analysis is conceptually simple, the analytic expressions are very complex to work with. Information about the stability and oscillatory behavior is obtained by numerical solutions. An example is considered to investigate the time evolution of gene activity and protein abundance in the XlnR regulon.

2.2.5.3 Bifurcation analysis

Bifurcation analysis relates to stability on the system parameters. Stability properties for the system without feedback are given by (2.10), where it was shown that the roots of the characteristic polynomial correspond to the degradation rate constants for the mRNA expression and protein abundance. As these constants are positive, the system is always stable. In the case that one of them equals to zero, then the system is critically stable.

For the network with feedback loop, consider the conjugate roots in (2.21) and (2.22) denoted by $\lambda_i(\cdot)$ for $i = 5, 6$. In the event that the conditions $|\mathcal{R}e(\lambda_i(\cdot))| = 0$ and $|\mathcal{I}m(\lambda_i(\cdot))| \neq 0$ are simultaneously fulfilled - then there exists a Hopf bifurcation for the corresponding genes and proteins. Such a bifurcation occurs when the root of the positive discriminant function in (2.23) equates to the sum of the degradation parameters for the *xlnR* gene and XlnR protein, i.e.

$$+ \sqrt{(\eta_1^2 - 2\eta_1 k_{1d} + k_{1d}^2 + 4\omega_1 r_1)} = \eta_1 + k_{1d} \quad (2.24)$$

or after working out becomes $\omega_1 = \eta_1 k_{1d} / r_1$. This example illustrates the case of a feedback in the cell at position (1, 4) of the matrix in (2.17). The analysis for the other entries of ω results in highly complex expressions, therefore a numerical analysis is preferred.

2.3 Results

2.3.1 System specification

The analysis is illustrated by an example case. Consider a regulon network of three genes given a perturbation of D-xylose. The pulse perturbation takes place at time $t = 0$. During fermentation, the D-xylose is consumed and the D-xylose concentration follows the expression $u(t) = u(0)(1/(\beta + e^{Kt}))$, where $u(t) \equiv [\text{D-xylose}]$ and $\beta > 0$, with $K = 0.3$ and $u(0) = 50$ mM as the initial D-xylose concentration.

The parameters used for the simulation are: $b_1 = 1$, $\rho_1 = 2e-3$, $\rho_2 = 2.5e-3$, $\rho_3 = 1e-3$, $k_{1d} = 0.5$, $k_{2d} = 0.4$, $k_{3d} = 0.3$, $k_{2s} = 5$, $k_{3s} = 6$, $k_{21} = 0.1$, $k_{31} = 0.1$, $r_1 = r_2 = r_3 = 0.5$, $\eta_1 = 1$, $\eta_2 = 1$ and $\eta_3 = 1$.

2.3.2 Stability and response analysis - without feedback

The expression for the characteristic polynomial, $\mathcal{P}(\cdot)$ in (2.10) is independent of the translation rate parameters r_i , the gene synthesis coefficient k_{is} and the

terms in the expression (2.8). From (2.10) it can be observed that without feedback, the system is globally stable (i.e. the conditions $\text{Trace}(\mathbb{J}_n(\cdot)) < 0$ and $|\mathbb{J}_n(\cdot)| > 0$ are satisfied for all \tilde{x} and \tilde{z}). The system stability behavior is dictated by how fast the translation and transcription rates are (i.e. magnitudes of k_{id} and η_i).

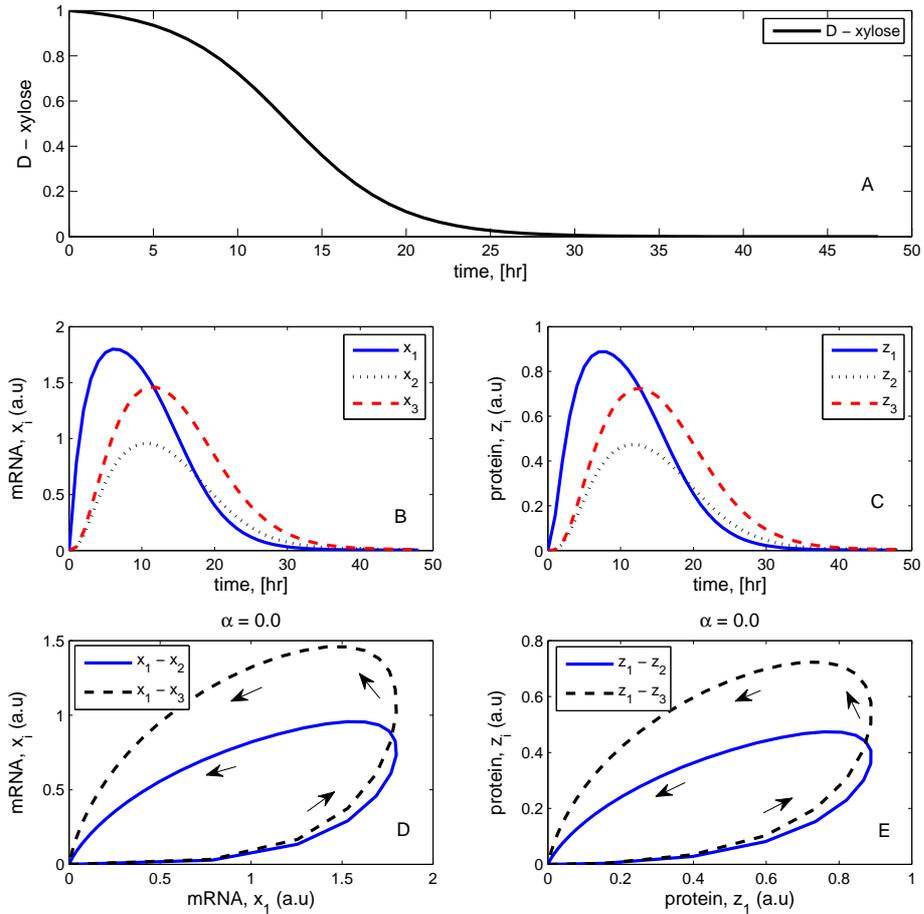


Figure 2.2: D-xylose consumption, gene expression, protein abundance and phase plots: **without feedback**. (A): The simulated trajectory for D-xylose consumption. (B): Gene expressions profiles. (C): Proteins abundance plots. (D): Phase plot for gene expression showing variation of mRNA concentrations of the *xlnR* gene and the other target genes, x_i . (E): Corresponding protein abundance phase plot.

In Figure 2.2 both the gene expressions in plot (B) and protein abundance plot (C) show similar behavioral dynamics. Moreover, with the chosen input pattern of D-xylose the target genes show phase plots similar in patterns but with variations that are dictated by individual gene or protein kinetic parameters. A relaxation time of $\tau_{R1} = 1/k_{1d} \approx 2$ hours is noticed for the master regulator and for the target genes, $\tau_{R1} < \tau_{R2}, \tau_{R3}$. The relaxation time is an

Chapter 2. Modeling and analysis of the dynamic behavior of the *XlnR* regulon in *Aspergillus niger*

approximation for the time required for the system to relax into steady state. This represents the time it takes a system to react to a persistent external input (D-xylose).

2.3.3 Feedback in the network

Since the presence of CreA is a strong repressor that inhibits the *xlnR* gene activity by blocking the promoter binding site, we chose to model this influence by considering a switch-like function with $H \in \{0, 1\}$. Here the assignment of $H = 0$ and $H = 1$ means CreA is present and absent respectively. In the absence of CreA the protein products from the target genes are involved in regulating the activity of the master regulator. These protein products may either inhibit or activate the *xlnR* gene.

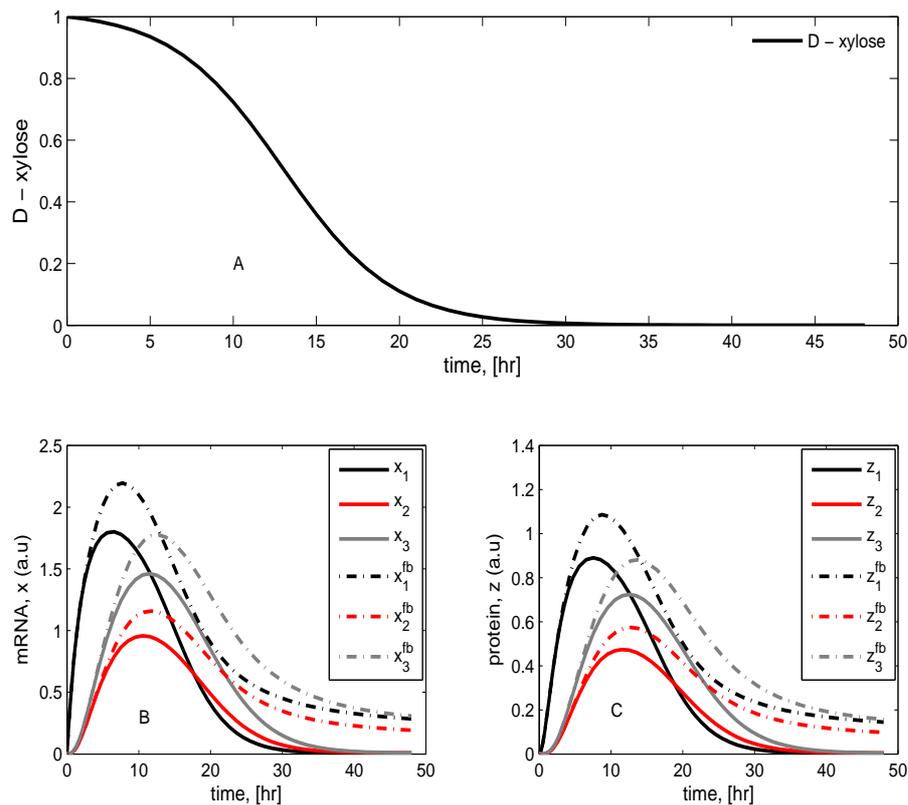


Figure 2.3: **D-xylose consumption, gene expression, protein abundance and phase plots with feedback.** (A): The simulated trajectory for D-xylose. (B): Gene expression profiles with solid lines (—) showing the expression profiles for the genes in the absence of CreA. The corresponding dotted lines (···) show the simulated effect of competitive feedback (with $\tau = 1$). (C): Protein abundance profiles (solid lines).

A comparison of the metabolite expression dynamics for the network with and without feedback loops in the absence of CreA is shown in Figure 2.3. The

same parameter values in the section **System specification** were used for the simulation with the extra parameters from (2.11) being $k_{RL} = 1$ and $k_{AL} = 1$ and the lumped synthesis parameter from (2.11) chosen as $k_{ls} = 1$. Figure 2.3 indicates the enhanced metabolite expression as a result of incorporating a feedback loop with delay (with $\tau = 1$) in the model - a result that is similar to what was observed by Maithreye et al. [65] during a theoretical kinetics analysis of the concentration of green fluorescent protein (GFP) in time.

2.3.3.1 Activating and repressing feedback

The expressions (2.18) and (2.19) have the potential to yield oscillatory behavior in the metabolite response profiles. The oscillatory behavior (if and when it exists) is purely governed by the values of the system mechanistic parameters. Such oscillatory behavioral patterns of gene expression may vary from organism to organism, and can be detected from time series data if enough samples are taken.

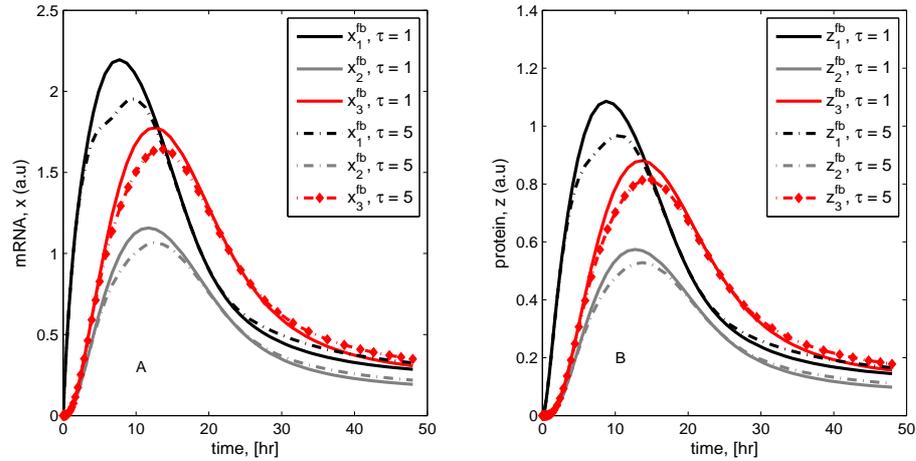


Figure 2.4: **Effect of time delay on expression.** (A)-(B): Plots showing the effect of variation in time delay in the feedback loops corresponding to the transcription and translation processes, respectively. The observed effect on the responses is small except for the slight deviation at the peak of the expression profiles.

To assess the effect of time delays in the transcription and translation processes, some cases were simulated. The results of the expression time-dynamics for both the genes and proteins are shown in Figure 2.4. The simulations were performed for specific cases of $\tau = 1$ hour and $\tau = 5$ hours and the subsequent outputs compared. The metabolite expression patterns from the two cases are nearly similar with the main difference occurring at the maximum level. Overall, longer time delays lead to a small and non significant reduction in expression values.

2.3.3.2 *xlnR* gene promoter site activity

The competitive effect of the activators and repressors for the promoter binding sites was also simulated. The effect of which transcription factor (TF) (either an activator or a repressor) wins occupancy of a promoter binding site depends partly on the strength of the synthesis parameter k_{ls} (Figure 2.5).

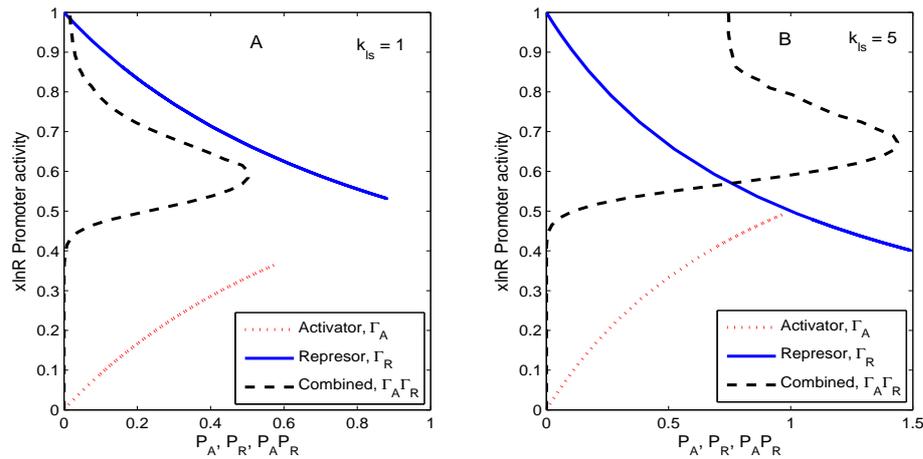


Figure 2.5: *xlnR* promoter region activity. Plot of the *xlnR* promoter region activities defined by Γ_A , Γ_R and $\Gamma_A \Gamma_R$ depending on the regulator. The term $\Gamma_A \Gamma_R$ - is the combined affect of competitive binding to promoter region by activators and repressors. Plots (A) and (B) show the influence of weak ($k_{ls} = 1$) and strong ($k_{ls} = 5$) synthesis parameters respectively.

The promoter is most active (activity around 50 – 80%) when the regulon is fully active. This corresponds with the time window at which the network is fully responsive to the external perturbation. We observe that the *xlnR* gene activator has a tendency of occupying most of the promoter sites at any given time (Figure 2.5).

2.3.3.3 Bifurcation and oscillatory behavior analysis

A range of values of activating and repressing parameters $\omega_1, \omega_2, \omega_3$, respectively on entries (1, 4), (1, 5) and (1, 6) in expression (2.17) was considered for analyzing the stability behavior of the network. It was observed that NFB on ω_1 gives a stable system and values of ω_2 and ω_3 below -977 results in an unstable systems, Figures 2.6 (A)-(C). The PFB effect of the XlnR protein on the *xlnR* gene leads to unstable system dynamics for $\omega_1 > 1$. This can be seen from Figure 2.6 (A). This result also leads to the conclusion that the XlnR regulon is unstable if the *xlnR* gene has a PFB from its own protein.

An analysis of how the various feedback parameters affect the oscillatory behavior of the gene and protein expression was also considered. The results show that there exist threshold values (or a range of parameter values) for the feedback parameters ω_j 's for which their oscillatory behavior may or may not occur (see Figure 2.7 (A)-(C)). The value $\omega_j = 0$ corresponds to no feedback in

the system and according to the previous analysis (under the subsection: stability and response analysis - without feedback), no oscillation occurs in this case. A transient oscillatory behavior is observed for values of the parameter $\omega_j \approx 0$ for all j , Figure 2.7 (B)-(C). The observed stability curve corresponding to the XlnR protein in the feedback loop (ω_1) is a near reflection of the corresponding resultant oscillation curve (Figure 2.6 (A) and Figure 2.7 (A)).

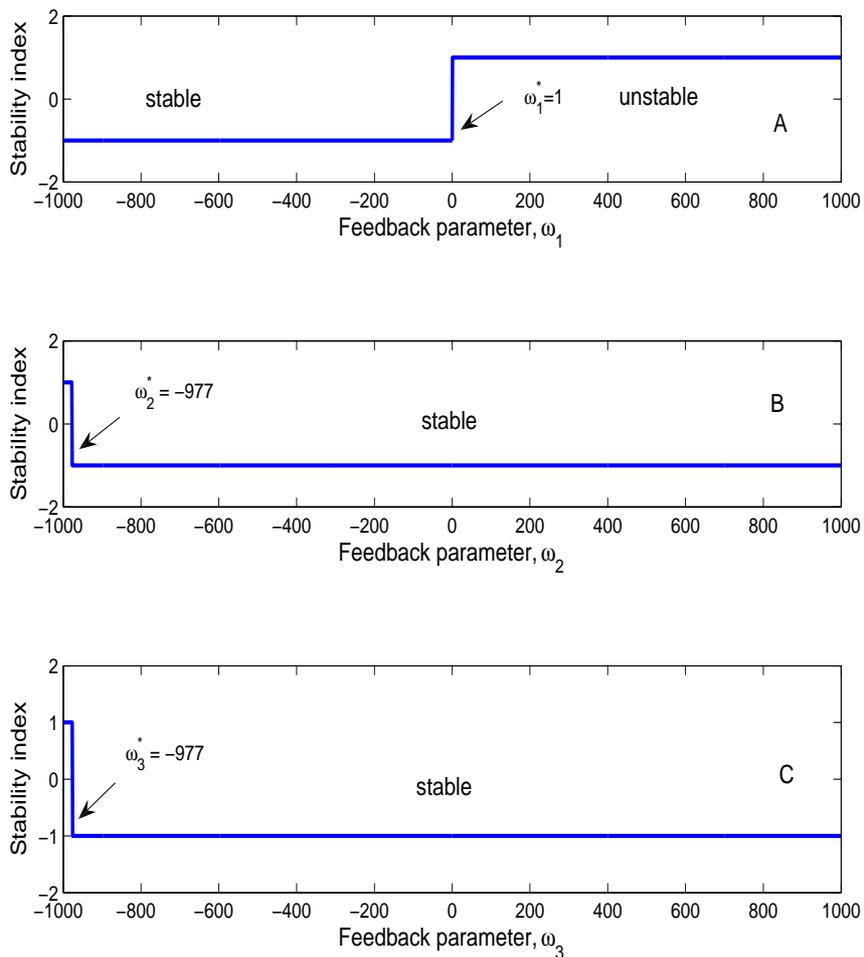


Figure 2.6: Stability index curves. (A)-(C): Plots of the stability indices for various values of $\omega_j \in (-1000, 1000)$ for $j = 1, 2, 3$. A stability index value of -1 and $+1$ indicates stable and unstable responses of gene and protein expression in time, respectively. The value $\omega_j < 0$ represents repressing feedback and $\omega_j > 0$ is the activating feedback effect. Simulations performed using parameters from **System specification** and the pseudo steady state expression values at which the Jacobian is estimated.

Chapter 2. Modeling and analysis of the dynamic behavior of the XlnR regulon in *Aspergillus niger*

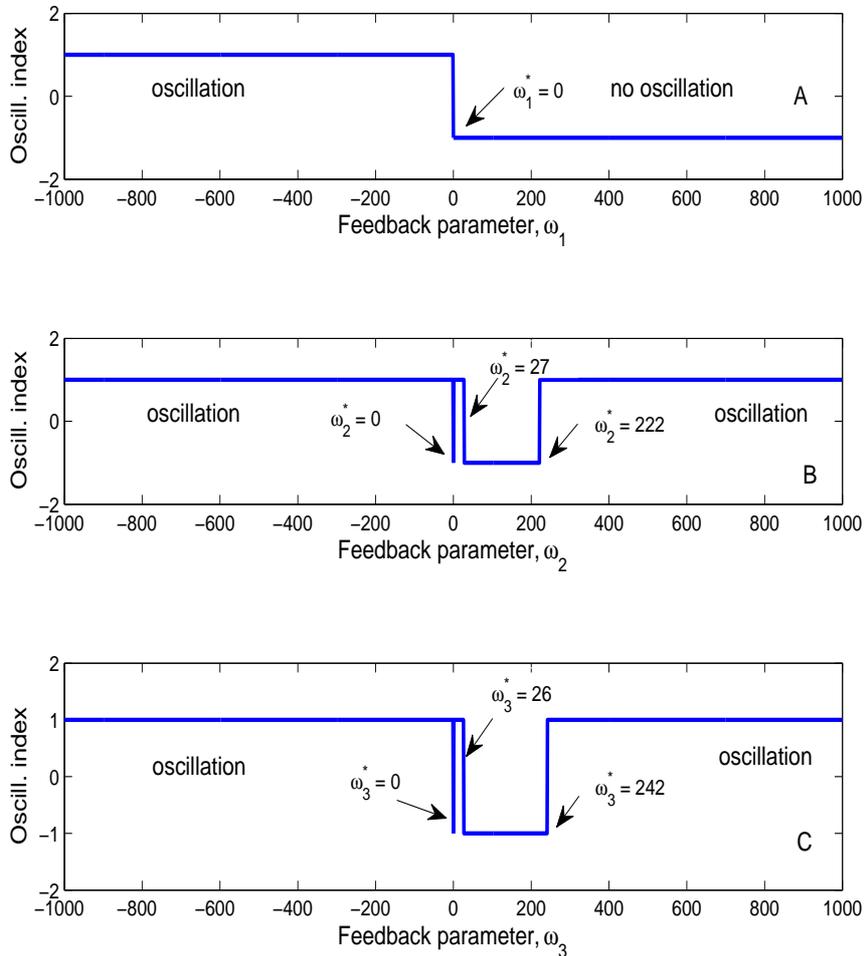


Figure 2.7: **Oscillation index curves.** (A)-(C): The plots indicate possible oscillatory behavior for various values of $\omega_j \in (-1000, 1000)$ for $j = 1, 2, 3$. Each ω_j is considered independently while the others are set to zero. The oscillation indices -1 and $+1$ represent non-existence and existence of oscillatory behavior in the response behavior of the metabolites, respectively.

2.4 Discussion

The model gives a better understanding of the rate limiting steps in the process of activating the XlnR regulon and therefore, helps to define the biological control points. Similarly this knowledge can be used to obtain strains that have enhanced xylanolytic enzyme production. These enzymes are industrially of importance as food and feed additives, but are also part of a system that is used to bleach paper pulp.

Given that the transcription rate and degradation rates have been shown to be the key parameters that dictate the systems dynamics for the XlnR regulon; this information is important for designing and sampling of time course exper-

iments. Once the transcripts are unstable, the proteins get quickly degraded; otherwise they remain stable. This observation is linked to the D-xylose uptake in fermentation experiments. The consumption of D-xylose also indirectly controls the regulation of the target genes and therewith the breakdown of sugars.

Simulations showed that the dynamics of the D-xylose input function considered in the examples has an important effect on the profiles of the individual metabolite concentrations. This is particularly dictated by the value of the parameters in the external input function $u(t)$. The larger the value of the K , the faster the consumption of D-xylose. This depends on the chemical reactions taking place in any given cell, or the saturation levels of the individual compounds in a cell.

Feedback significantly affects the response of the output profiles for the metabolites and changes the final steady state values (Figure 2.3). Further simulations showed that variations of the time delay in the feedback loop ($\tau = 1, 2, \dots, 5$ hours) have a small effect on the pattern of the response (Figure 2.4). The stability analysis subsection shows that the metabolite response dynamics exhibits no oscillatory behavior for a network without feedback loop. For a network with feedback loops, the results from numerical analysis showed that feedback conditions for which the system is stable or for which the system exhibits oscillatory behavior can be obtained (Figure 2.7 (A)-(C)). In modeling the feedback loop, time delay was accounted for and included in the model.

According to Bliss et al. [49] and Thomas and d'Ari. [66], including time delay in modeling biological networks is considered important because many biological systems exhibit some delays in their feedback loops. However, according to our finding (Figure 2.4), incorporating the time delay had no strong effect on the overall dynamics of the metabolite expression profiles.

The analysis shows that the existence or absence of oscillatory behavior is dictated by the numerical values of the individual mechanistic parameters. The conditions for oscillatory behavior follow from the eigenvalue spectra. The eigenvalue spectra analysis like that in (2.20) and the corresponding conditions for which all the eigenvalues are less than zero, gives also indication for the stability properties for the XlnR network with feedback loop. If all the eigenvalues satisfy the condition $\mathcal{R}e(\lambda(\mathbb{J}_n(\cdot))) < 0$ for all entries, then the system is stable, otherwise it is unstable.

Two scenarios can be considered: one in which the proteins involved in the feedback loop are activating and the other in which the proteins are repressing. The details of the expected behavioral dynamics from such a system requires a case by case analysis (like in Figure 2.6 (A)-(C)) of the effects of the proteins in the feedback loop. A similar analysis can be extended to study the stability in case of a combined effect of any two or more proteins of interest. When the number of network components become large, obtaining explicit analytic solutions and expressions for the eigenvalues and other quantities of interest increasingly become complex - in which case the alternative of numerical methods can be used (see Figure 2.6 (A)-(C)).

Thomas and d'Adri [67], and Thomas et al. [66] investigated the properties

Chapter 2. Modeling and analysis of the dynamic behavior of the XlnR regulon in *Aspergillus niger*

of mathematical Boolean net Modeling Genetic Networks works - investigations that provided significant insight into genetic network dynamics. In their work they showed the importance of NFB loops for maintaining homeostasis in levels of gene products. Our analysis leads to the observation that having a NFB loop stabilizes the response of the metabolite expressions (Figure 2.6 (A)-(C)). However, there exists certain ranges of values of the strength of feedback effects that make the system unstable. This sets constraints to the feasible parameter for the system if instability is not observed. In some cases having a PFB loop yields a stable network response, Figures 2.6 (B) and (C). This result is in agreement with that found in a study by Maithreye et al. [65]. In their investigations they found that NFB loops provide stability and withstand considerable variations and random perturbations of biochemical parameters.

The effect of time delay on stability can be analyzed from a transfer function of the model in the " s " domain, or by a transformation to the " z " domain. In these cases the delay time is considered as a finite dimensional system. Stability analysis can be done by searching for stability properties in the " s " domain or " z " domain. Examples of other methods that deal with the delay times are given for state estimation in the work by Liu et al. [68] and Yu et al. [69].

The adaptive filtering approach developed in [69] is based on the adaptive synchronization setting, for estimating unknown delayed genetic regulatory networks with noise disturbance. Using this approach, no exclusive knowledge of system parameters is required, e.g. those lacking in the XlnR regulon and many other biological networks. Liu et al. [68] proposed an adaptive feedback control approach for simultaneously identifying unknown (or uncertain) network topological structure, unknown parameters of uncertain general complex networks with time delay from available mRNA data and estimation of protein concentration. The effectiveness and applicability of their approach was shown using *in silico* numerical simulations. In contrast to [68] and [69], we study the XlnR regulon dynamics and do not focus on the system structure identification and parameter estimation.

According to Balsa-Canto et al. [70], powerful mathematical analytic tools highlight the value for successful study of many biological systems. However, such success can mainly be attributed to the unrelenting endeavors for an in-depth understanding of both computational methods and the biological problems of interest. For the case of the XlnR regulon, our analysis provides a basis for understanding the behavioral dynamics of genes and proteins after network perturbation. This will form a basis for future wet-lab experiments, particularly with the genes from the XlnR regulon. Given that the metabolite expression dynamics are known, this study provides a basis for strategic thinking in line with experimental design.

The modeling approach used in this paper provides good information for understanding network behavioral dynamics particularly for small-sized networks. This is illustrated by the XlnR regulon in which even the simplest of structures can yield interestingly complex dynamics. Therefore, a reason for having limited our focus to the regulon dynamics. Having detailed infor-

mation regarding the basal parameters and the other mechanistic parameters might further improve the analysis and investigations into the network dynamics. Nevertheless, with informed parameter guesses, simulation studies provide good information into the systems behavior.

2.5 Conclusions

The investigations in this paper considers the XlnR regulon as a dynamic system instead of a static system. Our study provides insight into the dynamic properties of the XlnR regulon. By studying this system, it has become more clear that the transcription and translation degradation rate parameters and the D-xylose consumption profile dictate most of the dynamics in the regulation properties of the network.

The existence of oscillatory behavior depends on the conditions of the mechanistic parameters in the feedback loop - conditions that cannot always be generalized analytically and therefore, must be treated by numerical analysis. The role played by feedback in the network dynamics was found to be significant on the expression dynamics of genes and proteins. This means that the effect of the feedback should be considered in the model if there is sufficient supportive biological need or evidence from data. Just like for most biological systems, this is no doubt an important piece of information for the accurate modeling of biological network.

The analysis of the network dynamics has provided useful information for future *in vitro* experimental work. Particularly the potential for hypothesis testing basing on this work and the design of related perturbation experiments to generate time series data. Once there are available techniques for the network structural identification and parameter estimation for the XlnR regulon can be investigated.

Authors' contributions

LHdG provided the biological knowledge that was used for the model formulations. JO performed the modeling, data analysis and wrote the manuscript. GvS and AJBvB also contributed in calculations and critical review of the methods used in the analysis. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work is supported by the graduate school VLAG and the IPOP program of Wageningen University.

**Evaluation of design strategies
for time course experiments in
genetic networks: case study of
the XlnR regulon in *Aspergillus
niger***

Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*

Abstract.

One of the challenges in genetic network reconstruction is finding experimental designs that maximize the information content in a data set. In this work, the information value of mRNA transcription time course experiments was used to compare experimental designs. The study concerns the dynamic response of genes in the XlnR regulon of *Aspergillus niger*, with the goal to find the best moment in time to administer an extra pulse of inducing D-xylose. Low and high D-xylose pulses were used to perturb the XlnR regulon. Evaluation of the experimental methods was based on simulation of the regulon. Models that govern the regulation of the target genes in this regulon were used for the simulations. Parameter sensitivity analysis, the Fisher Information Matrix (FIM) and the modified E-criterion were used to assess the design performances. The results show that the best time to give a second D-xylose pulse is when the D-xylose concentration from the first pulse has not yet completely faded away. Due to the presence of a repression effect the strength of the second pulse must be optimized, rather than maximized. The results suggest that the modified E-criterion is a better metric than the sum of integrals of absolute sensitivity for comparing alternative designs.

Keywords: Experimental design strategies, genetic network, trigger experiments, time course data, parameter estimation, XlnR regulon, *A. niger*.

This chapter is based on:

J. Omony, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten and A.J.B. van Bortel. Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(5), pp. 1316-1325, 2012.

3.1 Introduction

Obtaining information from gene expression data sets requires the highly involving task of reverse engineering a given network. The reverse engineering of biological networks involves both structural identification and parameter identification. Our work deals particularly with the parameter identification. Parameter values that characterize dynamic gene responses are essential for understanding genetic networks and are also needed to predict systems behavior; hence, the interest in parameter identification in Systems Biology [71, 72, 73, 74]. On the basis of a case study for *A. niger* this paper explores options for experimental designs that maximize parameter identifiability. One particular area of interest is the need to obtain estimates for a system of equation parameters that best describe measured gene expression trajectories in time.

A. niger is a fungus species used for the production of organic acids and a variety of other commercial enzymes. Other industrial applications of *A. niger* include the pulp and paper, detergents, textile, food and beverages (see [13, 14, 15, 17] for details). The organism and one of its main regulons (the XlnR regulon) have been comprehensively studied. The genetic response of the XlnR regulon in *A. niger* is dynamic, but in previous studies it has always been considered as a static problem. A first step to quantify the dynamics of the regulon was given by Omony et al. [75]. The response is induced by D-xylose, then it takes some time before the target genes in the XlnR regulon are fully expressed and the gene expression levels vary in time. Previous studies of the XlnR regulon of *A. niger* were based on 1 mM D-xylose triggers and gene expression at a fixed time point after induction [35]. To gain more knowledge on the regulon dynamics, it is vital to quantify the gene transcription in time. An approach to get a strong quantification of the response is to apply multiple triggers in time. The key questions of interest are of what strength and at what time instants these triggers should be applied to ensure efficient parameter identification.

Genetic response to an external stimulus can be gained by using real-time fermentation experiments and quantitative PCR analysis of the samples. The principles used in optimal experiment design (OED) are then used to interpret the information obtained from the XlnR regulon response. For example, Vanrolleghem et al. [76] showed that the information content and quality of an experiment is highly dependent on the design, and major improvements can be made by the varying experimental decision variables during the experiment. More information can be obtained by changing the experimental conditions at a specific time during experimentation. Model-based OED strategies aimed to maximizing parameter precision are regularly applied in industry and academia [77]. OED is gaining popularity in the field of modeling biological networks in Systems Biology. Some example applications in Systems Biology are [78, 79, 80, 81, 82, 83]; however, applications in modeling real time course gene expression data sets are still limited. An overview of the working principles of OED can be found in the review article by Kreutz and Timmer [84].

Administering additional pulses on genetic networks has been hypothe-

sized to provide more informative experiments [83]. Following this idea, in this work we explore the potential of various experimental design strategies, involving (i) a single pulse, (ii) two pulses, with a varying level of the second pulse. *In silico* experiments are performed by simulating the mRNA time course over 8 hours of selected test-case genes from the XlnR regulon of *A. niger*. Also an experiment is performed to investigate the possibility of reducing the number of samples by increasing the sample time. The information content of these experiments is compared on the basis of two alternate criteria for experimental design: the sum of integrals of absolute parameters sensitivities and the modified E-criterion. The results are tested by looking at the relative estimation error of the parameter estimates.

3.2 Methods

3.2.1 Target gene transcription in the XlnR regulon

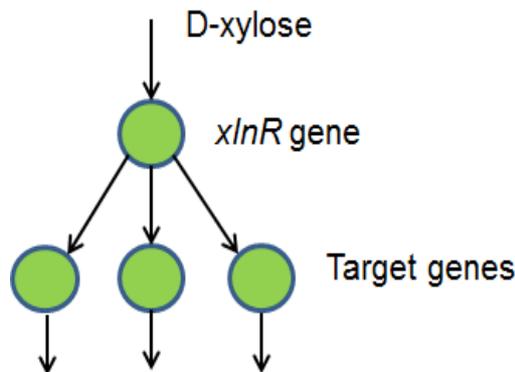


Figure 3.1: A graphical representation of a branching network of the XlnR regulon.

Typically, genetic networks consist of many genes. In the XlnR regulon the *xlnR* gene controls the activity of all the target genes in a branching topology (Figure 3.1). It is assumed none of the target gene products play a role in regulating the other target genes. Considering a single branch in the network in Figure 3.1 for a target gene i an enhanced scheme showing the key components are shown in Figure 3.2. This representation is derived from the work of Omony et al. [75] in which they modelled the dynamics of the XlnR regulon. After administering each trigger, the D-xylose is consumed and its concentration corresponds to an exponential decay function of the form

$$u_1(t) = u_1(0)(1/(\beta + e^{Kt})), \text{ given } u_1(0) \quad (3.1)$$

Using wet-lab experimental data [85], it was observed that the D-xylose consumption follows such a curve. The nonnegative parameters β and K can be estimated given availability of D-xylose data. Typically, the parameter β is infinitesimally small. Let us consider the model describing the dynamics for

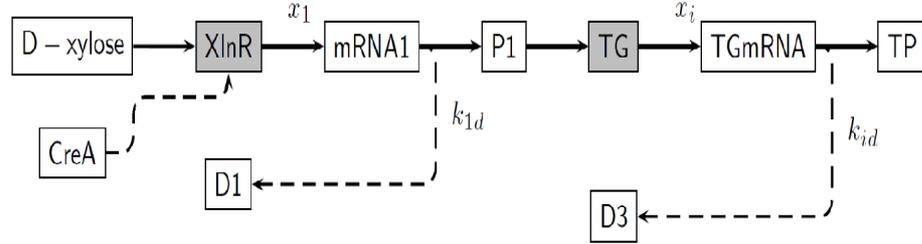


Figure 3.2: A scheme of the XlnR regulon showing induction with D-xylose. The terms D1 and D3 represent the mRNA degradation for the *xlnR* gene and the other target genes, respectively. The term mRNA1 and P1 represent the mRNA levels and the translated proteins for the *xlnR* gene. CreA is a catabolic repressor protein. The target gene mRNA is represented by TGmRNA and the corresponding proteins as a result of the translation from the target genes is represented by TP. The transcription level of the gene *xlnR* gene is represented by x_1 .

the *xlnR* gene. From the work of Omony et al. [75], the effects of any post-translational modifications are not considered in the modeling. The model that describes the regulatory dynamics of the *xlnR* gene is given by

$$\dot{x}_1 = \rho_1 - k_{1d}x_1 + b_1u_1, \quad x_1(0) = x_{10} \quad (3.2)$$

At induction of the XlnR regulon with D-xylose, the catabolic repressor protein CreA has an important effect on the expression of the target genes. At high D-xylose concentrations CreA represses the expression of the target genes, while at low D-xylose concentrations (in the range below 1 mM) the influence of CreA is absent. The repressing effect of CreA is modeled by the Hill function $1/(1 + k_{i2}u_1)$. Since the XlnR protein regulates all the target genes, the equation structure for all target genes is the same. The model describing the regulation for target gene i at high D-xylose concentration is given by

$$\dot{x}_i = \rho_i + k_{is} \frac{k_{i1}x_1}{1 + k_{i1}x_1} \frac{1}{1 + k_{i2}u_1} - k_{id}x_i, \quad x_i(0) = x_{i0} \quad (3.3)$$

Next, we sought to write this expression as a function of the D-xylose concentration. To do this, we need to find a relationship between the expression of the *xlnR* gene and the D-xylose concentration (u_1). From real time qPCR experimental studies, the mRNA levels of the *xlnR* gene were observed to be low and constitutively expressed in time [85]. We used the pseudo-steady state assumption in (3.2) and set $\dot{x}_1 = 0$ with $\rho_1 = 0$. This leads to the expression $\tilde{x}_1 = b_1u_1/k_{1d}$. This pseudo-steady state assumption is associated to the fast reaction of the *xlnR* gene compared to the target genes. By substituting for \tilde{x}_1

Chapter 3. Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*

in (3.3), the differential equation becomes

$$\dot{x}_i = \rho_i + k_{is} \frac{k'_{i1} u_1}{1 + k'_{i1} u_1} \frac{1}{1 + k_{i2} u_1} - k_{id} x_i, \quad x_i(0), \quad u_1(0) \quad (3.4)$$

for all i where $k'_{i1} = k_{i1} b_1 / k_{id}$. The inverse of this parameter determines the threshold of any existing switching behavior. The basal parameter is assumed to be negligible (therefore, $\rho_i = 0$ for all i). This is because of the reference time point $t = 0$ and the gene transcription values are relative. The production term in (3.4) is a Hill type function which is a popular formalism for modeling biological networks [55, 56]. This production term expresses direct activation of a target gene by D-xylose $k'_{i1} u_1 / (1 + k'_{i1} u_1)$ and a repression term $1 / (1 + k_{i2} u_1)$ for CreA which is active at a high D-xylose concentration. The effect of this on (3.4) is that it leads to repressed gene activity at high D-xylose concentration and enhanced gene activity at low concentration [19].

The model representation in (3.4) is derived from the full model for the XlnR regulon in Figure 3.2 (a simplification of the scheme given in [75]) by using the pseudo-steady state assumption for the *xlnR* gene response. For experiments with low D-xylose levels, the term $1 / (1 + k_{i2} u_1)$ in (3.4) approaches a limiting value of 1. Then we have

$$\dot{x}_i = k_{is} \frac{k'_{i1} u_1}{1 + k'_{i1} u_1} - k_{id} x_i, \quad x_i(0) \quad (3.5)$$

which can be seen as a special case of (3.4).

3.2.2 Parameter sensitivity functions

In literature, OED has been used to study various biochemical networks (e.g. in the work of Zi et al. [86], and Zhang and Rundel [87]). Sensitivity analysis can be used to classify which parameters are important in a given system [88]. Using the models that we have described, the parameters of interest in the XlnR regulon are k_{is} , k'_{i1} , k_{i2} and k_{id} . Expression (3.4) is differentiated partially with respect to each parameter in the model. The resulting (local) sensitivity function is given by the expression

$$\dot{x}_{i,\vartheta_r}(t) = \frac{\partial \phi_i}{\partial x_i} x_{i,\vartheta_r}(t) + \frac{\partial \phi_i}{\partial \vartheta_r} \quad (3.6)$$

where ϕ_i is the right hand side of (3.4) and $\partial \phi_i / \partial x_i$ is a diagonal system Jacobian matrix (all other terms being zero due to the structure of (3.4) and (3.5)). For more examples of applications of sensitivity analysis see the work of Gunawan et al. [89]. The partial derivatives with respect to each of the parameters are the sensitivity functions. Using (3.4) and (3.6) for each of the parameters, the time-dependent parameter sensitivities are computed from the system of equations

$$\frac{d}{dt}S_{k_{is}}(t) = -k_{id}S_{k_{is}}(t) + \frac{k'_{i1}u_1}{(1+k'_{i1}u_1)(1+k_{i2}u_1)} \quad (3.7)$$

$$\frac{d}{dt}S_{k'_{i1}}(t) = -k_{id}S_{k'_{i1}}(t) + \frac{k_{is}u_1}{(1+k'_{i1}u_1)^2(1+k_{i2}u_1)} \quad (3.8)$$

$$\frac{d}{dt}S_{k_{i2}}(t) = -k_{id}S_{k_{i2}}(t) - \frac{k_{is}k'_{i1}u_1^2}{(1+k'_{i1}u_1)(1+k_{i2}u_1)^2} \quad (3.9)$$

$$\frac{d}{dt}S_{k_{id}}(t) = -k_{id}S_{k_{id}}(t) - x_i \quad (3.10)$$

with all the initial conditions satisfying $S_{\bullet}(t_0) = 0$. Here, the subscript notation " \bullet " represents a parameter. To design the time course experiments, the sensitivity functions in (3.7)-(3.10) were solved in conjunction with (3.4).

By using (3.6) for parameters in the expression for the low D-xylose concentration (3.5), the system of differential equations for the sensitivity functions are

$$\frac{d}{dt}S_{k_{is}}(t) = -k_{id}S_{k_{is}}(t) + \frac{k'_{i1}u_1}{1+k'_{i1}u_1} \quad (3.11)$$

$$\frac{d}{dt}S_{k'_{i1}}(t) = -k_{id}S_{k'_{i1}}(t) + \frac{k_{is}u_1}{(1+k'_{i1}u_1)^2} \quad (3.12)$$

$$\frac{d}{dt}S_{k_{id}}(t) = -k_{id}S_{k_{id}}(t) - x_i \quad (3.13)$$

given the initial conditions, $S_{\bullet}(t_0) = 0$.

3.2.3 Experimental evaluation, comparison and parameter estimation

To evaluate the experimental designs, we used the Fisher Information Matrix which is defined by the quadratic function

$$\text{FIM}_i = \sum_{j=1}^N \left(\frac{\partial y_i}{\partial \vartheta_r}(j) \right)^T \Lambda_i \left(\frac{\partial y_i}{\partial \vartheta_r}(j) \right) \Big|_{y_i=y_i(j, \hat{\vartheta}_r)} \quad (3.14)$$

where Λ_i is a symmetric positive definite weighting matrix, $\vartheta_r \in \{k_{is}, k'_{i1}, k_{i2}, k_{id}\}$ is the parameter set to be identified. These are subject to the nominal parameter values ϑ being the maximum likelihood estimates. We write $\partial y_i / \partial \vartheta_r$ for the output sensitivity function for the parameter ϑ_r . It represents a change in the data output measurement value y_i per unit change in a given parameter ϑ_r . Often, the matrix Λ_i is unknown and is user-defined or estimated from a given data set. By considering a single output variable (with white measurement noise on the data), then $\Lambda_i = I \in \mathbb{R}^{N \times N}$, the identity matrix. In this case the noise vector entries are independent and therefore, have zero covariance, $\text{Cov}(\varepsilon_i, \varepsilon_k) = 0$ for all $i \neq k$.

Chapter 3. Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*

It is clear from (3.14) that the sensitivity functions (3.7)-(3.10) or (3.11)-(3.13) directly appear in the computation of the Fisher Information Matrix. The Fisher Information Matrix is equivalent to the inverse of the estimated parameter covariance matrix [90], and is expressed as

$$\begin{aligned} \text{FIM}_i = \Sigma_i^{-1} &= \begin{pmatrix} \hat{\sigma}_{\hat{k}_{is}}^2 & \hat{\sigma}_{\hat{k}_{is}\hat{k}'_{i1}} & \hat{\sigma}_{\hat{k}_{is}\hat{k}_{i2}} & \hat{\sigma}_{\hat{k}_{is}\hat{k}_{id}} \\ & \hat{\sigma}_{\hat{k}'_{i1}}^2 & \hat{\sigma}_{\hat{k}'_{i1}\hat{k}_{i2}} & \hat{\sigma}_{\hat{k}'_{i1}\hat{k}_{id}} \\ \vdots & & \hat{\sigma}_{\hat{k}_{i2}}^2 & \hat{\sigma}_{\hat{k}_{i2}\hat{k}_{id}} \\ & \dots & & \hat{\sigma}_{\hat{k}_{id}}^2 \end{pmatrix}^{-1} \\ &= \left(\frac{\xi_i^T \xi_i}{N - q} \left(J_i^T(\hat{\vartheta}) \Lambda_i J_i(\hat{\vartheta}) \right)^{-1} \right)^{-1} \end{aligned} \quad (3.15)$$

This is a representation for the four parameter case $q = 4$. Similarly, for $q = 3$ a Fisher Information Matrix can be defined by "omitting" the third column and third row in (3.15). The estimated parameter standard deviations ($\hat{\sigma}_{\hat{\vartheta}_r}$) are obtained by taking the square root of the diagonal elements of the matrix Σ_i ; where the off-diagonal entries are the parameter covariances. The desired situation for efficient identification is that the assumption of independence holds, i.e. $\mathbb{E}(\hat{\sigma}_{\hat{\vartheta}_r, \hat{\vartheta}_l}) = 0$ for all $r \neq l$.

Numerous OED evaluation criteria can be found in literature, e.g. the A - optimal design criterion: $\min(\text{Tr}(\text{FIM}_i^{-1}))$, the modified A - optimal design criterion: $\max(\text{Tr}(\text{FIM}_i))$, the D - optimal design criterion: $\max(\det(\text{FIM}_i))$, the E - optimal design criterion: $\max(\lambda_{\min}(\text{FIM}_i))$, and the modified E - optimal design criterion (see [91, 92, 93]). All together, these design criteria can be grouped into four major categories [93]. We used the modified E - optimal design criterion which is given by

$$E_{\text{modified}} = \min_i \left(\frac{\lambda_{\max}(\text{FIM}_i)}{\lambda_{\min}(\text{FIM}_i)} \right) \quad (3.16)$$

The design quantifier specified in (3.16) is a measure of the shape of the sum of squared errors from a model fit - during the systems identification. Some OED design criteria (e.g. the A - and modified A - criterion) are rarely used because they have been reported to occasionally yield non-informative experiments [94]. The advantage of using the modified E - criterion is that it keeps check of both the upper and lower eigenvalue and leads to informative experiments. It ensures that the condition number is not inflated. The modified E - criterion sacrifices accuracy in parameter estimates for improved numerical properties [92].

The model fits to data were based on a minimization of the sum of squared errors between the measured and estimated data. This is given by the goal function

$$\mathcal{J}_i(\hat{\vartheta}) = \arg \min_i \sum_{j=1}^N (y_{ij} - \hat{y}_{ij}(\hat{\vartheta}))^2 \quad (3.17)$$

where y_{ij} and \hat{y}_{ij} are the measured and estimated mRNA level of target gene i at time instant j , respectively. The singular value decomposition method was used to compute the covariance matrix. The residual covariance matrix was obtained from $\Theta_i = \xi_i^T \xi_i / (N - q)$. The parameter variances were computed using the expression $\Sigma_i = \Theta_i (\mathcal{J}_i^T \mathcal{J}_i)^{-1}$. The standard deviations of the parameters were obtained by taking the square root of the entries in the estimate of the diagonal matrix Σ_i , i.e.

$$\hat{\sigma}_{\hat{\vartheta}} = \text{diag}(\sqrt{\hat{\Sigma}_i}) = \{\hat{\sigma}_{\hat{k}_{is}}, \hat{\sigma}_{\hat{k}'_{i1}}, \hat{\sigma}_{\hat{k}_{i2}}, \hat{\sigma}_{\hat{k}_{id}}\} \quad (3.18)$$

The Coefficient of Variation is defined as the ratio of the standard deviation of a parameter to the parameter mean value, i.e. $c_v = \hat{\sigma}_{\hat{\vartheta}_r} / \hat{\vartheta}_r$. The smaller the c_v value, the less the dispersion in the variable. It aims to describe the extent of dispersion of a variable in a way that is independent of the measurement units of a variable.

3.2.4 Data generation and evaluation procedure

The model of the test regulatory mechanism in the XlnR regulon is used for data generation. For the models in (3.4) and (3.5), the systems outputs are given by

$$\begin{cases} \dot{x}_i(t) = f(x_i(t), u_1(t), \vartheta), & \text{given } x_i(0) \\ y_i(j) = x_i(j) + \varepsilon_i(j) \end{cases} \quad (3.19a)$$

$$(3.19b)$$

where $f(x_i(t), u_1(t), \vartheta)$ corresponds to the right hand side of (3.4) for the high D-xylose concentration and (3.5) for the low trigger concentration case. The gene expression was sampled at time instants given by j and white noise was added to the output. The simulated data was then used for the parameter estimation.

The above data simulations are all subject to:

$$\Gamma := \begin{cases} \text{Pulse}_1, \text{Pulse}_2 : u_{1,min} \leq \|u_1\|_1 \leq u_{1,max}, & (3.20a) \\ : t_L \leq \delta t < t_L + 1, & (3.20b) \\ : t_L \leq t_{\text{Pulse}_2(j)} \leq t_R & (3.20c) \end{cases}$$

where $t_{\text{Pulse}_2(j)}$ is an indicator of the time instant at which the second pulse is given. The notations t_L and t_R represent the left and right window of the second pulsing time instant, respectively.

To evaluate the experimental strategies, we consider a time course experiment with one target gene and two 1 mM D-xylose pulses. The first pulse is given at t_0 and a second pulse later on in time when the genes are beginning to exhibit decreased activity levels. Suppose that time course experiments have a duration of 8 hours, in which measurements are taken every 20 minutes (for a total of $N = 25$ samples). Next we sought to find the best moment in time to administer the second trigger. Then between the second and fourth hour-mark there is a total of six possibilities to administer the second pulse. The effect of

these pulses on the systems dynamics was simulated.

The experimental design to be evaluated data input scheme for $u_1(t)$ is taken to fulfill (3.20a)-(3.20c) for practical reasons. The sampling time steps for example have to be restricted to 20 or 30 minutes. Gaussian noise was added to the simulated noise-free data set, leading us to the matrix of time course gene transcription for the genes. The design evaluations involved parameter estimation by considering as if the synthetic data were real. The parameter estimates and their corresponding confidence bounds obtained. This together with the parameter sensitivity results and the modified E values were used for the design performance assessment.

3.2.5 Experimental strategy - double D-xylose pulse

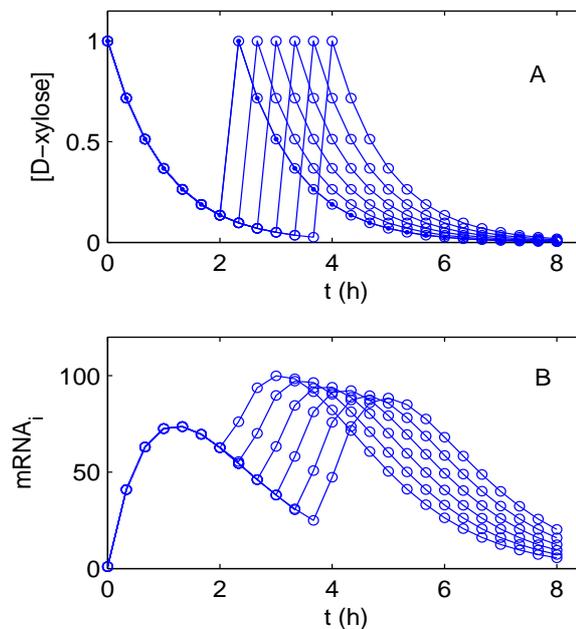


Figure 3.3: Simulation of a 2-pulse experiment. D-xylose concentration and the response of one gene to 2 pulse experiments with 6 different moments to administer the second pulse. A: D-xylose concentration, B: Noise-free expression profile of mRNA level in time. Nominal parameter values $k_{is} = 257$, $k_{i1} = 1.50$ and $k_{id} = 1.07$ (derived from the *xlnB* gene of the XlnR regulon) were used for data simulation.

From the *in silico* experimental data, model fits were performed to establish the right moment in time to administer the second pulse. This involves monitoring the Input-Output (or D-xylose - mRNA) data measurements. This choice of design ensures that when the gene expression levels begin to reduce, a pulse will once again enhance the mRNA levels. Figure 3.3 shows that an early pulse gives a stronger peak expression than a late one. We aim to find the most informative pulse combinations; therefore, the need to evaluate the influence of the second pulse on the identification.

3.2.6 Evaluation of parameter sensitivities

Since the parameters and state variables of a given biological network may span a wide range of values, sometimes it is preferable to use normalized sensitivities for comparing the parameter sensitivities. The normalized sensitivities are defined by

$$S_i^{\text{norm}} := (x_{i,\vartheta_r} / \bar{x}_i) \vartheta_r^{\text{nominal}} \quad (3.21)$$

Using the normalized sensitivity functions ensure that a fair comparison for the various design strategies. The ranking metric is computed from the normalized parameter sensitivity function

$$\Phi_{i,\vartheta_r} = \int_{t_0}^{t_f} \| S_i^{\text{norm}}(t) \|_1 dt \quad (3.22)$$

This integral sensitivity metric is based on time integrals [95]. The metric in (3.22) was used for comparisons of the parameter sensitivities from the expressions in (3.7)-(3.10) and (3.11)-(3.13).

Higher parameter sensitivities warrant better identification. To evaluate the results, an evaluation criterion was specified:

$$\mathbf{J}_{\text{TCE}i} := \sum_{r=1}^q \Phi_{i,\vartheta_r} \quad (3.23)$$

This expression is indicative of the cumulative area between the time axis and the individual sensitivity functions.

The variables to be evaluated are: (i) the D-xylose pulse strength, and (ii) the time instant of the second pulse administration. This evaluation aims to improve the kinetic parameter identification. To assess the performance of the design strategies, simulations were performed for various configurations of the single and double D-xylose pulses. Here, the aim is to identify the design strategies with the highest cumulative sensitivities.

3.2.7 Software used

The modeling, simulation and parameter estimation were all performed using the MATLAB[®] Version 7.12.0 (R2011a) software under the Windows[®] 7 environment. The *lsqnonlin* routine was used for the model fit to data. Overall, the time needed for data simulations and parameter identification ranged from a few seconds to a few minutes.

3.3 Results

3.3.1 Low D-xylose pulse experiments

In this section we present results from the simulations of the various time course experiments using pulse strategies described in the methods section. A set of

Chapter 3. Evaluation of design strategies for time course experiments in genetic networks: case study of the *xlnR* regulon in *Aspergillus niger*

parameters was considered as nominal, and used as the true values that describe the systems dynamics. These parameter values were obtained from a model fit to real time course experiment data for the *xlnR* gene obtained from a qPCR experiment. Parameter sensitivities for the low D-xylose concentrations (the three parameter model, (3.5)) were computed and a typical outcome is illustrated in Figure 3.4.

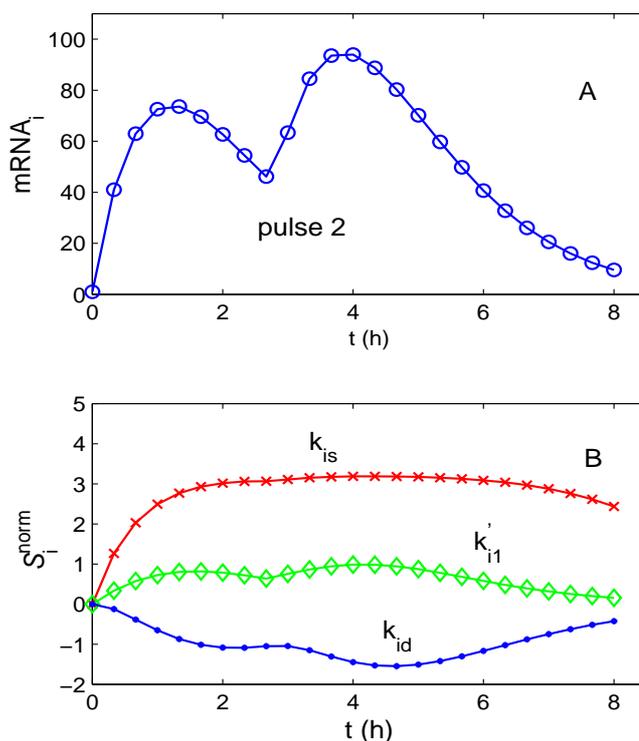


Figure 3.4: A: Plot of gene expression profile for which the D-xylose pulse is given at time point $j = 9$. B: Plots of normalized parameter sensitivities. The change in normalized sensitivities is significantly different for all the parameters.

By considering the normalized parameter sensitivities, the results in Figure 3.4 B show that the most sensitive (or easily identified) parameter is k_{is} followed by k_{id} and k'_{i1} , respectively. For each response in Figure 3.3, parameter sensitivities are obtained and the design performance evaluated (Table 3.1). For all these designs, the modified E-criterion has quite large values. Generally, it is desired to have values of modified E-criterion close to 1. The experiment with a second pulse at $j = 9$ has the lowest modified E value. Despite the high value, the preference was given to the experiment with a second pulse at $j = 9$.

Table 3.1 also shows that the modified E-criterion for a double trigger experiment at the different time instants $j = 7, \dots, 12$ have a significantly different value from that of the single pulse at $j = 1$. This shows that applying a second

3.3. Results

Single pulse at	E_{modified}	$J_{\text{TCE}i}$	$\Phi_{k_{is}} > \Phi_{k_{id}} > \Phi_{k'_{i1}}$
$j = 1, (t = 0)$	999.669	29.297	18.406 > 8.017 > 2.874
Second pulse at	E_{modified}		$\Phi_{k_{is}} > \Phi_{k_{id}} > \Phi_{k'_{i1}}$
$j = 7, (t = 2.0)$	578.564	36.637	23.365 > 8.017 > 5.255
$j = 8, (t = 2\frac{1}{3})$	582.575	35.876	22.722 > 7.955 > 5.199
$j = 9, (t = 2\frac{2}{3})$	541.848	35.886	22.780 > 7.892 > 5.214
$j = 10, (t = 3.0)$	574.344	35.711	22.680 > 7.789 > 5.242
$j = 11, (t = 3\frac{1}{3})$	595.379	36.284	23.259 > 7.729 > 5.296
$j = 12, (t = 3\frac{2}{3})$	625.022	35.462	22.689 > 7.609 > 5.164

Table 3.1: Design attributes used for selection of the most informative experimental design strategy. The design with the least E_{modified} value was considered as the best choice - here shown highlighted in gray. Data used with $\varepsilon_i = 0.05$ measurement noise.

pulse improves the information content of the data set and the overall design. The order in which the normalized parameter sensitivities are ranked is invariant of the time at which the second pulses were given (Table 3.1).

To test the performance of the identification technique used, the simulated data sets were used for parameter estimation. Using the simulated data sets, the parameter estimates for the two extreme pulse designs are shown in Table 3.2. The results in this table show that, despite the parameter correlation, for both experimental approaches the correct values are accurately found. The experiments with the second pulse at time instant $j = 9$ gives parameter values close to the true values and have the smallest Coefficient of Variation (c_v , see section 3.2.3) values (Table 3.2). This finding supports the results from the design evaluation procedure in Table 3.1.

Second pulse at	$\hat{\vartheta}_r \pm \hat{\sigma}_{\hat{\vartheta}_r}$ for $\varepsilon_i = 0.05$	c_v
$j = 7,$ $(t = 2.0)$	$\hat{k}_{is} = 256.70 \pm 60.58$	0.236
	$\hat{k}'_{i1} = 1.56 \pm 0.79$	0.506
	$\hat{k}_{id} = 1.08 \pm 0.08$	0.074
$j = 9,$ $(t = 2\frac{2}{3})$	$\hat{k}_{is} = 255.37 \pm 18.58$	0.073
	$\hat{k}'_{i1} = 1.49 \pm 0.02$	0.011
	$\hat{k}_{id} = 1.071 \pm 0.004$	0.004
$j = 12,$ $(t = 3\frac{2}{3})$	$\hat{k}_{is} = 250.94 \pm 37.55$	0.150
	$\hat{k}'_{i1} = 1.75 \pm 0.52$	0.297
	$\hat{k}_{id} = 1.16 \pm 0.03$	0.026

Table 3.2: Results from a double pulse experiment in which the second pulse is given at a selected set of time instants $j = 7, 9, 12$. The parameter estimates were performed at $p = 0.05$ and c_v is the coefficient of variation. The nominal parameter values were $k_{is} = 257$, $k'_{i1} = 1.50$ and $k_{id} = 1.07$.

Second pulse	$\hat{\vartheta}_r \pm \hat{\sigma}_{\hat{\vartheta}_r}$ for $\varepsilon_i = 0.05$	c_v	$\mathbf{J}_{\text{TCE}i}$	Ranked normalized sensitivities $\Phi_{k_{i_s}} > \Phi_{k_{i_d}} > \Phi_{k'_{i_1}} > \Phi_{k_{i_2}}$	E_{modified}
1 mM	$\hat{k}_{i_s} = 259.73 \pm 329.87$ $\hat{k}'_{i_1} = 1.46 \pm 1.79$ $\hat{k}_{i_2} = 0.21 \pm 1.04$ $\hat{k}_{i_d} = 1.08 \pm 0.17$	1.270 1.226 4.952 0.157	21.923	8.371 > 7.555 > 5.321 > 0.676	2.371e + 5
5 mM	$\hat{k}_{i_s} = 259.95 \pm 92.90$ $\hat{k}'_{i_1} = 1.49 \pm 0.78$ $\hat{k}_{i_2} = 0.20 \pm 0.16$ $\hat{k}_{i_d} = 1.07 \pm 0.02$	0.357 0.523 0.800 0.090	20.14	8.135 > 6.866 > 3.540 > 1.599	3.760e + 3
10 mM	$\hat{k}_{i_s} = 259.91 \pm 103.28$ $\hat{k}'_{i_1} = 1.48 \pm 0.96$ $\hat{k}_{i_2} = 0.20 \pm 0.10$ $\hat{k}_{i_d} = 1.08 \pm 0.07$	0.397 0.649 0.500 0.065	19.726	8.033 > 6.532 > 2.737 > 2.424	1.121e + 3
25 mM	$\hat{k}_{i_s} = 261.76 \pm 595.78$ $\hat{k}'_{i_1} = 1.48 \pm 2.36$ $\hat{k}_{i_2} = 0.20 \pm 0.91$ $\hat{k}_{i_d} = 1.08 \pm 0.14$	2.276 1.595 4.550 0.130	21.265	8.021 > 7.142 > 3.145 > 2.957	3.039e + 4
50 mM	$\hat{k}_{i_s} = 284.80 \pm 1866.20$ $\hat{k}'_{i_1} = 0.80 \pm 6.10$ $\hat{k}_{i_2} = 0.40 \pm 2.70$ $\hat{k}_{i_d} = 0.70 \pm 0.40$	6.553 7.625 6.750 0.571	22.015	8.190 > 7.499 > 3.635 > 2.691	2.340e + 4

Table 3.3: Design attributes for the two-pulse experiment with varying concentrations of D-xylose administered at the second pulse. Result from the parameter estimates at $p = 0.05$ and $N = 25$ samples in time. The second pulse of D-xylose was given at time instant $j = 12$ when the D-xylose concentration from the first pulse was approximately 0.1 mM. The term c_v is the coefficient of variation. The best design strategies are shown highlighted in gray. The nominal parameter values were $k_{i_s} = 257$, $k'_{i_1} = 1.50$, $k_{i_2} = 0.20$ and $k_{i_d} = 1.07$.

3.3.2 Experiments with varying pulse strength

The results from the following experiment provide insight into the effect of varying the strength of the second pulse. First, we present the sensitivity results followed by the design performances. The varying D-xylose concentrations on the second pulse can be given at any moment in time. By considering the 1 mM double pulse *in silico* experimental set-up, it was found that a second pulse given at the point at which the D-xylose concentration from the first pulse drops to about 0.1 mM was most effective (see Figure 3.3). This concentration of D-xylose discriminates between non-induced and induced states of gene activity [35]. Results for the corresponding design performances are shown in Table 3.3. Again the modified E-criterion from this table has high values as a consequence of the parameter correlation in the model. Nevertheless, experiments with a 5 and 10 mM second pulse are to be preferred.

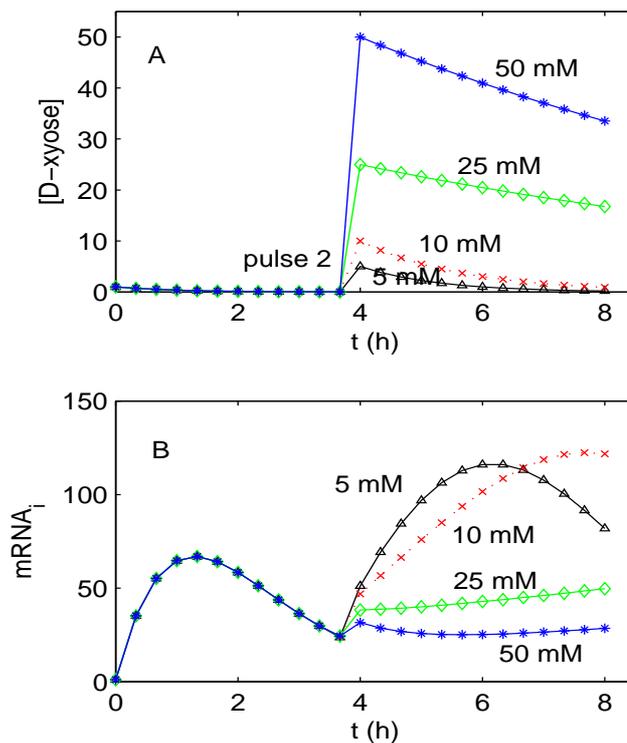


Figure 3.5: A: Noise-free plot of D-xylose concentration in time with various pulse strength, second pulses given at $j = 12$. B: Plot of corresponding gene expression levels in response to the pulse in A. The gene transcription level is more repressed with increasing D-xylose concentration.

Figure 3.6 illustrates how the parameter sensitivities change as a result of administration of the second D-xylose pulse. We see that the sensitivities of the parameters are strongly affected by the second pulse of D-xylose. The influence of the effect of the second pulse is stronger than when 1 mM was used in the

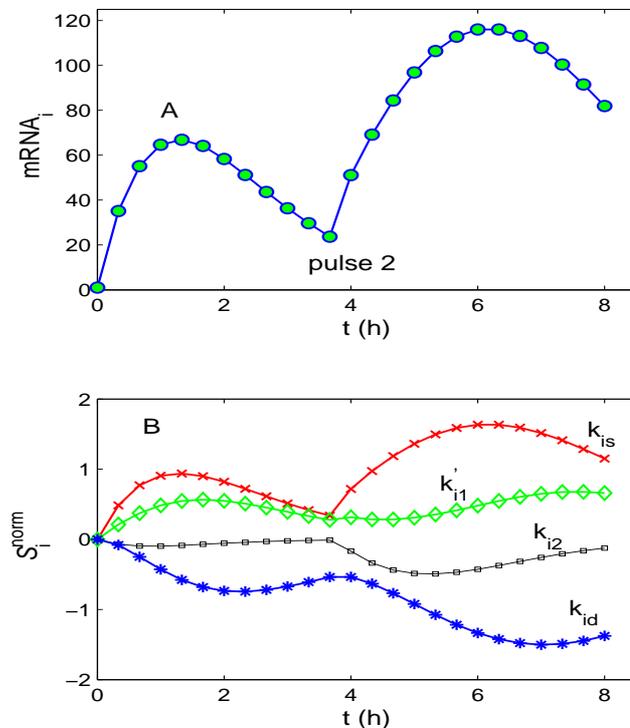


Figure 3.6: A: Noise-free plot of gene expression profile for which the second D-xylose pulse was given at time point $j = 12$ for the 5 mM D-xylose. B: Plots of normalized sensitivities functions for the various parameters. The second D-xylose pulse re-enforces the parameter sensitivities, particularly for k_{is} and k_{id} .

second pulse. Note that these D-xylose concentrations are just arbitrary and any other values could have been chosen for the induction. Parameter estimation results for the second pulse experiments with varying D-xylose concentrations are given in Table 3.3. All the kinetic parameters were accurately estimated. The experiments with a second pulse of 5 and 10 mM D-xylose are the most accurate and have the smallest coefficient of variation. The low gene response corresponding to the 25 and 50 mM D-xylose pulses are due to their repressing action through CreA (Figure 3.5).

3.3.3 Single versus double D-xylose pulses

Using the model (3.5), the results in Table 3.1 show that the two-pulse experiment is better than the single pulse experiment. In this section we verify if this result holds also for the model formulation (3.4). We compared the results from Table 3.3 with a single 1 mM pulse experiment (Table 3.4). The tables show that the two pulse experiments yield parameter values closer to the true values and with smaller confidence intervals. Due to the CreA repression the 50 mM experiment is an exception to this observation. The obtained parameter

Pulse strength	$\hat{\vartheta}_r \pm \hat{\sigma}_{\hat{\vartheta}_r}$ for $\varepsilon_i = 0.05$	c_v	E_{modified}
1 mM	$\hat{k}_{is} = 255.94 \pm 907.09$	3.544	$2.416e + 5$
	$\hat{k}'_{i1} = 1.25 \pm 5.59$	4.472	
	$\hat{k}_{i2} = 0.08 \pm 1.92$	24.000	
	$\hat{k}_{id} = 1.00 \pm 0.36$	0.360	

Table 3.4: Example of parameter estimates performed at a statistical significance level of $p = 0.05$, for a data set in which only a single pulse of D-xylose was administered at $t_0 = 0$ hours. The term c_v is the coefficient of variation. The nominal parameter values were $k_{is} = 257$, $k'_{i1} = 1.50$, $k_{i2} = 0.20$ and $k_{id} = 1.07$.

estimates have very large standard deviations. A comparison of these tables reaffirms that the second pulse experiment significantly improves the quality of the estimated parameters - especially at low (5 and 10 mM) D-xylose concentrations. Observing from the Coefficient of Variation (c_v), the accuracy and precision levels of the parameters differ greatly with k_{i2} and k'_{i1} being the worst identified.

3.3.4 Number of samples and inter-sampling time

Time course experiments with more data points during an experiment can be beneficial for precision and can be considered if resources are available [96, 97]. However, it is not a guarantee that the accuracy on the parameter estimates improves proportionally. In the examples in this paper, we consider a sufficiently and realistically large number of data points for the identification and sensitivity analysis. Even with correlation between k_{is} and k'_{i1} , most of the parameters were recovered accurately. Having assessed the performance of the designs strategies, we considered redesigning the experiment with 10 mM D-xylose as a second pulse. The total experimental time is planned for the same 8 hour duration but now with a sampling interval of 30 minutes (total of $N = 17$ samples).

Second pulse	$\hat{\vartheta}_r \pm \hat{\sigma}_{\hat{\vartheta}_r}$ for $\varepsilon_i = 0.05$	c_v	E_{modified}
10 mM	$\hat{k}_{is} = 256.75 \pm 50.79$	0.198	$1.404e + 3$
	$\hat{k}'_{i1} = 1.54 \pm 0.46$	0.299	
	$\hat{k}_{i2} = 0.22 \pm 0.08$	0.364	
	$\hat{k}_{id} = 1.00 \pm 0.01$	0.010	

Table 3.5: Identification results with a second pulse of D-xylose given at the fourth hour of experimentation ($j = 9$) at a statistical significance level of $p = 0.05$, sampling interval 30 minutes and $N = 17$. The term c_v is the coefficient of variation. The nominal parameter values were $k_{is} = 257$, $k'_{i1} = 1.50$, $k_{i2} = 0.20$ and $k_{id} = 1.07$.

Table 3.5 shows that the values of the E_{modified} criteria are close to each other

Chapter 3. Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*

(for the 20 minutes sampling interval $E_{\text{modified}} = 1.121e + 3$ and for the 30 minutes interval $E_{\text{modified}} = 1.404e + 3$) for the data set with measurement noise $\varepsilon_i = 0.05$, an increase that is not that very much significant. The benefit of considering an experimental design with a wider data inter sampling distance compared to experiment with the sampling interval of 20 minutes is that it has reduced costs, requires less number of samples and more sample handling time.

3.3.5 Comparing criteria for experimental design

Different D-xylose pulse combinations were simulated and the design evaluation of the parameter sensitivities was performed using (3.23). A summary of the parameter sensitivity evaluation and parameter identification results is shown in Table 3.3. The results show that for all the inducing D-xylose concentrations the coefficient of variation for the mRNA degradation parameter k_{id} is the lowest. However, it must be noted that the sensitivity of the parameter k_{id} is highly affected by the accuracy in the measurement of the state variable x_i , see expressions (3.10) and (3.13).

Looking at the ranked normalized sensitivities in Table 3.3 we see that the parameters k_{is} and k_{i2} have the highest and lowest sensitivity values, respectively. In this table we observed that changing the level of the second D-xylose pulse does not significantly change the sensitivity of the mRNA synthesis parameter k_{is} . The standard deviations of the parameters \hat{k}_{is} and \hat{k}'_{i1} are increasingly inflated in experiments with higher D-xylose concentration on the second pulse. The experimental scheme with the 5 and 10 mM D-xylose as a second pulse had the lowest coefficient of variation values $c_v < 1.0$ for all parameters. This shows the benefit of using the 5 and 10 mM D-xylose as a second pulse to trigger the XlnR regulon.

Why obtaining accurate values for all the estimated parameters may be a challenge can be explained by the results observed for the sensitivity integrals. The results show that an increase (or decrease) in sensitivity value for one parameter may coincide with a decrease (or increase) in sensitivity for another parameter, respectively. For example for the 5 mM second pulse $\Phi_{k'_{i1}} = 3.540$ and $\Phi_{k_{i2}} = 1.599$, while for the 10 mM second pulse $\Phi_{k'_{i1}} = 2.737$ and $\Phi_{k_{i2}} = 2.424$. This is particularly true for a set of correlated parameters as can be seen from the values of the "Ranked normalized sensitivities" column in Table 3.3 for the different D-xylose concentrations. The decrease, for instance, in integral sensitivity of k'_{i1} over the first three rows in Table 3.3 is accompanied by an increase in integral sensitivity of k_{i2} .

All the sums of sensitivity values in Table 3.3 fall in the range $\mathbf{J}_{\text{TCE}i} = 19.726$ to $\mathbf{J}_{\text{TCE}i} = 22.015$ which shows that there is little preference for one experimental scheme over the other. However, by using the modified E-criterion it is easier to decide on which design is to be preferred. For instance, the 10 mM D-xylose experiment yielded the best parameter estimates which coincide with the lowest value for the modified E-criterion ($E_{\text{modified}} = 1.121e + 3$), therefore, the modified E evaluation criteria is to be preferred to the sum of sensitivity functions.

We could not find any literature information with similar results that confirms the presence of the modified E-criterion as compared to the sum of sensitivity functions, but we suspect that the difference is caused by the presence of parameter estimate correlations, to which, by design, the modified E-criterion is directly sensitive in contrast to the sum of sensitivity integrals.

3.4 Discussion and conclusions

Experimental design is a valuable tool to improve the design of time course experiments with the goal to maximize information gain and improve identification. To successfully design a gene expression time course experiment, like that of the XlnR regulon we are faced with the following issues: (i) parameters and model are needed but the design is intended to help obtain these, (ii) single regulon, limited number of parameters, low complexity in our study. Nevertheless, with these limitations in mind, through strategic experimental designs, costs can be reduced by taking fewer samples, informative data sets can be obtained and the identification of the network kinetic parameters can be improved.

This work illustrates that the parameter estimates are improved by using a second inducer pulse instead of just using a single pulse. Overall, the use of the second pulse improved the experimental design. The second inducer pulse should be given at an effective moment in time; in this case when the inducer concentration became low. The results also indicate that for time course experiments, data sampling intervals can be optimized with respect to the accuracy of the parameters, available resources and labor costs.

Regarding the criteria used for experimental design, this work suggests that the use of the modified E-criterion, which encompasses the effects of estimation correlations resulting from the design, is better than basing the design on comparison of parameter sensitivities.

This work was focused on the XlnR regulon of *A. niger*. In the multi-pulse experiments, giving a second D-xylose pulse of 10 mM yields better identification results. Now we are faced with the challenge of extending these results to studying the multiple target gene problem in the XlnR regulon.

Despite the potential benefits of OED in Systems Biology, there are some limitations. Successful application of OED requires the presence of a model and a nominal parameter set, but to obtain these is the very purpose of the experimental design. Hence, OED can best be applied in an iterative framework, where first preliminary experiments are performed based on biological insight to set up an early model, after which OED is employed to design further informative experiments.

Acknowledgments. This work is supported by the VLAG graduate school and the IPOPOP program of Wageningen University. We thank Birgit Jovanovic for her assistance in the HPLC analysis.

Chapter 3. Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*

Competing interests. The authors declare that they have no competing interests. All authors read and approved the final manuscript.

Appendix A

A. 1 Greek symbols

β	shape parameter.
$0 < \gamma_i \leq 1$	parameter sensitivity weighting coefficient.
Γ	set of constraint variables.
ε_i	Gaussian distributed measurement noise.
Θ_i	residual covariance matrix.
λ_{min}	minimum eigenvalue of Fisher Information Matrix.
λ_{max}	maximum eigenvalue of Fisher Information Matrix.
Λ_i	square weighting matrix (inverse of measurement error covariance matrix).
ϑ	is a parameter vector.
ϑ_r	is an element of the parameter vector.
$\vartheta_r^{nominal}$	nominal parameter value.
$\xi_i = y_i - \hat{y}_i(\hat{\vartheta})$	residual vector from model fit to data.
ρ_1	basal transcription rate for the <i>xlnR</i> gene.
ρ_i	target gene basal transcription rate.
Σ_i	parameter covariance matrix.
ℓ	index of experiment simulation run.
ϕ_i	function describing the dynamical system.
Φ_{i,ϑ_r}	integral of parameter sensitivity function over time.

A. 2 General terms

b_1	external input (stimulus) coefficient.
$c_v = \hat{\sigma}_{\hat{\vartheta}_r} / \hat{\vartheta}_r$	Coefficient of Variation.
det	matrix determinant.
δt	sampling time interval (h).
$\mathbb{E}(\bullet)$	expectation operator.
$E_{modified}$	modified E-criterion design strategy.
FIM	Fisher Information Matrix.
k_{is}	mRNA synthesis parameter.
k'_{i1}	affinity constant for gene 1 activating gene i (mM^{-1}).
k_{i2}	inverse of Hill constant for the repressive effect of higher D-xylose concentrations (mM^{-1}).

A. 2 General terms

k_{id}	mRNA degradation parameter.
K	is the D-xylose depletion parameter.
$i = 1$	indexes the master regulator gene (x_{lnR}).
$i = 2, \dots, n$	target gene index.
I	identity matrix.
$j = 1, \dots, N$	is the discrete time index.
\mathbf{J}_{TCEi}	sum of parameter sensitivities corresponding to a time course experiment for target gene i .
\mathcal{J}_i	model fit to data goal function.
$J_i \in \mathbb{R}^{N \times q}$	resulting Jacobian matrix from model fit.
n	total number of genes in a network.
N	maximum number of samples taken in time.
p	statistical p -value.
q	total number of parameters to be estimated.
$r = 1, \dots, q$	parameter index.
$S_{\bullet}(t)$	parameter sensitivity function.
$S_i^{\text{norm}} := (x_{i,\vartheta_r} / \bar{x}_i) \vartheta_r^{\text{nominal}}$	normalized parameter sensitivity function.
t	time in hours.
t_0	time zero (time of first induction).
t_f	final time point, in continuous time domain.
TCE	time course experiment
Tr	matrix trace.
Pulse ₁ , Pulse ₂	first and second pulse indicators, respectively.
$t_{\text{Pulse}_2(j)}$	time at which second pulse is administered (h).
t_L, t_R	left and right window of second pulsing time instant, respectively.
u_1 or [D-xylose]	D-xylose concentration (in mM).
$u_{1,min}, u_{1,max}$	minimum and maximum D-xylose pulse, respectively.
x_i	gene activity state.
$\dot{x}_i = dx_i/dt$	transcription rate.
\bar{x}_i	mean of gene activity.
$x_{i,\vartheta_r} := \partial x_i / \partial \vartheta_r$	partial derivative of the state vector w.r.t a parameter.
y_i	vector of measured gene expression.
\hat{y}_i	vector of estimated gene expression.
\mathbb{R}	matrix space.
$\ \bullet\ _1$	absolute value (Manhattan or L_1 norm).

**D-xylose
concentration-dependent
hydrolase expression profiles and
the according role of CreA and
XlnR in *A. niger***

D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*

Abstract.

Aspergillus niger is an important organism for the production of industrial enzymes such as hemicellulases and pectinases. The xylan-backbone monomer, D-xylose, is an inducing substance for the coordinate expression of a large number of polysaccharide-degrading enzymes. In this study, the responses of 22 genes to low (1 mM) and high (50 mM) D-xylose concentrations were investigated. These 22 genes encode enzymes that function as xylan backbone-degrading enzymes, accessory enzymes, cellulose-degrading enzymes, or enzymes involved in the pentose catabolic pathway in *A. niger*. Notably, genes encoding enzymes that have a similar function (e.g., xylan backbone-degradation) respond in a similar manner to different concentrations of D-xylose. Although low D-xylose concentrations provoke the greatest change in transcript levels, in particular, for hemicellulase-encoding genes, transcript formation in presence of high concentrations of D-xylose was also observed. Interestingly, a high D-xylose concentration is favorable for certain groups of genes. Furthermore, the repressing influence of CreA on the transcription and transcript levels of a subset of these genes was observed regardless of whether a low or high concentration of D-xylose was used. Interestingly, the decrease in transcript levels of certain genes in high D-xylose concentrations, is not reflected by the transcript level of their activator, XlnR. Regardless of the D-xylose concentration applied and whether CreA was functional, *xlnR* was constitutively expressed at a low level.

This chapter is based on:

Astrid R. Mach-Aigner, **Jimmy Omony**, Birgit Jovanovic, Anton J.B. van Buxtel, Leo H. de Graaff. D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*. *Applied and Environmental Microbiology*, 78(9), pp. 3145-3155, 2012.

4.1 Introduction

Aspergillus niger is a filamentous ascomycete that degrades plant biomass wherever available due to its saprophytic activity. Therefore, this fungus produces a broad spectrum of hydrolytic enzymes; examples of enzymes that degrade the xylan backbone are endoxylanases and a β -xylosidase as well as the accessory enzymes that remove the modifications of the backbone residues. The sugars that are released after degradation are mostly pentoses such as D-xylose and L-arabinose that are metabolized intracellularly. Table 4.1 summarizes the *A. niger* enzymes and their encoding genes that are involved in the degradation of xylan and metabolism of the D-xylose that is released.

Some of these enzymes are produced in industrially relevant amounts; *A. niger* is used as a native production host for enzymes such as pectinases [98, 99, 100] and hemicellulases like xylanases or arabinases [101, 102]. Consequently, the induction of hemicellulase expression using inexpensive substances is an important issue. In the past, the xylan-backbone monosaccharide, D-xylose, was shown to trigger expression of xylanases and D-xylose metabolizing enzymes in *A. niger*. Therefore, D-xylose is commonly used to induce the expression of these enzymes [35, 103, 104, 105]. Moreover, the co-regulation of the xylanolytic and the cellulolytic system via the inducer D-xylose has been reported [106, 107, 108].

The *A. niger* protein XlnR is a binuclear zinc finger protein [109] that belongs to a class of transcription factors specific for fungi, such as the *Saccharomyces cerevisiae* protein GAL4p [110]. XlnR functions as a transactivator responsible for a wide range of target genes such as those encoding xylan-degrading enzymes or the enzymes involved in D-xylose metabolism [34, 104, 105, 108], endo-cellulases and cellobio-hydrolases [107], and some of the galactosidase-encoding genes [111]. Within the pool of possible XlnR target genes, only the *bglA* gene was clearly shown to not be under the control of XlnR [107]; there are contradicting reports for the *xkiA* gene [104, 112]. Whereas XlnR functions as the transactivator of polysaccharide-degrading enzyme expression, CreA has been described as a wide-domain regulator mediating carbon catabolite repression and is best studied in *A. nidulans* (reviewed by [113, 114] and [115]). In *A. niger*, repression of xylanase gene expression in the presence of glucose was correlated to CreA influence [22, 116]. In 1999, de Vries et al. reported that CreA also mediates carbon catabolite repression of genes involved in xylan degradation on D-xylose concentrations higher than 1 mM [19].

In this study, we investigated how genes encoding arabinoxylan-degrading enzymes that are classified by their particular activity (xylan backbone-degrading enzymes, accessory enzymes, cellulose-degrading enzymes, D-xylose metabolizing enzymes) respond to high and low D-xylose concentrations. We also investigated whether the response of enzymes within the group is homogenous and whether the response of different groups of enzymes changes over time.

Enzyme	EC no.	GH family no.	Gene	GenBank accession no.	Reference
Xylan backbone-modifying enzymes					
Endo-xylanase B	EC 3.2.1.8	11	<i>xlnB</i>	D38071	[117]
Endo-xylanase C	EC 3.2.1.8	10	<i>xlnC</i>	C5J411	M.M.C. Gielkens, J. Visser and L.H. de Graaff, unpublished data
β -xylosidase	EC 3.2.1.37	3	<i>xlnD</i>	Z84377	[102]
Xylan accessory enzymes					
α -L-arabinofuranosidase B	EC 3.2.1.55	54	<i>abfB</i>	L23502	[118]
α -galactosidase B	EC 3.2.1.22	27	<i>aglB</i>	Y18586	[111]
α -glucuronidase	EC 3.2.1.139	67	<i>aguA</i>	AJ290451	[119]
Acetyl xylan esterase	EC 3.2.1.72		<i>axeA</i>	A22880	[120]
Arabinoxylan arabinofuranohydrolase	EC 3.2.1.55	62	<i>axhA</i>	Z78011	[101]
Esterase A			<i>estA</i>	AY456379	
Ferulic acid esterase	EC 3.1.1.73		<i>faeA</i>	Y09330	[121]
β -galactosidase	EC 3.2.1.23	35	<i>lacA</i>	L06037	[111]
Cellulose-degrading enzymes					
β -glucosidase	EC 3.2.1.21	3	<i>bglA</i>	AF121777	[34]
Cellobiohydrolase A	EC 3.2.1.91	7	<i>cbhA</i>	AF156268	[107]
Cellobiohydrolase B	EC 3.2.1.91	7	<i>cbhB</i>	AF156269	[107]
Endoglucanase A	EC 3.2.1.4	12	<i>eglA</i>	AJ224451	[34]
Endoglucanase B	EC 3.2.1.4	5	<i>eglB</i>	AJ224452	[34]
Endoglucanase C	EC 3.2.1.4	74	<i>eglC</i>	AY040839	
D-xylose metabolism enzymes					
L-Arabitol dehydrogenase	EC 1.1.1.12		<i>ladA</i>	AJ854040	[122]
Transaldolase B	EC 2.2.1.2		<i>talB</i>	XM001391362	
Xylitol dehydrogenase	EC 1.1.1.9		<i>xdhA</i>	AJ854041	[112]
D-xylulose kinase	EC 2.7.1.17		<i>xkiA</i>	AJ305311	[123]
Xylose reductase	EC 1.1.1.21		<i>xyrA</i>	AF219625	[108]

Table 4.1: *A. niger* enzymes involved in the degradation of xylan.

Gene	Primer sequence (5' → 3')	
	Forward	Reverse
<i>abfB</i>	GATAGTCCACCACGCTGAAGG	CTGCTTCGTGCCATCGTTGG
<i>aglB</i>	GATATCACGGCAACCTGGTCC	CTTCATCATCGCCCAAAGGG
<i>aguA</i>	CACAGTTCGGATCTTAGGCAAGG	GTCCATGTTGTCCCATTGATTCACC
<i>axeA</i>	CTATGATGACGAACGTAATGGCAGC	GACGTCTCCTTGTGCGCAG
<i>axhA</i>	CTCCAGCACTTTCACCTACCG	CACAGTCTGATCAATGGCACCG
<i>bglA</i>	GACCGCCAGTGTCTCTCTG	GATCGTGTTGTTGCAGTTGCTAGC
<i>cbhA</i>	CACTGGCAACACCTGGGATG	GTCAGTGGTGACACCGTAGG
<i>cbhB</i>	CAATGTGCGACGGTGACTCC	GTGATGAACTGGGTGACGACG
<i>eglA</i>	GTGGTGAGGGAACAGTAAAAGC	GACATCGGCGTTGACGTTGG
<i>eglB</i>	GACTGGTTCATACGACGAGGAG	GTATCAAACATGACCAAGTCGTTATC
<i>eglC</i>	ACGGACTACGGCGAGACATG	CACCACACCGACAAAGATGCG
<i>estA</i>	CAAAGTGCCACGGAGTACGG	GTAGTTTCGATTGAGATGTGGCTGTGG
<i>faeA</i>	ACCGTGACAGGCCATAGTCTG	GGAAGTACTGGGTCGTCTCC
<i>hist</i>	ATCTTGCGTGACAACATCCA	CACCCTCAAGGAAGGTCTTG
<i>lacA</i>	CTGGATGCTACAGATAATTACGTCTCTC	CATACTGCATGTACACTGGATCGG
<i>ladA</i>	GAATCTGCGGGTCCGACGTG	GCAAATGATGTTAGGCTCGACGG
<i>talB</i>	GCACCACAGTCGCCAAAACC	GGTATTCTTGGATGAGGGCTTTGG
<i>xdhA</i>	CAGGTATATGTGGTAGTGATGTTTCATTATTGG	CATTGCAACGCAGTCTCCTACC
<i>xkiA</i>	AGCAAGGCACACCATAGATTTACAGG	CATCAGAGATGTCCATCGGAGC
<i>xlnB</i>	CAACTTTGTCCGGTGGAAAGG	GGGTAGCCGTGTAGATATCG
<i>xlnC</i>	GTGGAGGTGCTGGAATTTCTGG	CAGGAATCCGGGTCAGCAAC
<i>xlnD</i>	TAATCTACGCCGGTGGTATC	TTCTTGAGCGAAGAGGAATC
<i>xlnR</i>	CTCATTGCATTGAATTCGGACTGACC	GCTGTCTCCGACGTTTCGC
<i>xyrA</i>	ACGGCGCCTGCGACTATG	TGCGGCAAATGGGCTCGAC

Table 4.2: Primers used for qPCR in this study.

Chapter 4. D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*

Furthermore, we addressed the question of whether the carbon catabolite repressing signal of D-xylose is mediated by CreA and whether this occurs via a concentration-dependent mechanism. Finally, we examined whether a D-xylose concentration-dependent effect on the transcript levels of target genes is reflected by the transcript level of their main activator encoding gene, *xlnR*, and if regulation of *xlnR* expression by CreA might be responsible for such an effect.

4.2 Materials and methods

4.2.1 Strains and growth conditions

The *A. niger* strains N400 (CBS 120.49) and NW283 (*cspA1*, *fwnA1*, *pyrA6*, *lysA7*, *creA^{d4}*) were maintained for generating spores on complete medium plates [124] supplemented with uridine and lysine. Strain NW283 exhibits a CreA - derepressed phenotype (*creA^{d4}* [d, derepressed; 4, allele 4]) and was previously used for studying CreA effects in *A. niger* [19, 63, 112, 125, 126]. Strain NW283 is a derivative of the NW145 strain (bearing additionally a *pyrA* marker) and is therefore a descendant of the N400 strain [114] to which it was consequently compared.

Cultivation of the strains in 2.2-liter benchtop fermenters (Applikon Biotechnology, Schiedam, The Netherlands) was carried out in duplicate using 2 l of medium at pH 6.0 supplemented with uridine and lysine and containing 1.2 g NaNO₃, 0.3 g KH₂PO₄, 0.1 g KCl, 0.1 g MgSO₄·7 H₂O, 2 g yeast extract, 4 g casamino acids, and 2 ml Vishniac solution [127]. The carbon source was 100 mM sorbitol. To avoid excessive foam formation, 1 ml antifoam (Sigma Aldrich, St. Louise, MO) was added to the medium. Inoculation of the cultures using 1×10^6 spores per ml, cultivation conditions, and induction with 1 mM or 50 mM D-xylose were performed as previously published [35]. After induction, the reference sample was taken within 30 s (preinduction sample, identical to the 0-h sample). After induction, samples were collected each hour (10 ml each, postinduction samples) and analyzed by microscope for infection control. Culture supernatant and mycelia were separated by filtration through Miracloth (Calbiochem, [part of Merck, Darmstadt], Germany). Mycelia were immediately snap-frozen in liquid nitrogen.

4.2.2 RNA extraction and reverse transcription (RT)

Harvested mycelia were homogenized in 1 ml of peqGOLD TriFast DNA/RNA/protein purification system reagent (PEQLAB Biotechnologie, Erlangen, Germany) using a FastPrep FP120 BIO101 ThermoSavant cell disrupter (Qbiogene, Carlsbad, CA). RNA was isolated according to the manufacturer's instructions, and the concentration was measured using a NanoDrop 1000 (Thermo Scientific, Waltham, MA).

After treatment with DNase I (Fermentas, part of Thermo Fisher Scientific,

St. Leon-Rot, Germany), cDNA was synthesized from 0.45 μg mRNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas,); all reactions were performed according to the manufacturer's instructions.

4.2.3 qPCR analysis

All quantitative PCRs (qPCRs) were performed in triplicate in a Rotor-Gene Q cycler (QIAGEN, Hilden, Germany). The amplification mixture (final volume 15 μl) contained 7.5 μl 2 \times Absolute QPCR SYBR Green Mix (ABgene, [part of Thermo Fisher Scientific], Cambridge, United Kingdom), 100 nM forward and reverse primer and 2.5 μl cDNA (diluted 1 : 100). Primer sequences are provided in Table 4.2. Each experiment included a template-free control and an amplification-inhibited control (0.015% SDS added to the reaction mixture). The cycling conditions comprised a 15 min initial polymerase activation at 95°C, followed by 40 cycles of 95°C for 15 s, 59°C for 15 s, and 72°C for 15 s. Initially, four previously used reference genes [35] have been compared to evaluate how stable their expression is under the applied conditions. Therefore, a synthetic control RNA transcript, a bacterial kanamycin synthetase-encoding gene fused to a eukaryotic poly(A) tail (Promega), was added to the total RNA prior to cDNA synthesis to correct for various efficiencies of the reverse transcription or PCR itself [128]. The differences in the expression of all four candidate reference genes were below 5% under all conditions. Consequently, the number of the genes used for normalization of the expression data was reduced to two, which were randomly chosen: the histone-encoding gene (termed "*hist*") and the gene with similarity to *Schizosaccharomyces pombe* gene *dma1*, which is a component of the spindle assembly point involved in mitotic division (short-termed "*spas*"). Calculations were performed using the equation from Pfaffl [129]: the transcript ratio equals $E_{\text{target}}^{\Delta C_t(\text{pre-post})_{\text{target}}} / E_{\text{ref}}^{\Delta C_t(\text{pre-post})_{\text{ref}}}$, where E denotes the amplification efficiency, and $\Delta C_t(\text{pre} - \text{post})_{\text{target}}$ denotes the difference in threshold cycles between pre- and post-induction samples.

4.2.4 Statistical analyses of qPCR data

The log-transformed data values of the relative transcript level ratios were used for the comparative data analysis. Sample standard deviations were computed for every data point for the entire data set using the averaged log-relative transcript level ratios from the duplicate experiments. The expression for computing the unbiased sample standard deviation is given by

$$\text{std}(\log(y_{ijd})) = \sqrt{\frac{1}{m-1} \sum_{d=1}^m (\log(y_{ijd}) - \overline{\log(y_{ijd})})^2}$$

where $\log(y_{ijd})$ and $\overline{\log(y_{ijd})}$ are the log-relative transcript level and mean log-relative transcript level ratio, respectively. The indices i , j and d represent the profiled genes, time instant and the duplicates, respectively. In this case, $j =$

Chapter 4. D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*

1, ..., $N(= 6)$ corresponds to the zero- to fifth-hour mark (where N is the total number of samples taken in time) and $d = 1, \dots, m(= 2)$. Overlapping standard error bars indicate no significant difference in transcript levels between groups. This approach provides a standard platform for comparing the transcript levels of the genes as a result of induction by 1 mM and 50 mM D-xylose. A similar analysis was used to compare gene transcript levels between the wild-type and CreA mutant strain.

4.2.5 HPLC analysis

Analyses were performed using a Thermo Finnigan Surveyor high-pressure liquid chromatography (HPLC) instrument (Thermo Fisher Scientific, MA). All 10- μ l samples were injected onto a Repro-Gel Pb column (9 μ m, 150 \times 8 mm; Dr. Maisch, Germany). Water was used as the mobile phase and elution was followed at 50°C at a flow rate of 1.0 ml/min for 20 min. The concentration was determined using xylitol as an internal standard.

4.3 Results

4.3.1 Response of xylan backbone-degrading enzyme expression to different D-xylose concentrations

A. niger strain N400 was cultivated in a benchtop fermenter and a low (1 mM final concentration) or high (50 mM final concentration) concentrations of D-xylose were added. Samples were taken immediately after the pulse and then every 60 min after the initial sample to compare transcript formation. The genes encoding endoxylanase B (*xlnB*) and endoxylanase C (*xlnC*) as well as β -xylosidase (*xlnD*) were analyzed by RT-qPCR. The use of 1 mM D-xylose leads to high transcript levels for all three genes after 1 h, with a slight decrease in the transcript levels of these genes after 2 h (Figure 4.1 A to C). HPLC analysis showed that D-xylose was not detectable after 2 h (Table 4.3). Interestingly, after 3 h, the transcript formation trend clearly differs: while considerable amounts of the *xlnB* transcript could be detected, which decreased at later a time point (Figure 4.1 A), *xlnC* transcript levels dropped significantly and then remained at a low constant level (Figure 4.1 B), and *xlnD* transcript formation was reduced below the preinduction level (Figure 4.1 C). These differences might be due to differences in the mRNA stability of each transcript. In contrast to 1 mM D-xylose, the addition of 50 mM D-xylose to the culture induces *xlnB* and *xlnC* at a constant level over the entire period investigated (Figure 4.1 D and E). However, transcription of *xlnD* is induced for only 2 h and then decreases to nearly preinduction levels (Figure 4.1 F). Because there are other genes known to be induced by D-xylose, we also investigated their response to different concentrations of D-xylose.

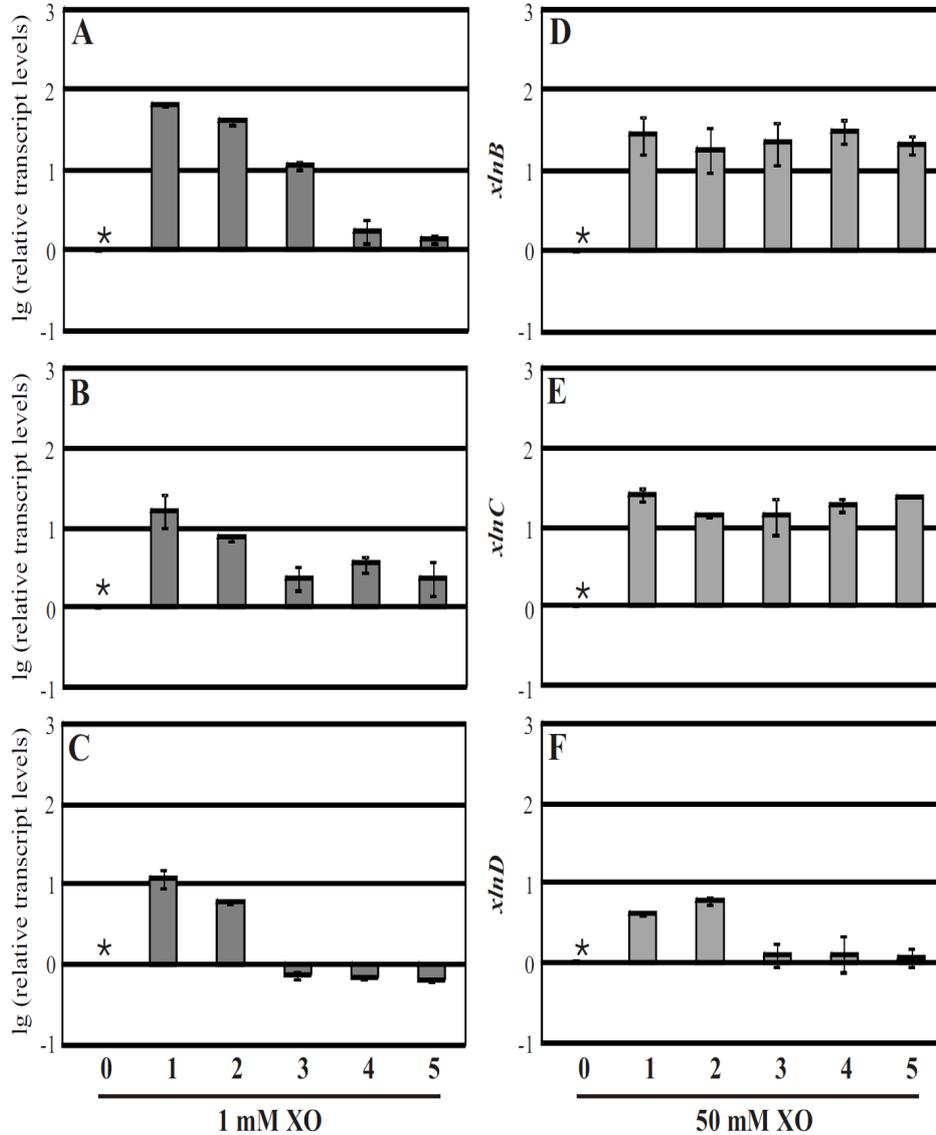


Figure 4.1: Transcript levels of genes encoding xylan backbone-degrading enzymes in *A. niger*. The strain N400 was induced using 1 or 50 mM D-xylose (XO) and samples were taken hourly for 5 h. Transcript level analysis of *xlnB* (A and D), *xlnC* (B and E), and *xlnD* (C and F) was performed by qPCR using a histone-like and a *Schizosaccharomyces pombe dma1*-like gene transcript for normalization. Transcript levels always refer to the reference sample (taken directly after induction) within an experiment, which is indicated by an asterisk. The results are presented as the relative transcript levels in logarithmic scale (*lg*). The values are means of two independent biological experiments measured in triplicate. Error bars indicate the standard deviation.

Chapter 4. D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*

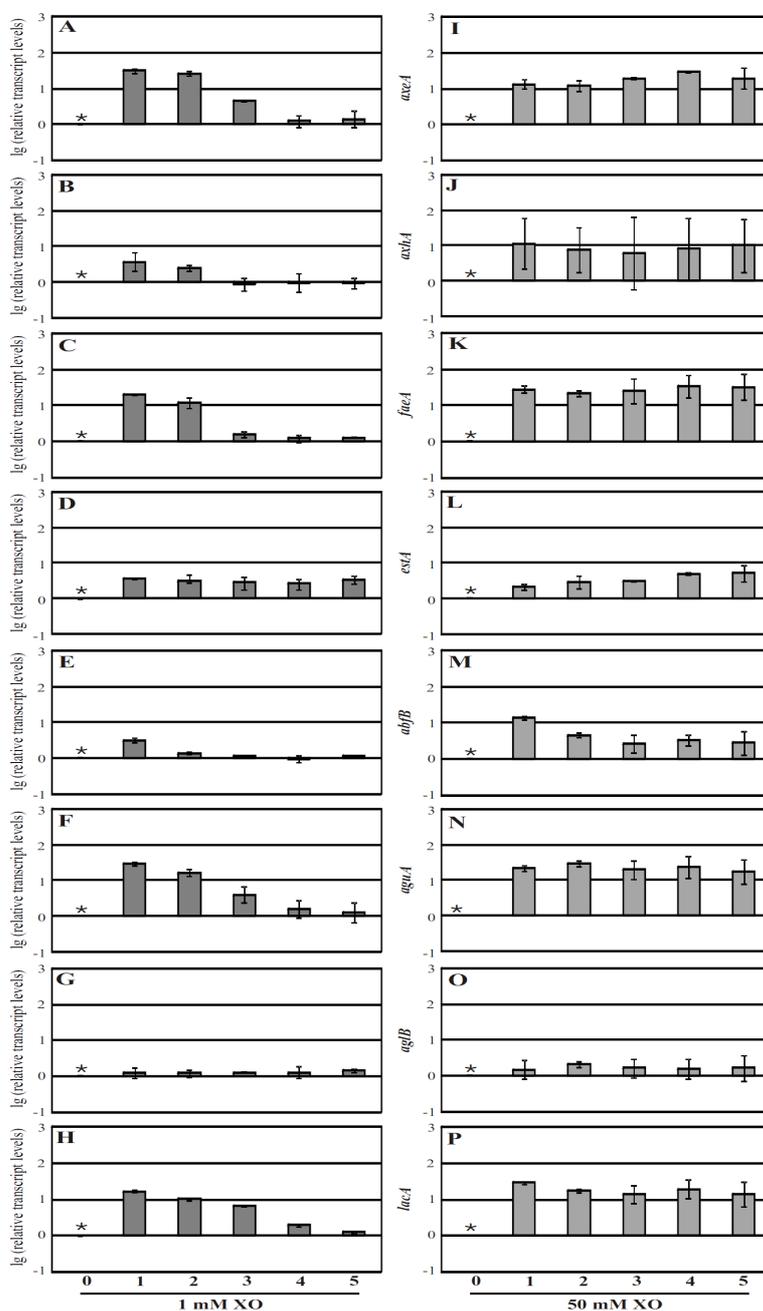


Figure 4.2: Transcript levels of genes encoding other accessory enzymes in *A. niger*. The strain N400 was induced using 1 or 50 mM D-xylose (XO) and samples were taken hourly for 5 h. Transcript level analysis of *axeA* (A and I), *axhA* (B and J), *faeA* (C and K), *estA* (D and L), *abfB* (E and M), *aguA* (F and N), *aglB* (G and O), and *lacA* (H and P) was performed by qPCR using a histone-like and a *Schizosaccharomyces pombe dma1*-like gene transcript for normalization. Transcript levels always refer to the reference sample (taken directly after induction) within an experiment, which is indicated by an asterisk. The results are presented as the relative transcript levels in logarithmic scale (*lg*). The values are means of two independent biological experiments measured in triplicate. Error bars indicate the standard deviation.

4.3. Results

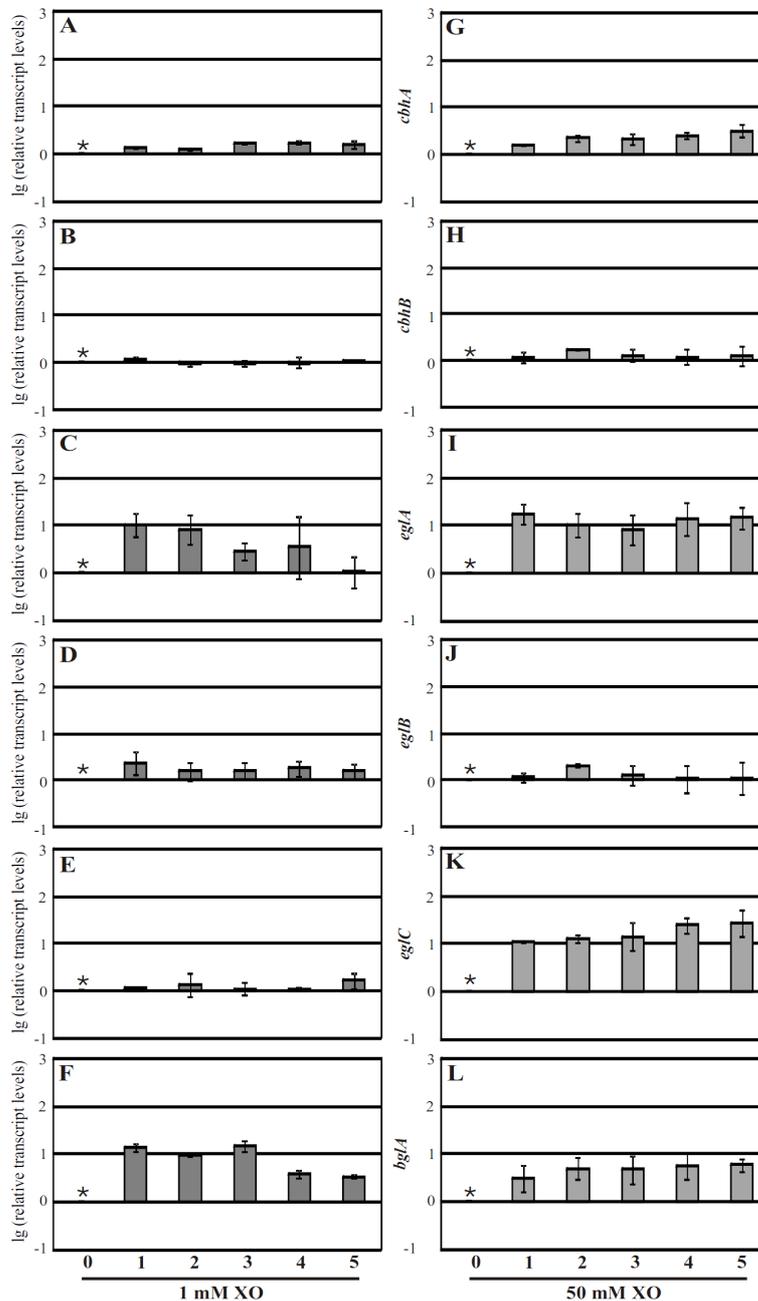


Figure 4.3: Transcript level of genes encoding cellulose-degrading enzymes in *A. niger*. The strain N400 was induced using 1 or 50 mM D-xylose (XO) and samples were taken hourly for 5 h. Transcript level analysis of *cbhA* (A and G), *cbhB* (B and H), *eglA* (C and I), *eglB* (D and J), *eglC* (E and K), and *bglA* (F and L) was performed by qPCR using a histone-like and a *Schizosaccharomyces pombe dma1*-like gene transcript for normalization. Transcript levels always refer to the reference sample (taken directly after induction) within an experiment, which is indicated by an asterisk. The results are presented as the relative transcript levels in logarithmic scale (*lg*). The values are means of two independent biological experiments measured in triplicate. Error bars indicate the standard deviation.

Chapter 4. D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*

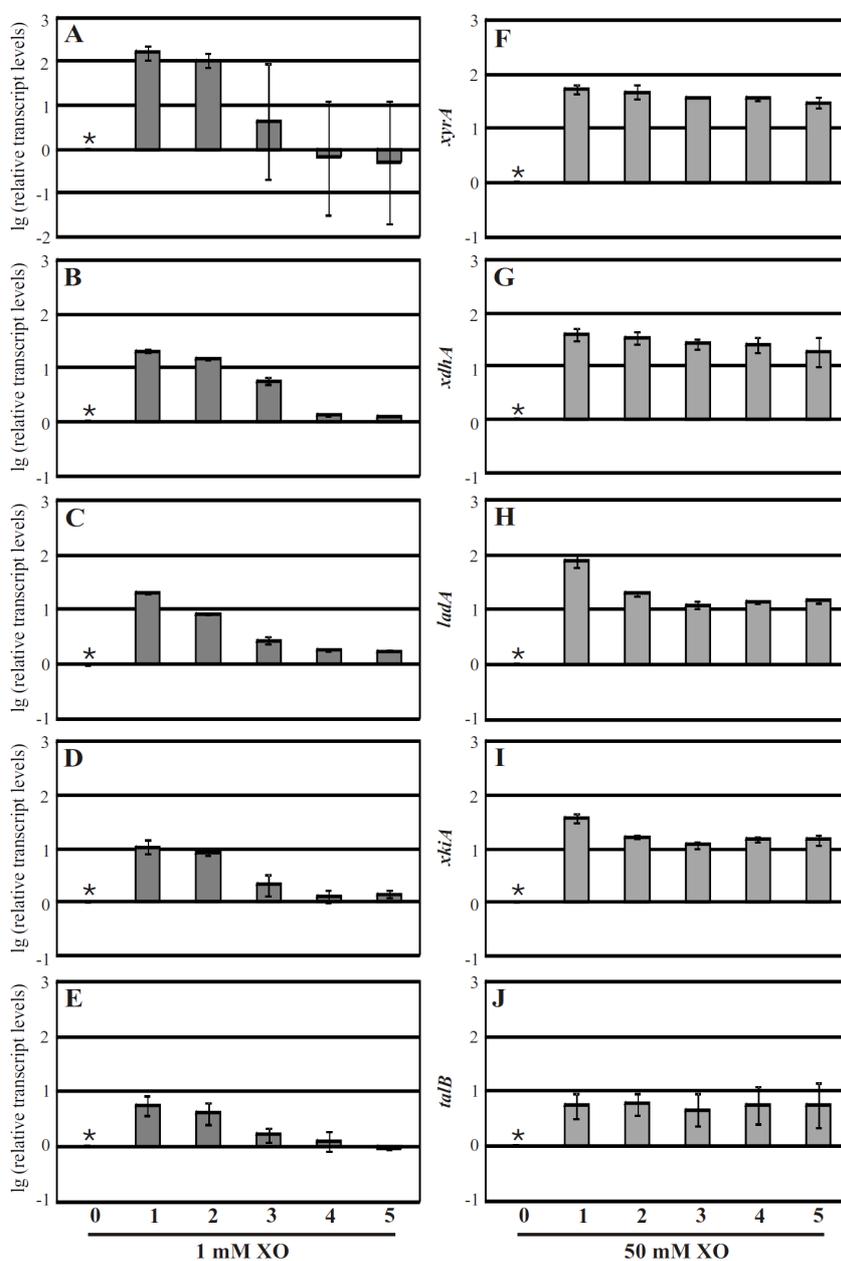


Figure 4.4: Transcript levels of genes encoding enzymes of the pentose metabolic pathway in *A. niger*. The strain N400 was induced using 1 or 50 mM D-xylose (XO) and samples were taken hourly for 5 h. Transcript level analysis of *xyrA* (A and F), *xdhA* (B and G), *ladA* (C and H), *xkiA* (D and I), and *talB* (E and J) was performed by qPCR using a histone-like and a *Schizosaccharomyces pombe dma1*-like gene transcript for normalization. Transcript levels always refer to the reference sample (taken directly after induction) within an experiment, which is indicated by an asterisk. The results are presented as the relative transcript levels in logarithmic scale (*lg*). The values are means of two independent biological experiments measured in triplicate. Error bars indicate the standard deviation.

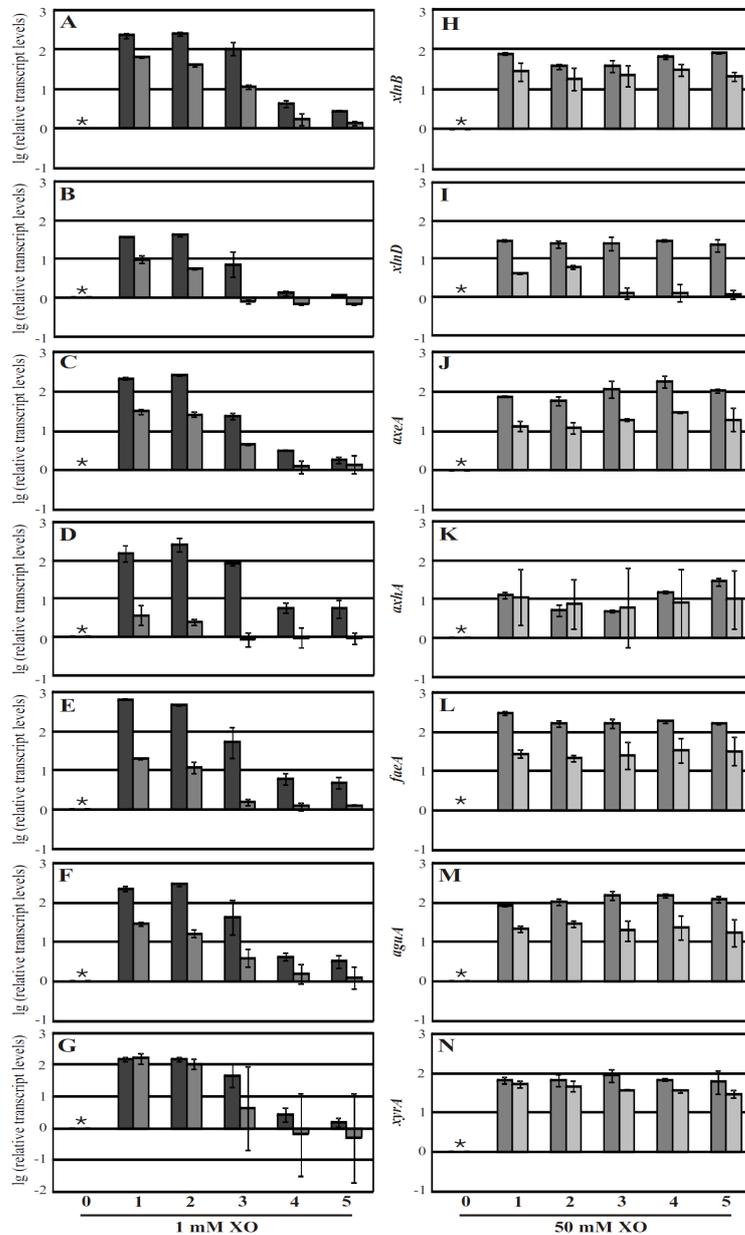


Figure 4.5: Influence of CreA on the transcription of genes involved in xylan-degradation and D-xylose metabolism in *A. niger*. The strain NW283 (left, dark gray bars) was induced using 1 or 50 mM D-xylose (XO) and samples were taken hourly for 5 h. Transcript level analysis of *xlnB* (A and H), *xlnD* (B and I), *axeA* (C and J), *axhA* (D and K), *faeA* (E and L), *aguA* (F and M), and *xyrA* (G and N) was performed by qPCR using a histone-like and a *Schizosaccharomyces pombe dma1*-like gene transcript for normalization. Transcript levels always refer to the reference sample (taken directly after induction) within an experiment, which is indicated by an asterisk. The results are presented as the relative transcript levels in logarithmic scale (*lg*). The values are means of two independent biological experiments measured in triplicate. Error bars indicate the standard deviation. The transcript levels of the wild-type strain (right, light gray bars) are pictured again for comparison.

Chapter 4. D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*

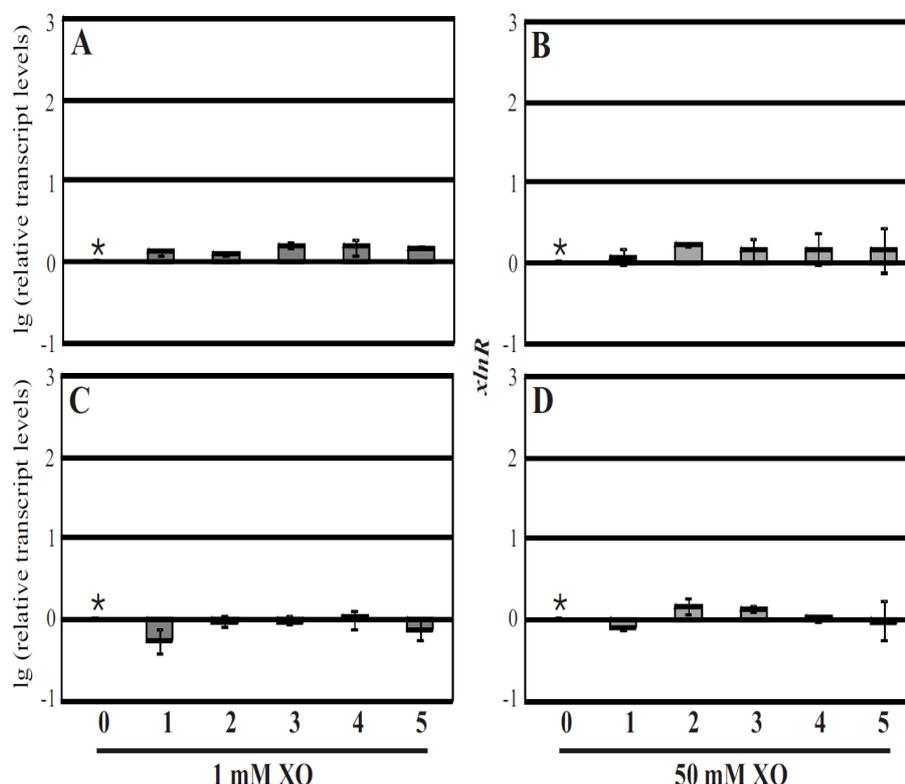


Figure 4.6: Influence of CreA on the transcription of *xlnR* in *A. niger*. The strains N400 (A and B) and NW283 (C and D) were induced using 1 or 50 mM D-xylose (XO) and samples were taken hourly for 5 h. Transcript level analysis of *xlnR* was performed by qPCR using a histone-like and a *Schizosaccharomyces pombe dma1*-like gene transcript for normalization. Transcript levels always refer to the reference sample (taken directly after induction) within an experiment, which is indicated by an asterisk. The results are presented as the relative transcript levels in logarithmic scale (*lg*). The values are means of two independent biological experiments measured in triplicate. Error bars indicate the standard deviation.

4.3.2 Expression of accessory enzyme-encoding genes is mostly favored using a high D-xylose concentration

We used the samples from *A. niger* N400 mycelia generated as described in the methods to analyse expression of different genes encoding accessory enzymes, namely *axeA*, *axhA*, *faeA*, *estA*, *abfB*, *aguA*, *aglB*, and *lacA*. Overall, we found that the change in transcript level of *axeA* after one and 2 h after induction as well as the levels of *estA* and *aguA* after 1 h are higher in the samples induced with 1 mM D-xylose (Figure 4.2 A, D, and F) than in those induced with 50 mM D-xylose (Figure 4.2 I, L, and N). For all the other genes, in particular, during the later sampling time points (3, 4, and 5 h after the addition of D-xylose), we detected higher transcript levels in the 50 mM D-xylose culture (compare Figure 4.2 A to H and Figure 4.2 I to P). This finding again correlates with the depletion of D-xylose in the culture after 2 h if the initial concentration was

1 mM (compare Table 4.3). An exception to this trend was *estA* and *aglB* because they maintained constant transcript levels even after 3 h (Figure 4.2 D and G). However, *aglB* is not generally significantly induced. Overall, in particular, *axeA* and *aguA* respond to different D-xylose concentrations in the same way: the greatest change in transcript levels occurs with 1 mM D-xylose after 1 or 2 h, but a constant, highly induced level over time was obtained with 50 mM D-xylose. However, all genes show a prolonged induction in the presence of the high D-xylose concentration.

4.3.3 Potential of D-xylose to induce expression of cellulase-encoding genes

Because it was reported that the xylanolytic and cellulolytic enzyme systems are coordinately expressed in *A. niger* (reviewed by [106] and [18]), we investigated the response of genes encoding cellulose-degrading enzymes to low and high D-xylose concentrations. Using 1 mM D-xylose triggered the induction of *eglA* and *bglA* (Figure 4.3 C and F) and low induction of *cbhA* and *eglB* (Figure 4.3 A and D), and the expression of *cbhB* and *eglC* was similar to that in the pre-induction sample (Figure 4.3 B and E). While *cbhB* remained at a very low transcript level after adding 50 mM D-xylose (Figure 4.3 H), *eglC* transcript formation increased significantly (Figure 4.3 K). Compared to the low D-xylose concentration, the higher D-xylose concentration resulted in increased transcript levels of *cbhA* (Figure 4.3 G) and *eglA* (Figure 4.3 I), while *eglB* (Figure 4.3 J) and *bglA* (Figure 4.3 L) exhibited reduced expression. Altogether, we found that D-xylose is not a very good inducer of cellulases, except for *eglA*, *bglA*, and at higher concentrations, *eglC*. The use of higher D-xylose concentration appears to be primarily beneficial for *eglC* transcript expression.

4.3.4 High D-xylose concentrations are preferable for inducing the gene expression of enzymes in the pentose metabolic pathway

According to microarray analysis data reported for *A. niger*, the expression of some genes coding for enzymes involved in the pentose metabolic pathway are up-regulated on D-xylose compared to other carbon sources [35, 103]. In this study, we investigated the influence of the concentration of D-xylose, which is one of the two starting metabolites of the pentose metabolic pathway, on regulation. We found that transcript formation of *xyrA*, *xdhA*, *ladA*, *xkiA*, and *talB* can be induced using 1 mM D-xylose for the first 2 h after induction and that after 2 h, expression strongly decreases (Figure 4.4 A to E). If 50 mM D-xylose is used for induction, the transcript formation of *xdhA*, *ladA*, and *xkiA* is even higher after 1 h than with 1 mM D-xylose. The observed transcript levels stay at this high or a slightly reduced level for the entire time period investigated (Figure 4.4 G to I). *xyrA*, which encodes for the enzyme that directly metabolizes D-xylose (to xylitol), is an exception; after 1 h of induction, the change in transcript levels is significantly higher with 1 mM D-xylose than with 50 mM D-xylose (Figure 4.4 F). *TalB* showed similar transcript levels after the first

Chapter 4. D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*

hour independent of the D-xylose concentration, but after 2 h, 50 mM D-xylose yielded enhanced transcript levels (Figure 4.4 J). Overall, the induction of the genes involved in pentose metabolism is stronger using a high D-xylose concentration, in particular, if observed for longer than 1 h.

4.3.5 Absence of a functional CreA positively influences expression of genes encoding xylan-degrading enzymes independent of D-xylose concentration

Previously, CreA-dependent decreased transcript levels of four xylanolytic enzyme - encoding genes on high D-xylose concentrations was observed by growth of *A. niger* in different D-xylose concentrations [19]. To determine if and to what extent CreA influences the expression of genes encoding certain hemicellulases (i.e., *xlnB*, *xlnD*, *axeA*, *axhA*, *faeA*, *aguA*) and *xyrA* on D-xylose, we grew the CreA-mutant strain NW283 under the same conditions as the wild-type strain and analyzed gene expression using RT-qPCR. In the CreA mutant strain, the change in transcript levels of the observed genes generally increased (up to two orders of magnitude) relative to the wild-type strain, regardless of whether 1 mM or 50 mM D-xylose was used for induction (Figure 4.5). The only exception was *xyrA*, which showed no significant difference in changes in transcript levels when comparing the two strains (Figure 4.5 G and N). Interestingly, when comparing the transcript levels in the CreA mutant strain between samples derived from 1 mM D-xylose and the ones derived from 50 mM D-xylose, we found higher transcript levels after 1 and 2 h of induction if 1 mM D-xylose was used for induction (Figure 4.5 A to G) for all the genes investigated. Between 2 and 3 h, the D-xylose was depleted in the 1 mM D-xylose initial concentration (compare Table 4.4), which correlates with the distinct decrease in transcript levels after 3 h. In the presence of 50 mM D-xylose, slightly lower transcript levels were obtained but remained stable over time (Figure 4.5 H to N). In summary, a CreA deficiency leads to a strong enhancement of transcript formation that is independent of the concentration of D-xylose. This finding leads to the conclusion that CreA is involved in mediating the D-xylose repression signal via a concentration-independent mechanism.

4.3.6 The *xlnR* gene is constitutively expressed independent of CreA and D-xylose concentration

Because the repressor CreA is involved in regulation of hydrolase-encoding gene expression on D-xylose, we asked if CreA might exert its effect indirectly by regulating the expression of the activator XlnR. Furthermore, it was unclear if the difference in the response to different D-xylose concentrations of certain genes, in particular, the ones encoding xylan backbone-degrading enzymes and some accessory enzymes, is provoked by a D-xylose concentration-dependent expression of their transcriptional activator. A comparison of *xlnR* transcript levels using 1 mM or 50 mM D-xylose shows similar constitutive transcript levels over the whole period investigated (Figure 4.6 A and B). Analysis of *xlnR*

4.3. Results

Target D-xylose concn (mM)	Time (h) after induction	Concn (mM) ^a	
		D-xylose	Sorbitol
1	0	0.7±0.2	85.4±2.3
	1	0.6±0.1	77.2±1.0
	2	ND ^b	70.9±1.0
	3	ND	65.7±2.7
	4	ND	56.1±3.5
	5	ND	48.0±3.3
50	0	55.1±0.8	91.2±3.7
	1	52.8±0.9	85.2±2.5
	2	52.1±0.5	83.4±4.6
	3	47.9±1.7	78.7±8.3
	4	45.5±0.2	75.3±5.8
	5	43.1±0.1	70.1±5.3

Table 4.3: HPLC analysis of D-xylose and sorbitol concentrations in supernatants of *A. niger* N400 cultivations in benchtop fermenters. ^aValues are means and standard deviations from two biological experiments. ^b“ND”, not detected; the limit of quantification was < 0.03 mM.

Target D-xylose concn (mM)	Time (h) after induction	Concn (mM) ^a	
		D-xylose	Sorbitol
1	0	0.9±0.1	90.3±5.8
	1	0.9±0.2	94.3±1.1
	2	0.5±0.1	94.8±2.3
	3	ND ^b	90.8±1.0
	4	ND	91.3±2.7
	5	ND	88.5±3.1
50	0	48.7±4.3	96.1±3.4
	1	47.2±2.8	99.5±5.2
	2	45.6±0.2	96.7±0.6
	3	46.5±1.9	96.0±4.8
	4	45.2±4.0	97.5±6.1
	5	44.2±0.1	95.6±1.0

Table 4.4: HPLC analysis of D-xylose and sorbitol concentrations in supernatants of *A. niger* NW283 cultivations in benchtop fermenters. ^aValues are means and standard deviations from two biological experiments. ^b“ND”, not detected; the limit of quantification was < 0.03 mM.

transcript levels in the CreA mutant also yielded no significant differences in expression, regardless of whether *xlnR* expression on different D-xylose concentrations within the CreA mutant strain was compared (Figure 4.6 C and D)

or whether *xlnR* expression between both strains was compared (Figure 4.6). Briefly, *xlnR* is stable and constitutively expressed, regardless of whether CreA is present or the amount of D-xylose that is applied.

4.4 Discussion

A. niger is an important organism for the production of native enzymes such as hemicellulases. Because D-xylose is often used to induce the expression of hemicellulases, we studied the D-xylose induction kinetics of the system and D-xylose-mediated carbon catabolite repression. In this study, we demonstrated that the use of D-xylose provokes a concentration-dependent response that is mostly specific for genes encoding enzymes with a particular hydrolytic function. Because D-xylose is depleted after 2 h if a concentration of 1 mM was applied for induction, it is most useful to focus on the transcript levels detected in samples taken 1 h after induction for comparison of the inducing potential of high and low D-xylose concentrations. Nevertheless, we think it was also necessary to investigate transcript levels during D-xylose depletion to determine if the constant presence of D-xylose (high concentrations) might be beneficial. At all sampling time points, the cultures were in the log phase with respect to their oxygen consumption.

Although in some cases, the use of 1 mM D-xylose provoked the greatest change in transcript levels, it is important to note that with 50 mM D-xylose, transcript levels could be detected that were clearly higher than those from the pre-induced samples. With respect to *xlnB* and *xlnD*, this result is consistent with the study published by de Vries et al. in 1999. In the case of *aguA* and *faeA* on 50 mM D-xylose, hardly any transcript was detected in that study [19], but otherwise, in the same year, de Vries et al. observed clear transcript formation of the same genes regardless of whether high (66 mM) or low (2 mM) D-xylose concentrations were applied [105]. This discrepancy might be due to the use of Northern Blot analysis, which is difficult to standardize. However, the later results are more consistent with the results of this study. Nevertheless, it should also be taken into account that in the present study, all data were obtained similarly from growth conditions, while in the previous studies, growth conditions (high D-xylose concentrations) were compared to cell-resting conditions (low D-xylose concentrations). The experimental design of this study allows a quantitative comparison of 1 and 50 mM D-xylose concentrations, even if rather small differences in transcript levels are observed.

In general, the utilization of a high D-xylose concentration was found to be beneficial, in particular, for the induction of hemicellulase-encoding genes, because their transcript formation is maintained for a longer time. However, *xlnD* clearly poses an exception because its transcript level decreases after 2 h even in the high D-xylose background. Interestingly, in *A. nidulans*, β -xylosidase activity induced by 66 mM D-xylose was detected after 2 h, but decreased afterwards [130]. In the same study, a strong *xlnD* transcript was detected after 1 h on D-xylose, which decreased after 2 h [130]. However, we observed that the use of a

CreA mutant strain completely restored the prolonged transcript formation on 50 mM D-xylose.

The release from CreA influence elevated the induction of the six investigated hemicellulase-encoding genes compared to the wild-type strain by 0.5 to 2 orders of magnitude. Notably, this increase was observed independent of whether 1 mM or 50 mM D-xylose was used. Initially, this finding might lead to the conclusion that a relief from carbon catabolite repression provokes a general increase in transcript formation in a sorbitol background. However, a recent study of *A. niger* demonstrated that sorbitol is a non-inducing carbon source compared to D-xylose with respect to the expression of hemicellulases and has no inducing effect on the expression of carbohydrate-active enzymes [131]. Furthermore, evidence was provided that sorbitol does not repress xylanase expression in *A. tubigensis* and *A. niger*, while glucose and fructose do repress expression [116]. The previously mentioned study of de Vries et al. also reported a D-xylose concentration-independent (ranging from 1 - 100 mM) increase in transcript formation in a CreA mutant, although they transferred the strain to medium containing D-xylose as the only carbon source after pre-growth on fructose [19]. Likewise, in *Hypocrea jecorina* (*Trichoderma reesei*), the transfer of a pre-grown *cre1* (the *creA* orthologue) deletion strain to D-xylose as the sole carbon source led to increased transcript levels of xylanase-encoding genes compared to the parental strain, independent of the D-xylose-concentration [132]. Taken together, these findings strongly support the assumption that the effects observed in the CreA deficient strain during this study result from the use of D-xylose as an inducer. Finally, it should be mentioned that *xyrA* was the only gene investigated whose transcript formation did not significantly respond to release from the influence of CreA. Even if there is a single CreA-binding site (CTGGGG) present at a position 757 bp upstream from the *xyrA* structural gene, this does not necessarily mean that it is functional *in vivo* under the applied conditions.

Even if the XlnR regulon is inducible by D-xylose, we demonstrated that the transcript levels of *xlnR* do not reflect this induction. This result is consistent with what has been reported for the expression of the XlnR-orthologue, Xyr1, from *H. jecorina* [133]. In those studies it was found that inducer molecules such as D-xylose, L-arabitol or xylobiose cannot induce *xyr1* transcription although they are used to induce xylanase expression [134, 135]. Because the induction signal is not mediated/communicated by transcriptional regulation of the *xlnR* gene, XlnR *de novo* synthesis-based regulation or an XlnR-activating post-translational modification would be a possible alternative regulatory mechanism. Recently, Noguchi et al. provided the first support for such a hypothesis when they reported that D-xylose triggered reversible phosphorylation of XlnR in *A. oryzae* [136].

Previously, it was reported for *A. nidulans* that CreA represses the expression of *xlnA* and *xlnB* on glucose by an indirect mechanism [137, 138], which was more recently identified as a CreA-mediated repression of the *xlnR* gene [139]. Moreover, in *H. jecorina*, repression of *xyr1* transcription by Cre1 was re-

Chapter 4. D-xylose concentration-dependent hydrolase expression profiles and the according role of CreA and XlnR in *A. niger*

ported on D-glucose as well as on D-xylose [132, 135]. However, in this study, we could detect no significant differences in *xlnR* transcript levels in an *A. niger* CreA mutant strain compared to the wild-type strain, regardless of whether low or high D-xylose concentrations have been used. This result suggests that other regulatory mechanisms must be acting in the functional regulation of XlnR activity and that they might be present at the post-transcriptional level.

Acknowledgements

This work is supported by the graduate school VLAG and the IPOP program of Wageningen University.

**Quantification of the effect of the
carbon catabolite repressor CreA
on transcription in the XlnR
regulon of *Aspergillus niger***

Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*

Abstract.

In this work measurement data from time course experiments was used to study the effect of the carbon catabolite repressor (CreA) on the transcription of the genes from the XlnR regulon of *A. niger*. The transcription of genes encoding xylanolytic and cellulolytic enzymes in the wild type strain (Wt) and CreA mutant strain (Mt) was studied at 1 mM and 50 mM D-xylose inducer concentrations. Transcription and repression levels depend on the inducing D-xylose concentration. The time course data showed that repression by CreA varies over time and differs amongst the genes. To quantify the percentage change in transcription over time, the CreA effect ratio (CER) was defined; i.e. the degree of target gene repression by CreA. The CER was used to compare transcription across strains, conditions and time. In the CreA mutant strain different effects were found amongst the target genes compared to the Wt. Genes with a single CreA binding site exhibited complete de-repression of transcription while genes with multiple CreA binding sites exhibited partial de-repression in time. The regulatory mechanisms of the regulon involving D-xylose induction through CreA were examined. Some indications were found for additional transcription factors other than the XlnR and CreA proteins that contribute in regulating the target genes. The study also shows that judging repression effects on the basis of a single time point observation can be misleading.

Keywords: CreA effect ratio, XlnR regulon, *A. niger*, D-xylose, time course data.

This chapter is based on:

J. Omony, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten and A.J.B. van Boxtel. Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*. 2012. *Submitted for publication*.

5.1 Background

Transcription of xylanolytic and cellulolytic enzymes in *A. niger* is controlled by the xylanolytic activator protein (XlnR). The target genes of the XlnR regulon encode the main xylanolytic enzymes xylanases B and C and β -xylosidase, and accessory enzymes such as α -glucuronidase A, acetylxylan esterase A, arabinoxylan arabinofuranohydrolase A, and feruloyl esterase A. In addition to these xylanolytic enzymes, the target genes also encode cellulolytic enzymes like endo-glucanases A, B, C and cellobiohydrolases A and B.

The carbon catabolite repressor protein (CreA) also mediates the transcription of the target genes in the XlnR regulon in *A. niger* [19, 85]. The effect of CreA was described by Arst et al. [140] in a related fungus, *A. nidulans*, where it represses the transcription of many target genes in response to the presence of sugars like e.g. D-glucose, D-fructose and D-xylose. Over the years, further studies on the repressive role of CreA have been done (e.g. [113, 141]). Mach-Aigner et al. [85] studied the effect of CreA on the transcription of the genes in the XlnR regulon in *A. niger* by comparing the response in two strains: *A. niger* N400 (Wt) and *A. niger* N283 (CreA), which is a partial loss-of-function mutant (Gielkens et al. [107] and van Peij et al. [34]). In the work of Mach-Aigner et al. [85] the sensitivity in response to D-xylose was measured in time course experiments and the response of the target genes was compared qualitatively.

Measuring the transcription profiles in time, as done in the work of Mach-Aigner et al. [85], enables a detailed comparison of the transcription dynamics. These dynamics may be missed out if comparisons of transcription are based on single rather than multiple time points. Since the administration of D-xylose activates the CreA protein, it is important to see how the effect of CreA repression varies in time. In previous studies this effect has been qualitatively described for one or a few points in time (e.g. [19, 85, 139]). Much as it is important to have a qualitative description of the effect of CreA on transcription of the target genes, it is also important to quantify the effect as well. Therefore, the goals of this work are to (i) quantitatively describe the role of CreA in the transcription of the target genes in the XlnR regulon, (ii) investigate whether there is extra information to be gained by using time course data sets as opposed to using single time points for comparative studies for transcription. In this work data from the time course experiments from Mach-Aigner et al. [85] is used.

5.2 Materials and methods

5.2.1 Description of experiments

Experiments 1 and 2: These time course experiments concerned the *A. niger* N400 (Wt) with 1 mM D-xylose as an inducer, and sampling on an hourly basis for a total of 6 data points. The setup consisted of two fermentors that were used in parallel. Three measurements were taken from each duplicate and averaged to yield the data values used for the analysis. Experiment 2, which had a similar

Chapter 5. Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*

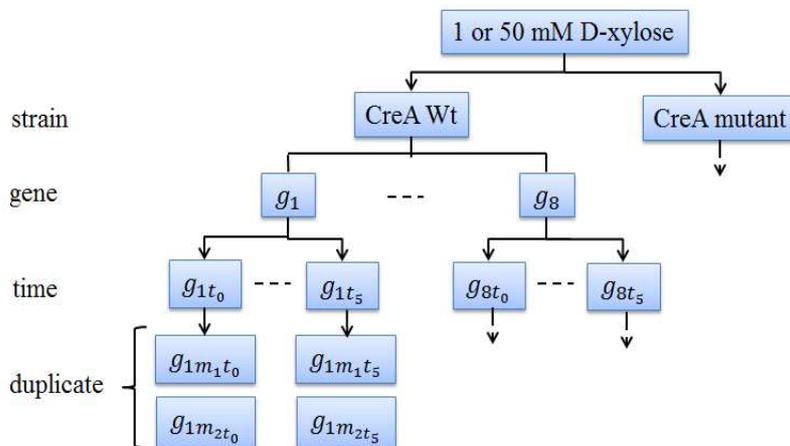


Figure 5.1: Schematic representation of the experimental design. Both the 1 mM and 50 mM D-xylose concentrations were used to induce the system. The terms m_1 and m_2 represent the measurement duplicate levels, t_0 and t_5 are the time indices, and g_1, \dots, g_8 represent the eight genes used in the study. A branching design similar to that used for the CreA Wt was also used for the CreA Mt strain.

experimental set-up, involved inducing the system with 50 mM D-xylose. An overview of the scheme used for the experiments is given in Figure 5.1. The transcription measurements were performed using real-time qPCR analysis.

Experiments 3 and 4: The *A. niger* NW283 (CreA) strain was used in these time course experiments. A D-xylose concentration of 1 mM was used to induce transcription in experiment 3. Experiment 4 was performed with a 50 mM D-xylose inducer. The same number of data samples was collected for all the genes. A total of 6 data points in 5 hours was used (Figure 5.1). As in experiments 1 and 2, experiments 3 and 4 also had two duplicates for each time point.

5.2.2 Data analysis criteria

The hypothesis that was tested is that transcription of a gene is not CreA repressed. If this hypothesis holds true then we expect to observe no significant difference in transcription in the Wt and the CreA strains. Let $y_{i,Mt}(j)$ and $y_{i,Wt}(j)$ represent the average observed transcription of gene i over all samples and duplicates at time point t_j in the CreA strain and in the Wt strain, respectively. The notations Mt and Wt indicate the mutant and wild type, respectively.

To normalize the transcription values, a histone-encoding gene (*hisT*) was used as a reference gene [35]. The calculations were performed using the equation found in [129]. The computed transcription ratio was obtained from the expression:

$$y_{i,\bullet} = E_{\text{target}}^{\Delta C_t(\text{pre-post})_{\text{target}}} / E_{\text{ref}}^{\Delta C_t(\text{pre-post})_{\text{ref}}} \quad (5.1)$$

where the subscript " \bullet " in $y_{i,\bullet}$ is an indicator of the strain used, "target" and

“ref” represent the target and reference gene samples, respectively. The term E denotes the amplification efficiency, and $C_t(\text{pre} - \text{post})$ denotes the difference in threshold cycles between pre- and post-induced samples. The expression Δ denotes a deviation in a variable and the term C_t refers to the cycle threshold value at constant fluorescence level.

The mean transcription of gene i over all time points in the CreA strain and Wt is represented by $\bar{y}_{i,\text{Mt}}$ and $\bar{y}_{i,\text{Wt}}$, respectively. The null hypothesis to be tested is that there is no significant difference in the average transcription of gene i in the Mt and Wt strains, i.e. $H_0 : \bar{y}_{i,\text{Mt}} = \bar{y}_{i,\text{Wt}}$ and the alternative hypothesis is $H_1 : \bar{y}_{i,\text{Mt}} > \bar{y}_{i,\text{Wt}}$. The two sample t -test was used to test the hypothesis. The number of data points (N) was the same for all the genes. The test statistic for the two sample t -test is given by

$$t = \left[\frac{(\bar{y}_{i,\text{Mt}} - \bar{y}_{i,\text{Wt}}) / s_{i,\text{Mt,Wt}} \sqrt{2/N}}{\sqrt{2/N}} \right] \sim t_{2(N-1),\alpha} \quad (5.2)$$

This test is based on the assumption that the data from the two experimental conditions has the same variance structure. It tests the evidence against the null hypothesis that a gene is not differentially expressed across conditions (e.g. [142, 143, 144]). The pooled variance term in (5.2) is given by

$$s_{i,\text{Mt,Wt}}^2 = \frac{1}{N} (s_{i,\text{Mt}}^2 + s_{i,\text{Wt}}^2) \quad (5.3)$$

The square root of this term represents the “pooled” standard deviation $s_{i,\text{Mt,Wt}}$. The terms $s_{i,\text{Mt}}^2$ and $s_{i,\text{Wt}}^2$ are the estimated variances on the transcription from the mutant and wild type samples, respectively.

To ensure a fair comparison the same reference (histone-encoding gene, *hisT*) was used to normalize the data sets, which is similar to that used in [35]. The sampled data points corresponding to the same time instants (0, 1, 2, . . . , 5 hours) were used for the data analysis. For the CreA mutant we expect a significant increase in transcription in both the 1 mM and 50 mM D-xylose experiments. Sample processing variations may arise from experimental duplicates, fermenters, RNA sample preparation, and time effect in handling the samples [145]. This potentially narrows down the power of the statistical tests required to detect significant differences between variables of interest. Hence, a p -value of 0.10 was used as a threshold significance level.

5.2.3 D-xylose uptake in time

The medium for growth contained 100 mM as a carbon source. After pulse inducing using 1 mM or 50 mM, the D-xylose concentration begins to reduce in time. Upon induction, the D-xylose consumption rates were observed to be different. Almost 80% of the initial concentration was left after the fifth hour for the 50 mM D-xylose induced system compared to near zero percent of the 1 mM D-xylose (see Table 5.1). This high inducer concentration leaves an excess of D-xylose that drives the transcription process [146].

Chapter 5. Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*

Time (h)		0	1	2	3	4	5
D-xylose concentration	1 mM	1.00	0.84	0.4225	0.1549	0.0423	0.01
	50 mM	50.00	47.97	47.09	41.98	40.97	38.79

Table 5.1: D-xylose concentration as a function of time. The concentrations were measured following a pulse at time zero (0 hours).

5.2.4 The CreA effect ratio (CER) statistic

To quantify the percentage change in transcription over time, we define a term that specifies the degree of target gene repression by CreA. This term is the CreA effect ratio (CER). To quantify the average ratio of the effect of CreA in the Wt and in the CreA mutant strain on transcription of the target genes in the XlnR regulon, we consider a statistic that measures the average percentage fold change in time. The statistic is defined as

$$\bar{\mu}_{i,\text{CreA}} = \frac{1}{N} \sum_{j=1}^N \mu_{i,\text{CreA}}(j) \quad (5.4)$$

where

$$\mu_{i,\text{CreA}}(j) = y_{i,\text{Wt}}(j)/y_{i,\text{Mt}}(j) \quad (5.5)$$

The statistic (5.5) quantifies the effect of CreA on transcription at time instant j . The CER metric ($\mu_{i,\text{CreA}}$) lies in the domain $0 \leq \mu_{i,\text{CreA}}(j) < +\infty$. If no significant differential transcription of a gene expressed in the Wt and CreA strain is observed, then the ratio $\mu_{i,\text{CreA}}(j)$ is expected to be equal to 1 ($\mu_{i,\text{CreA}}(j) = 1$). If the CER lies in the interval $0 \leq \mu_{i,\text{CreA}}(j) \ll 1$, then a gene is considered to be significantly less expressed in the Wt strain. Thirdly, when $\mu_{i,\text{CreA}}(j) \gg 1$, then a gene is significantly expressed in the Wt strain compared to the CreA strain. The above inference for the ratio $\mu_{i,\text{CreA}}(j)$ also implicitly applies to $\bar{\mu}_{i,\text{CreA}}$ in (5.4).

5.3 Results

5.3.1 Comparing transcription dynamics

The transcription data was analyzed to test for differential transcription across the Wt and CreA strains. In each of experiments 3 and 4, a total of eight genes were profiled, including the transcriptional regulator *xlnR* gene. In the work of Mach-Aigner et al. [85], the log-ratio data was used, while in this work the transcription ratio data is used. The use of the transcription ratio data is convenient for computing the CreA effect ratio. From Figures 5.2 and 5.3, we see that the regions of peak transcription vary widely. At low D-xylose concentrations the genes attain their highest level of transcription between the first and third hour. This provides a distinct characteristic transcription profile for the target

5.3. Results

genes in time. In contrast, for the 50 mM D-xylose inducer the regions of peak transcription vary between genes, especially in the CreA Mt strain.

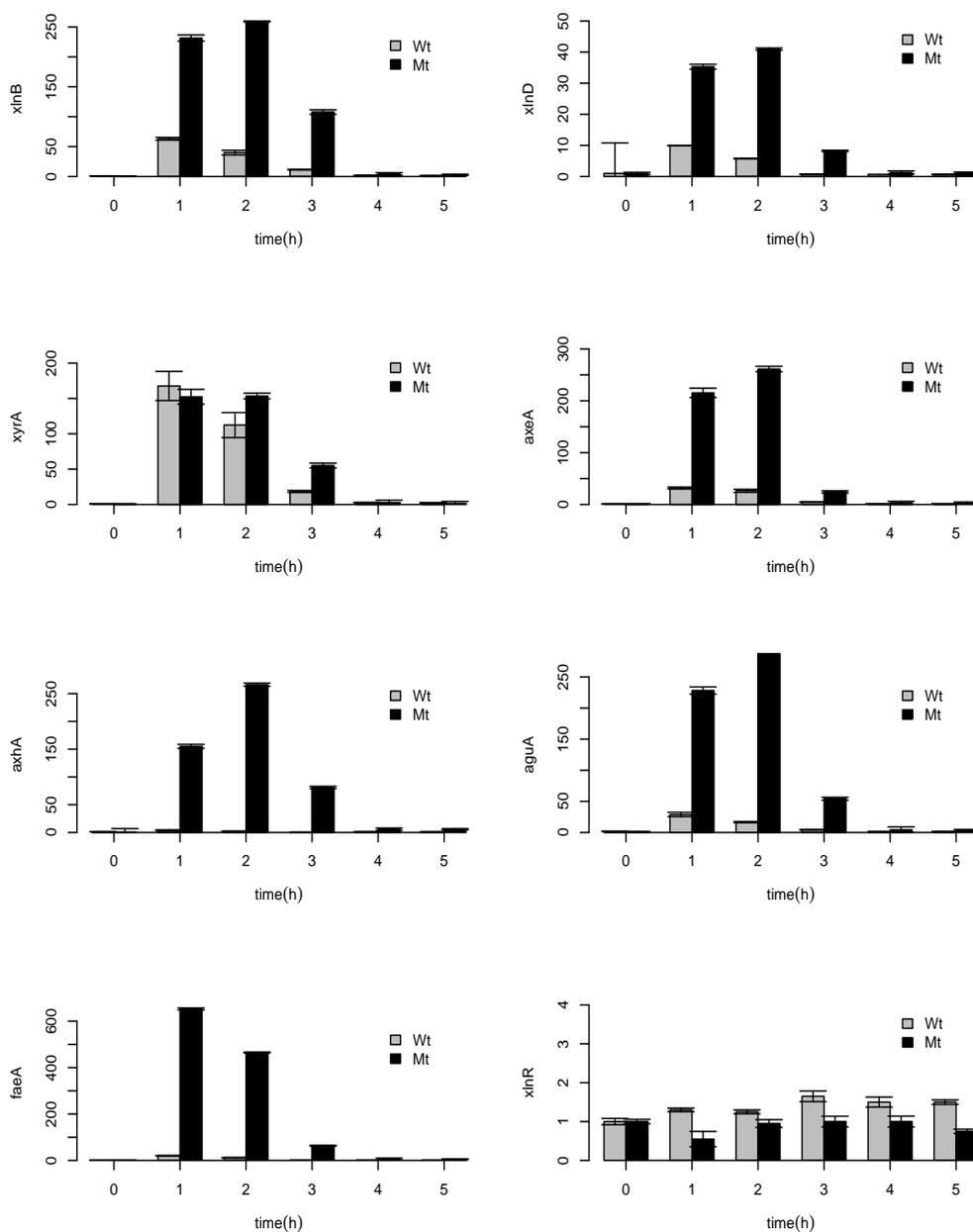


Figure 5.2: The 1 mM D-xylose inducer experiment. A comparison of the transcription profiles from the CreA Wt and Mt experiments. Nearly all the genes (except for D-xylose reductase *xyrA* and the transcriptional activator *xlnR*) showed wide differences between the experiments with the CreA Mt and Wt strains. The highest activity levels are observed between the first and third hour. Transcription of the *xlnR* gene remains steadily unchanged over time. Note: The vertical axes are scaled differently to aid proper visibility.

Chapter 5. Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*

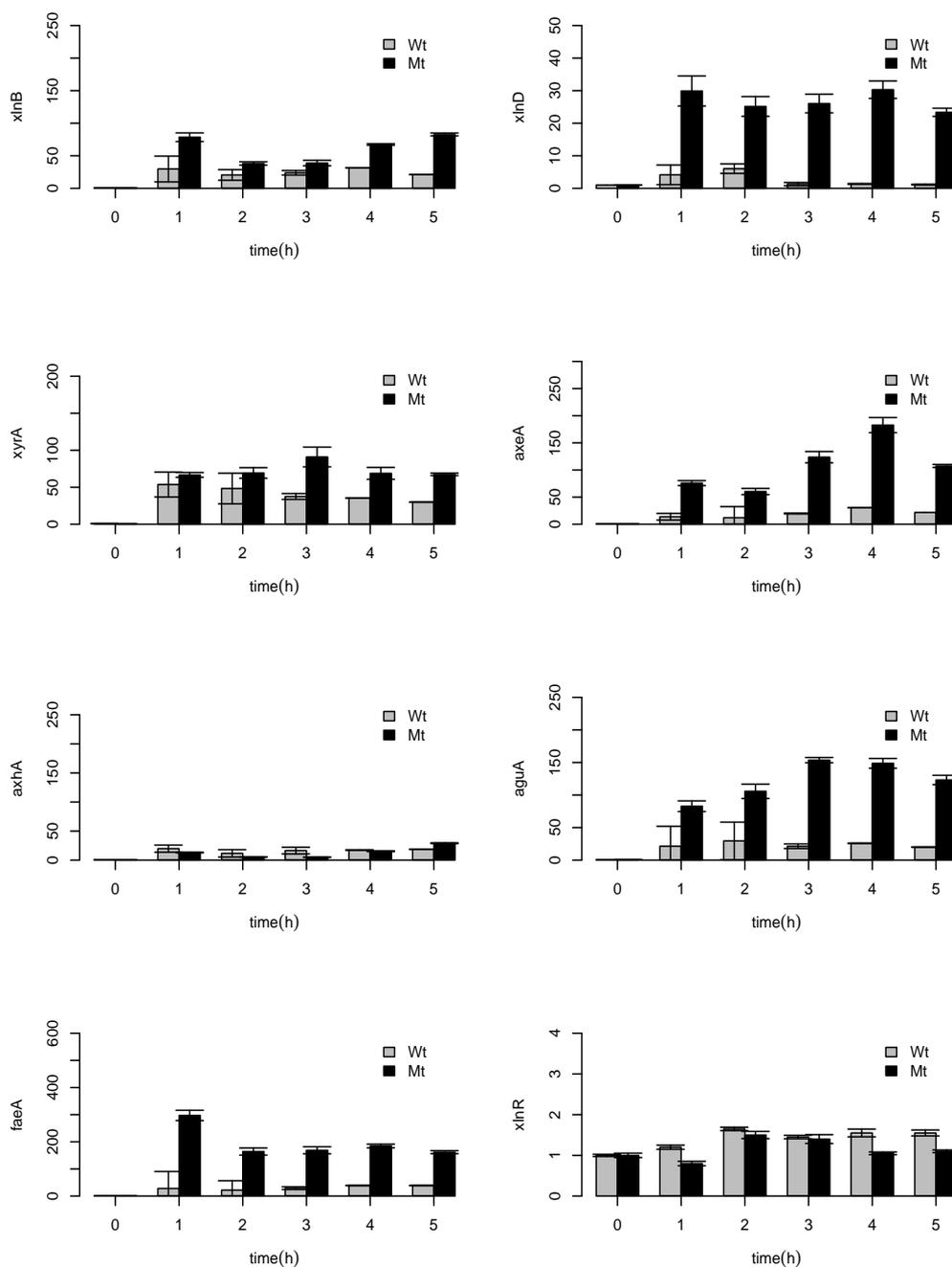


Figure 5.3: The 50 mM D-xylose inducer experiment. A comparison of the transcription profiles from the CreA Wt and Mt experiments. Higher D-xylose concentration reduces the transcription level of the arabinoxylan hydrolase *axhA* gene under the CreA Mt strain compared to the CreA Wt strain. Transcription of the *xlnR* gene remains steadily unchanged over time. Note: The vertical axes are scaled differently to aid proper visibility.

By considering each gene separately, the differences in the transcription profiles (Figures 5.2 and 5.3) are mainly attributed to the CreA mutation and the inducing D-xylose concentration. By using 1 mM D-xylose concentration, nearly all the available D-xylose is depleted by the fourth hour (see Table 5.1). This depletion in D-xylose leads to loss in transcription at about five hours of induction (Figure 5.2). In contrast, for the 50 mM D-xylose, after 5 hours of induction a large fraction of the initial D-xylose concentration (about 80%) remained (Table 5.1).

Gene	Enzyme function	1 mM D-xylose		50 mM D-xylose	
		$t_{10,0.10}$	$p = 0.10$	$t_{10,0.10}$	$p = 0.10$
<i>xlnB</i>	endo-xylanase B	1.63	0.067	2.21	0.026
<i>xlnD</i>	β -xylosidase	1.50	0.083	4.43	0.001
<i>xyrA</i>	D-xylose reductase	0.24	0.406	1.81	0.051
<i>axeA</i>	acetyl-xylanesterase	1.48	0.084	2.95	0.007
<i>axhA</i>	α -L-arabinoxylan hydrolase	1.92	0.042	-0.53	0.697
<i>aguA</i>	α -Glucuronidase	1.67	0.063	3.54	0.003
<i>faeA</i>	ferulic acid esterase A	1.65	0.065	3.50	0.003
<i>xlnR</i>	transcriptional activator of hemi-cellulases and cellulases	-4.05	0.999	-1.75	0.945

Table 5.2: **Test for differential transcription between genes expressed under 1 mM and 50 mM D-xylose.** A table showing p -values for genes with differential transcription in the CreA Wt and Mt strains. The p -values for the genes that are differentially expressed are indicated in bold. The term $t_{10,0.10}$ represents the t -statistic value at $N - 2$ degrees of freedom and a threshold p -value of 0.10.

Nearly all the profiled target genes showed significant changes in transcription folds (Table 5.2). Transcription for all the target genes was significantly higher in the CreA Mt compared to the Wt (exception for the *xyrA* and *axhA* genes in the 1 and 50 mM D-xylose inducer conditions, respectively, Table 5.2). There was no significant difference in transcriptional level between the Wt and Mt strain for the *xlnR* gene. Previous studies by de Vries et al. [105], involving transcription of the *faeA*, *aguA*, *xlnB* and *xlnD* genes in a CreA Wt strain and in CreA Mt strain, showed that D-xylose concentrations higher than 1 mM lead to repression of the xylanolytic genes tested mediated by CreA. The experiments revealed the time dynamics, the regions in time at which the genes are most expressed, and the peak transcription values. This result is in agreement with the results from previous experiments in which the highest transcription levels of the target genes were observed after induction with the 1 mM D-xylose concentration [105].

5.3.2 CreA effect ratio (CER)

The CreA effects were computed for the individual time points for all genes using the expression in (5.5). In literature the role of CreA has mostly been described for one time point only; in our work we present the effects of the CreA protein on the transcription levels of the target genes over a time horizon.

Chapter 5. Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*

This shows how the level of a gene's transcription in the CreA Wt strain varies in comparison to its transcription in the CreA Mt strain. The low transcription level observed for the Wt can possibly be explained by the effect of the CreA protein, which represses transcription by blocking the promoter binding sites of the XlnR target genes, preventing the XlnR protein from binding. For each gene CER was calculated as a single measure of the effect of CreA over the five hours experimental time frame. Therefore, CERs were calculated using (5.5) and the results are shown in Table 5.3. Hence, less transcription as a result of the repression (Table 5.3). A similar regulation mechanism by the binding of transcription factors to target gene promoter binding site was shown in *A. nidulans* (e.g. [145, 147, 148]).

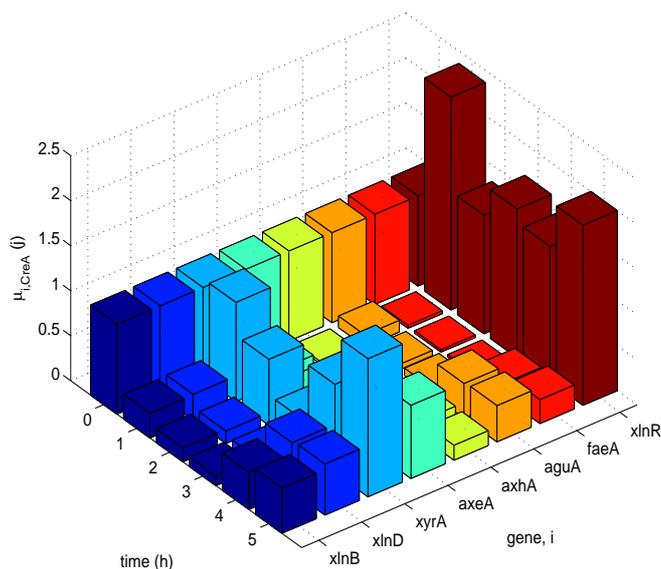


Figure 5.4: Bar plots for the CreA effect ratio (CER) in the 1 mM D-xylose inducer experiment. Plot for the CER effect in the 1 mM D-xylose inducer. Considering the entire space, the largest number of the ratios falls within the unit interval $0 < \mu_{i,CreA}(j) < 1$. Compared to the other genes, only the D-xylose reductase gene (*xyrA*) and the xylanolytic activator gene (*xlnR*) have a CER value exceeding the threshold value 1 at a few time points.

To further assess the CreA effect, consider the CER for the *axhA* gene shown in the Figures 5.4 and 5.5. At time zero, the CreA effect values are equal for both low and high D-xylose inducing conditions, this value quickly changes in time. At low D-xylose concentrations the ratio decreases within one hour and remains constant, while at high D-xylose concentrations the CreA ratio quickly increases to a peak of $\mu_{i,CreA}(j = 4) \approx 2.5$ before later-on reducing in time. This pattern shows how the transcription ratio varies across the two strains.

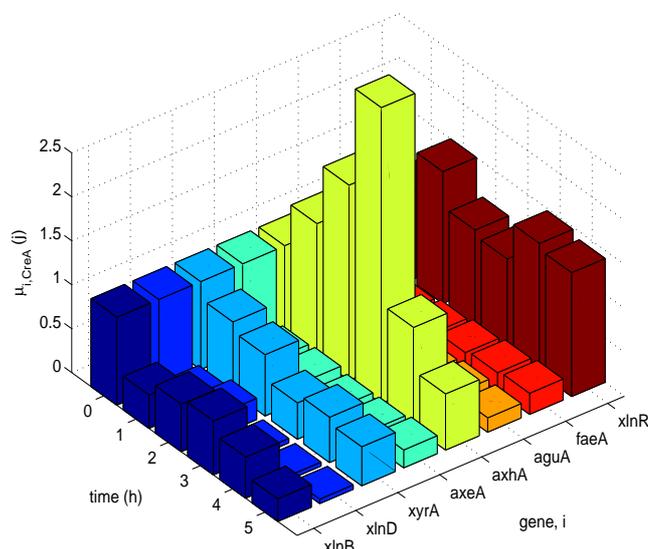


Figure 5.5: Bar plots for the CreA effect ratio (CER) in the 50 mM D-xylose inducer experiment. Plot for the CER effect in the 50 mM D-xylose inducer. Considering the entire space, the largest number of the ratios falls within the unit interval $0 < \mu_{i, \text{CreA}}(j) < 1$. Unlike the rest of the genes, the CER corresponding to the *axhA* gene far exceeds the value of 1 especially from the first to the third hour after induction.

5.4 Discussion and conclusion

D-xylose triggers transcription of target genes of the XlnR regulon. Previous studies involved single time point measurements on the role of CreA in controlling transcription (Tamayo et al. [139], Drysdale et al. [22], and Ruijter and Visser [125]). In contrast to these works we used the results from Mach-Aigner et al. [85], where the transcription is given over a time horizon, to analyze the transcription patterns over time. As a first result, the analysis shows by comparing the transcription in the Wt and Mt strains that over the whole time horizon most genes were significantly repressed by CreA (Table 5.2).

Assessment of the CreA effect at a single point, without the dimension time, would most likely lead to the interpretation that the CreA effect is constant. The analysis of the time course experiments shows the extent to which each gene in the Wt and CreA mutant strains is transcribed in time (Figure 5.2 and 5.3). From the time course experiments, we conclude that comparing transcription data (Wt and Mt) at a single time point is indeed limiting. By performing an experiment with measurements at one time point, information on gene behavior at other time points is missed. This is amongst others illustrated by the 1 mM experiments where the time averaged CER for *xlnB* is 0.41 (Table 5.3), while the ratio at 1 hour is 0.27 (Figure 5.4). For all the other genes the time

Chapter 5. Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*

averaged CERs differed by at least 0.10 from the corresponding ratios at 1 hour (CERs for *xlnD*, *xyrA*, *axeA*, *axhA* and *aguA* (1 mM D-xylose, Table 5.3) were 0.44, 0.95, 0.44, 0.23 and 0.34 with corresponding CERs of 0.28, 1.10, 0.15, 0.03 and 0.13 at 1 hour, respectively).

In Figure 5.2, the *xyrA* gene is more expressed at 1 hour in the Wt than in the Mt. For the same gene this expression pattern is reversed at 2 and 3 hours. It shows that there is a risk of falsely concluding that a gene is significantly more expressed in one condition than in the other. A similar situation occurs for *axhA* at the end of the 50 mM experiments. Figure 5.4 and 5.5 further confirm the limitation of measuring at one time point, where the CER values are given over time. In these graphs the CER values differ in time points and show the dynamic character of the system.

In the *alc* regulon of *A. nidulans*, the multiplicity of potential binding sites has been shown to be relevant to CreA mediated repression [145]. To link the number of CreA binding sites and transcription in the XlnR regulon of *A. niger*, we performed a sequence search. The search showed that the *xlnD* gene had up to 18 CreA binding sites, while the other genes had a significantly lower number of CreA binding sites (e.g. the *faeA* gene with a single CreA binding site, Table 5.3). The presence or absence of a transcription factor binding site also affects its transcription. For instance the *faeA* gene with one CreA binding site had the highest transcription values in the mutant (Figure 5.2), while the *xlnD* gene (with 18 CreA binding sites) had the lowest transcription levels amongst the target genes. These differences in transcription values can be seen around the 1 to 3 hour time window. A mutation in the CreA protein hinders its proper functionality as a repressing transcription factor. This implies that for genes such as *faeA* with a single CreA binding site, a mutation in the CreA protein leads to complete de-repression. For genes with multiple CreA binding sites, mutation of the CreA protein leads to partial de-repression of target gene activity. We found that the least transcribed genes had multiple promoter binding sites for the CreA protein (Table 5.3).

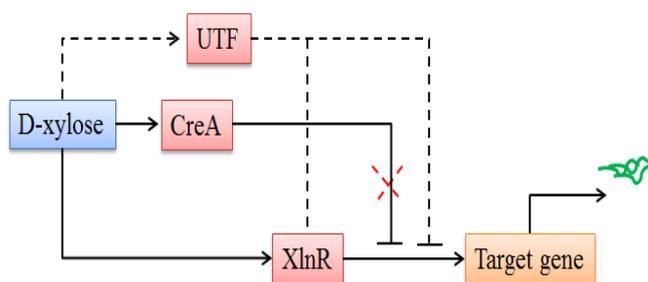


Figure 5.6: A scheme showing the effect of CreA and D-xylose. The transcription levels are dependent on the D-xylose concentration and the CreA strain used. Gene regulation in the XlnR regulon after D-xylose induction is XlnR and CreA mediated. The effect of CreA is lost in the Mt strain, here represented by the presence or absence of the dotted cross. The CreA partial loss-of-function is here shown with the cross. The unknown transcription factor (UTF) effect is indicated by the dotted line. This UTF potentially forms a complex with the XlnR protein, a complex that is ultimately involved in regulation of transcription.

Gene	Enzyme function	1 mM D-xylose $\bar{\mu}_{i,CreA}$	50 mM D-xylose $\bar{\mu}_{i,CreA}$	Number of CreA sites	Position ¹	Sequence [149]
<i>xlnB</i>	endo-xylanase B	0.4074	0.5424	4	226, 337, 499, 555	CTCCAC,CTCCAC, CTCCAC,CCCCAC
<i>xlnD</i>	β -xylosidase	0.4361	0.2538	18	18, 42, 55, 79, 87, 98, 124, 389, 499, 536, 674, 758, 800, 968, 992, 1093, 1098, 1237	CCCCAC,CCCCGG, CCCCAG,CTCCGC,CTCCGC,CTCCCG, CCCCGC,CTGGGG,GTGGAG,CTGGGG, CCGGGG,CCGGAG,GCGGAG,CCGGAG, GCGGAG,GCGGAG,GCGGAG,CTCCGC
<i>xyrA</i>	D-xylose reductase	0.9466	0.6453	2	164, 233	CTCCAG,CCCCAG
<i>axeA</i>	acetyl-xylanesterase ²	0.4441	0.3187	–	–	–
<i>axhA</i>	α -L-arabinoxylan hydrolase	0.2302	1.6451	3	234, 498, 758	GTGGGG,CCGGAG,CTGGGG
<i>aguA</i>	α -Glucuronidase	0.3415	0.3360	4	110, 128, 445, 507	GCGGGG,CCCCGC,CTCCAG,CCCCGG
<i>faeA</i>	ferulic acid esterase A	0.2572	0.3080	1	379	CTCCGG
<i>xlnR</i>	transcriptional activator of hemi-cellulases and cellulases	1.6382	1.2535	8	20, 240, 259, 481, 501, 567, 659, 916	CTCCAC,CTCCGG,CTCCGG, CCCCGG,CTGGGG,CCGGGG, GTGGAG,CCCCGC

Table 5.3: CreA effect ratios (CER), $\bar{\mu}_{i,CreA}$ and CreA binding sites. The CER values were computed using the statistic defined in (5.4). These values provide the average percentage change (increase or decrease) in transcription over time under the CreA Wt compared to the CreA mutant. For instance the transcription of the *xlnB* gene under the 1 mM D-xylose inducer condition decreases on average by about $|(1 - 0.4074)| \times 100 \approx 59\%$. At the same D-xylose concentration, transcription of the *xlnR* gene increased by about $|(1 - 1.6382)| \times 100 \approx 64\%$. This table also indicates the positions and sequences of the CreA binding sites on the various genes. ¹Positions are relative to the translation start point. ²No CreA binding sites were found for the *axeA* gene.

Chapter 5. Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*

Van Peij et al. [34] showed that D-xylose acts as a signal molecule of the regulon in two ways: 1) as an inducer of the XlnR pathway, and 2) by stimulating carbon catabolite repression at high D-xylose concentration via CreA. These two mechanisms are confirmed by the analysis in this work. Omony et al. [150] considered therefore that target gene repression was only due to CreA. However, since the transcription of the genes in the Mt for the 50 mM experiments is significantly lower than that of the 1 mM experiments (Figure 5.2 and 5.3), we conclude that there is also another mechanism triggered by D-xylose. This unknown transcription factor (UTF) mechanism represses the target gene transcription. Figures 5.2 and 5.3 show amongst others that the effect is very strong for the *axhA* gene with a 10 fold lower expression in the 50 mM experiment compared to the 1 mM experiment. The other genes are affected in a similar, but less strong, way. Hence, transcription of the genes in the XlnR regulon is also affected by an extra, but not yet qualified transcription factor (UTF), see Figure 5.6. Similar results were observed by Noguchi et al. [136] for XlnR-mediated induction in *A. oryzae*, and for the promoter of the gene encoding for a xylanase in *Hypocrea jecorina* [132].

Acknowledgements

This work is supported by the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences) and the "Investeringsplan / Ondernemingsplan" (IPOP) program of Wageningen University. The authors would like to thank Birgit Jovanovic of the Institute of Chemical Engineering at Vienna University of Technology for the assistance with the HPLC analyses.

**On the transcription dynamics in
the XlnR regulon of *A. niger* to
D-xylose triggers**

On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

Abstract.

The XlnR regulon in *A. niger* is activated by D-xylose triggers. In this work we measured the time responses of transcription in the XlnR regulon during 5 hours at 20 minutes interval time at two concentration levels of D-xylose (1 mM and 50 mM). From literature, it was derived that after activation by a 1 mM D-xylose pulse the target gene expression reaches a maximum value at around 1 – 2 hours. The model predictions showed that afterwards the transcription goes down due to D-xylose depletion. For 50 mM experiments it was expected that the time response would go to a maximum value with a minor decrease due to the remaining excess of D-xylose. The observed transcription profiles from the 1 mM trigger experiments were as expected, while several genes in the 50 mM experiments showed different. These results illustrate that the effect of CreA or other unknown transcription factors (UTF) should be taken into account while modeling the regulon transcription dynamics. By using frequent data sampling and a sufficiently number of data points, we showed that our proposed model accounts for most of the crucial transcription dynamics in the experimental data.

Keywords: Dynamics, modeling, transcription, D-xylose, CreA, UTF, XlnR regulon, *A. niger*.

This chapter is based on:

J. Omony, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten and A.J.B. van Bortel. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers. 2012. *In preparation*.

6.1 Introduction

The fungus *A. niger* has numerous economic and ecological benefits. It is used in various fermentation processes and has many applications in the pharmaceutical, cosmetic and chemical industries [151]. The fungus is important for enzyme production and it is a precursor for several food products, such as citric acid and gluconic acid. Knowledge about the transcription on genetic level is regarded as one of the keys in advancing its application and controlling the conversions realized by *A. niger*. In contrast to bacterial systems, which are fast, gene transcription in *A. niger* has a response time of a few hours. To consolidate current knowledge on *A. niger* we work on explaining the response characteristics of the target gene transcription profiles in this fungus.

One of the most popular formalisms for modeling the time responses in biological networks is the use of differential equations (see [33, 51, 52, 53, 152, 153, 154, 155] for example applications). These differential equations comprise of mechanistic representations of the transcription rates in biological networks (see for example Klipp et al. [156]). Developing a dynamic transcription model starts by combining prior knowledge, which is mostly in the form of qualitative information with limited time course data. Predictions of the model can be verified or falsified in new time course experiments. Moreover, the early model can be used to find a good experimental design in order to maximize the information from the experiments. New time course experiments focusing on other aspects such as other essential transcription factors are required to enhance the transcription model. This procedure is repeated until the results of the modeling cycle are satisfactory.

In this work we study the transcription dynamics in the XlnR regulon of *A. niger*. The transcriptional regulator XlnR [34] controls regulation of xylanolytic and cellulolytic genes [109]. The XlnR protein also plays a major role in regulating cellulose and hemicellulose degradation and utilization [57, 107, 109]. The XlnR regulon consists of 20 to 40 target genes which are controlled at the level of transcription by XlnR. According to Mach-Aigner et al. [85], it was shown that *xlnR* is constitutively expressed, and the same gene is also known to be repressed by the carbon catabolic repressor protein (CreA).

An early model formulation for the expression of target genes in time by Omony et al. [157], which was based on the information from de Vries et al. [19], allowed partial repression by CreA. In that model repression by CreA was proportional to the D-xylose concentration. As a next step in the modeling cycle, the model was first evaluated to time course data (5 hour time horizon, sampling frequency 1 hour) for the wild type strain (Wt) and the CreA repressed mutant strain by Mach-Aigner et al. [85] and Omony et al. [157]. They showed that the CreA-repression is partial and varies amongst genes over time. They also showed that an additional mechanism is involved. The time resolution was, however, not high enough to cover the full dynamic response characteristics. Therefore as a new step in the modeling cycle, the Wt strain is again considered for a 5 hour time horizon with a sampling time of 20 minutes, a practical

Chapter 6. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

minimum time for sample handling. Moreover, the observation from Mach-Aigner et al. [85] and Omony et al. [157] that at high D-xylose concentration the transcription varies in time, is in conflict with the previous model formulation (see Omony et al. [75]). Therefore, in this work we search for model structures with the least disagreement with observed dynamics in the experimental data.

6.2 Materials and methods

6.2.1 Experimental procedure

Strains and growth conditions. The *A. niger* strains N400 (CBS 120.49) and NW283 (*cspA1*, *fwnA1*, *pyrA6*, *lysA7*, *creA*^{d4}; a strain with a CreA-derepressed phenotype (d, derepressed; 4, allele 4) that is a derivative of NW145, which bears additionally a *pyrA* marker [114]) were maintained for generating spores on complete medium plates [124] supplemented with uridine and lysine. Cultivation of the strains in 2.2-L-bench top fermenters (Applikon Biotechnology, Schiedam, The Netherlands) was carried out in duplicates using each 2 liter medium adjusted to pH 6.0 comprising 1.2 g NaNO₃, 0.3 g KH₂PO₄, 0.1 g KCl, 0.1 g MgSO₄ · 7H₂O, 2 g yeast extract, 4 g casamino acids, 2 mL Vishniac solution [127], and was supplemented with uridine and lysine. 100 mM sorbitol was used as a carbon source. 1 mL antifoam (Sigma Aldrich, St. Louise, MO) was added to the medium to avoid excessive foam formation. Inoculation was done using 1 × 10⁶ spores per mL and cultivation conditions have been adjusted as published earlier [35]. After induction with 1 mM or 50 mM D-xylose, the reference sample was taken within 30 s (pre-induction sample, "pre"). Then 20 minutes sample taking (each 10 mL, post-induction samples, "post") was followed by a microscopic analysis for infection control. Culture supernatant and mycelia were separated by filtration through Miracloth (Calbiochem, part of Merck, Darmstadt, Germany). Mycelia were immediately snap-frozen in liquid nitrogen.

RNA-extraction and reverse transcription. Harvested mycelia were homogenized in 1 ml of peqGOLD TriFast DNA/RNA/protein purification system reagent (PEQLAB Biotechnologie, Erlangen, Germany) using a FastPrep FP120 BIO101 ThermoSavant cell disrupter (Qbiogene, Carlsbad, US). RNA was isolated according to the manufacturer's instructions, and the concentration was measured using the NanoDrop 1000 (Thermo Scientific, Waltham, US). After treatment with DNase I (Fermentas, part of Thermo Fisher Scientific, St. Leon-Rot, Germany), synthesis of cDNA from 0.45 μg mRNA was carried out using the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas); all reactions were performed according to the manufacturer's instructions.

Quantitative PCR analysis. All quantitative PCRs (qPCRs) were performed in a Rotor-Gene Q cyclor (QIAGEN, Hilden, Germany). All reactions were

performed in triplicate. The amplification mixture (final volume 15 μL) contained 7.5 μL 2 \times Absolute QPCR SYBR Green Mix (ABgene, part of Thermo Fisher Scientific, Cambridge, UK), 100 nM forward and reverse primer and 2.5 μL cDNA (diluted 1 : 100). Each run included a template-free control and an amplification-inhibited control (0.015% SDS added to the reaction mixture). The cycling conditions comprised of a 15 min initial polymerase activation at 95°C, followed by 40 cycles of 95°C for 15 s, 59°C for 15 s, and 72°C for 15 s. As previously published, a histone-encoding gene was used as reference gene [35].

Calculations were performed using the equation from Pfaffl [129]. The HPLC analysis was performed using a Thermo Finnigan Surveyor HPLC instrument (Thermo Fisher Scientific, MA, US). All 10 μL samples were injected onto a Repro-Gel Pb column (9 μm , 150 \times 8 mm; Dr. Maisch, Germany). Water was used as the mobile phase and elution was followed at 50°C, applying a flow rate of 1.0 mL/min for 20 min. The concentration was determined using xylitol as internal standard.

6.2.2 List of genes used from the XlnR regulon

The state variables in the model are mRNA concentrations for the genes, namely: *xlnR* (active form of XlnR); the genes encoding for endoxylanases (*xlnB* and *xlnC*); β -xylosidase, *xlnD*; arabinoxylan arabinofuranohydrolase, *axhA*; acetylxylan esterase, *axeA*; α -glucuronidase, *aguA*; feruloyl esterase, *faeA*; the two endoglucanase (*eglA* and *eglB*) also regulated by XlnR [109]. The other genes regulated by the XlnR protein are: *eglC*, *talB*, *xdhA*, *ladA*, *estA*, the D-xylose reductase *xyrA*, *abfB*, *bglA*, *xkiA* (see [57] for details). The XlnR protein also regulates the α - and β -galactosidase genes (*aglB* and *lacA*), respectively [111]; two cellobiohydrolase-encoding genes *cbhA* and *cbhB* [107].

6.3 Results

6.3.1 Transcription measurements

Of the 22 target genes, seven target genes (*eglB*, *eglC*, *estA*, *abfB*, *aglB*, *cbhA*, and *cbhB*) remained around their steady state value and showed no relevant dynamics. The 15 other genes were used for interpretation of its transcription dynamics.

Experimental results for 1 mM and 50 mM induction by D-xylose for the remaining 15 genes are given in Figure 6.1 and 6.2. From these figures we see that the genes respond with different transcription patterns following the low and high D-xylose induction concentrations. Nearly all the patterns in Figure 6.1 for the 1 mM D-xylose concentrations are straightforward and easy to explain. Transcription is activated by D-xylose; it attains a maximum value and then decreases due to depletion of D-xylose in time. Compared to the other genes, the observed transcription profiles for *xlnD* and *bglA* show a sudden drop to zero and a bimodal response, respectively (Figure 6.1). These two transcription pro-

Chapter 6. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

files deviate from the expected pattern following low concentration D-xylose induction.

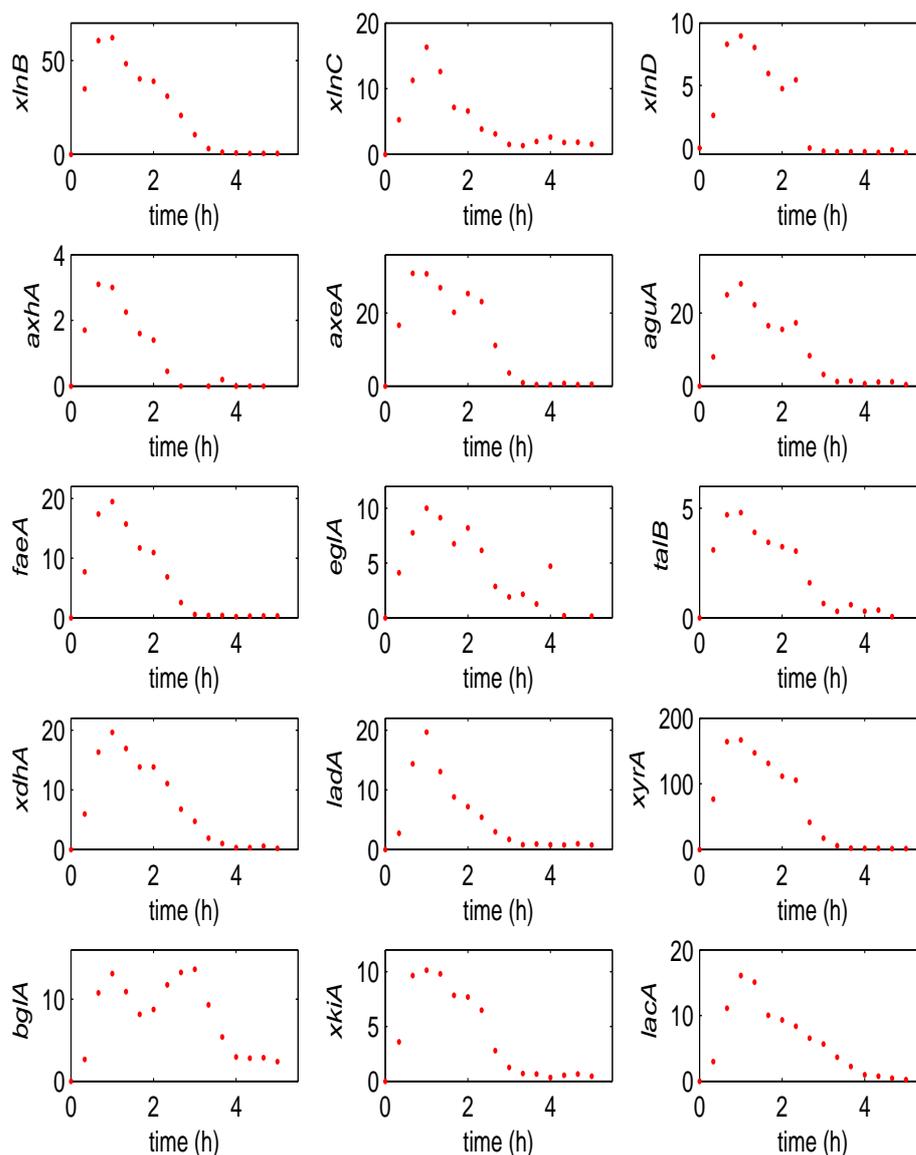


Figure 6.1: Plot of transcription data from the 1 mM D-xylose inducer experiment. Nearly all the D-xylose was metabolized 5 h after induction. The figures are plotted on different scales to aid visibility.

The responses to the 50 mM concentrations are given in Figure 6.2. Although the D-xylose concentration decreases during these experiments from 50 to 38 mM (see Figure 6.3), the level of D-xylose can still be considered as high. Therefore it was expected that the target gene response would follow a mono-

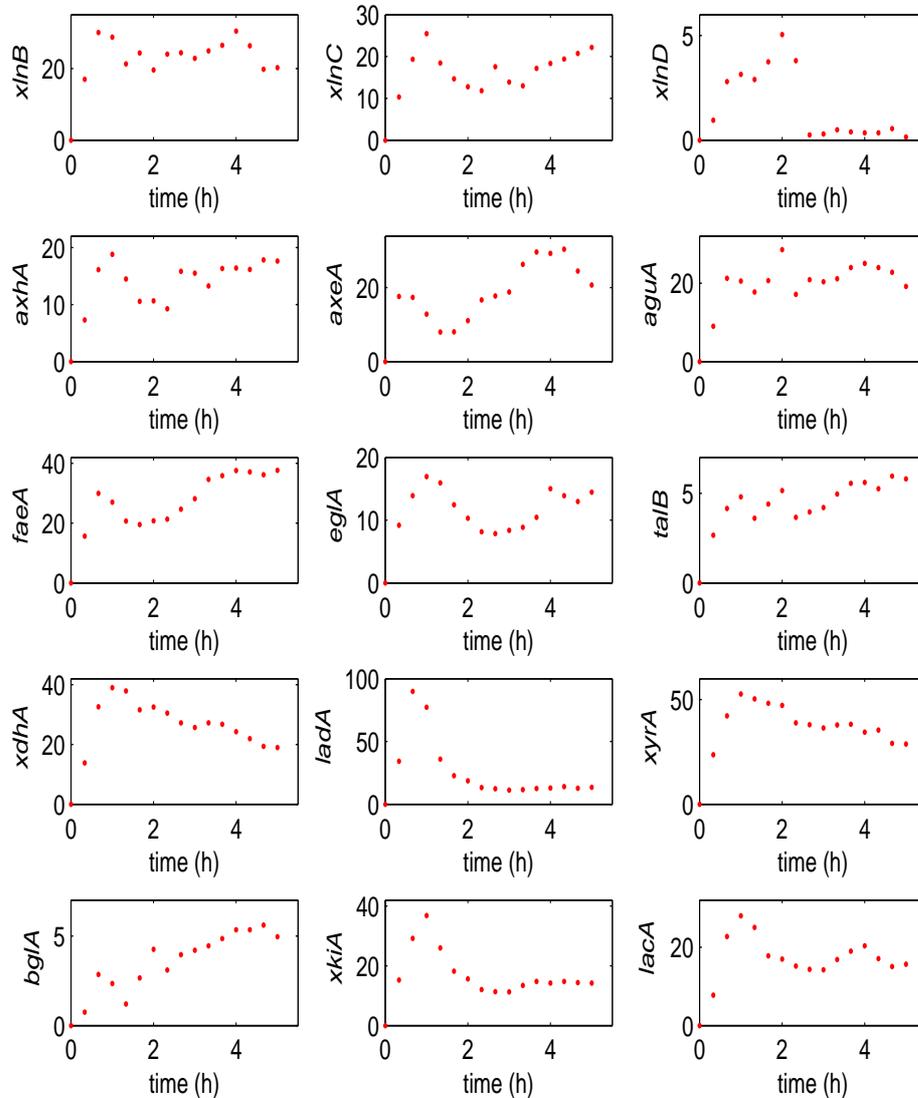


Figure 6.2: Plot of transcription data from the 50 mM D-xylose inducer experiment. On induction the genes show varying transcription profiles. The figures are plotted on different scales to aid visibility.

tonic increasing curve, with an option for a slow decrease due to the falling D-xylose concentration. The responses in Figure 6.2 deviate significantly from this expectation. Overall three different classes of transcription patterns are noticed (Figure 6.4, C_1 , C_2 , and C_3). Class C_1 is the expected monotonic increasing functions towards a steady state value. Transcription under C_2 reaches a maximum value after which it gradually decreases to a new steady state. In classification C_3 transcript numbers steadily rise until reaching a maximum, decrease and

Chapter 6. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

then again increase in the time window used. In the next section we discuss models for these three classifications of transcription profiles.

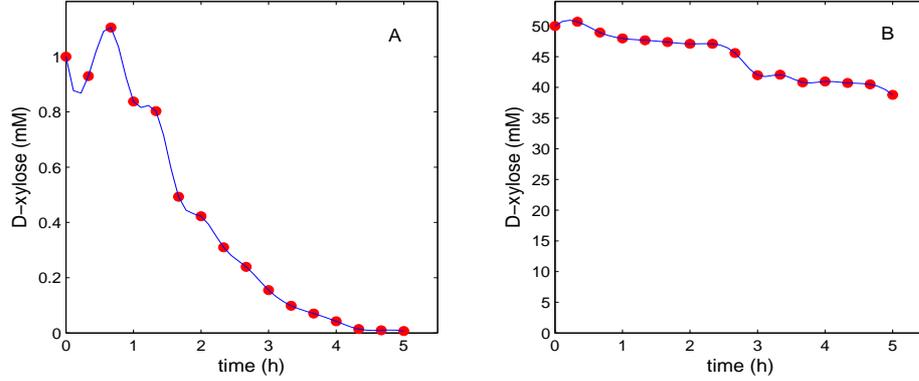


Figure 6.3: **D-xylose consumption.** A: For the 1 mM D-xylose concentration data, nearly all the D-xylose is depleted by 5 h. B: For the 50 mM D-xylose concentration data, about 80% of the initial pulse is left by 5 h.

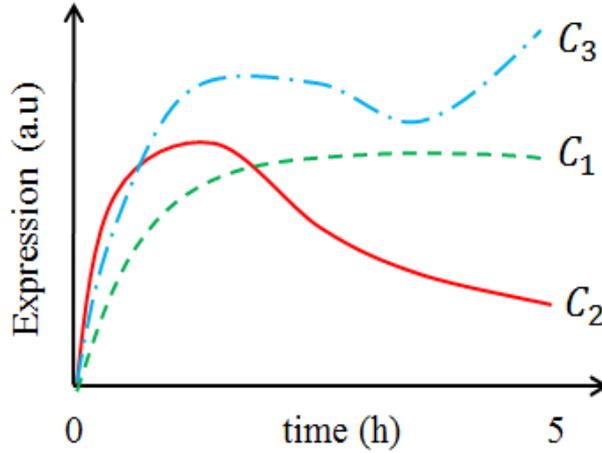


Figure 6.4: **Transcription profiles classification.** The classifications C_1 : a monotonic increasing function, C_2 : a function with a maximum and a lower steady state, and C_3 a function that steadily rises until reaching a maximum, decreases and then again increases.

6.4 Mathematical modeling and discussion

Based on qualitative literature information (de Vries et al. [19]), Omony et al. [150] formulated a first differential equation model for the regulation dynamics for the XlnR regulon:

$$\begin{cases} \dot{x}_1 = -k_{1d}x_1 + b_1u_1, \\ \dot{x}_i = k_{is} \frac{(k_{i1}x_1)^{h_1}}{1 + (k_{i1}x_1)^{h_1}} \frac{1}{1 + (k_{i2}u_1)^{h_2}} - k_{id}x_i, \end{cases} \quad \begin{array}{l} \text{given } x_1(0) = x_{10} \\ x_i(0), u_1(0) \end{array} \quad (6.1)$$

6.4. Mathematical modeling and discussion

In this model x_1 - activity state for the *xlnR* gene, u_1 - D-xylose concentration (in mM), b_1 - external input stimulus coefficient, and k_{1d} - mRNA degradation parameter for *xlnR*, x_i - the activity state for the target gene with index i . The other coefficients in equation (6.1) are defined in the appendix. In their work Omony et al. [150] applied the Hill coefficients $h_1 = h_2 = 1$. A description of the parameters is given in the Appendix section of this article. The model structure describing the regulation of all target genes is the same since all these genes are regulated by the XlnR protein.

The experimental results of Mach-Aigner et al. [85] and Omony et al. [150] showed that *xlnR* is constitutively transcribed i.e. its transcription values were found to be low and nearly constant in time. Using the quasi-steady state (or fast equilibrium) assumption that $\dot{x}_1 = 0$, the first equation in (6.1) gives $x_1 = \tilde{x}_1 = b_1 u_1 / k_{1d}$ and rewriting the expression for \dot{x}_i in (6.1) gives:

$$\dot{x}_i = k_{is} \frac{(k'_{i1} u_1)^{h_1}}{1 + (k'_{i1} u_1)^{h_1}} \frac{1}{1 + (k_{i2} u_1)^{h_2}} - k_{id} x_i, \quad x_i(0), u_1(0) \quad (6.2)$$

where $k'_{i1} = k_{i1} b_1 / k_{1d}$ following a substitution of x_1 in \dot{x}_i , see equation (6.1). For constant D-xylose concentrations, equation (6.2) exhibits a typical first order response; which is a monotonic increasing function. Equation (6.2) fits for the response class C_1 (in Figure 6.4). In the experiments the D-xylose concentrations reduce over the 5 hour time window from 50 to 38 mM (see Figure 6.5). However, this drop in D-xylose concentration is not strong enough, and therefore insufficient to explain the expression class C_2 and gives no explanation for class C_3 .

The model given by (6.2) assumes that catabolic repression of transcription results directly from D-xylose, and is not mediated by the CreA repressor protein. Moreover, D-xylose was also considered to bind as a monomer ($h_1 = 1$) and the possibility of cooperative binding (e.g. as a dimer) was not considered. Omony et al. [157] also showed the existence of a second repressive component other than CreA, which was classified as an unknown transcription factor (UTF). This information is not used in model (6.2). Omony et al. also showed that transcription was still observed in the mutant strain of *A. niger* following a high D-xylose concentration induction.

The response of the XlnR regulon takes time, and therefore it is also expected that it takes time to generate the CreA and UTF repressor proteins. Omony et al. [157] showed that the effect of CreA on transcription is not constant in time. To modify the model (6.2), it is assumed that after D-xylose induction the pools of CreA and UTF gradually build-up to a repressive level. Therefore, equation (6.2) is extended by equations that describe the build-up and repression of the protein concentrations for CreA and UTF:

Chapter 6. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

$$\begin{cases} \dot{x}_{\text{CreA}} = k_1 u_1 - k_2 x_{\text{CreA}} \\ \dot{x}_{\text{UTF}} = k_3 u_1 - k_4 x_{\text{UTF}} \\ \dot{x}_i = k_{is} \frac{(k'_{i1} u_1)^{h_1}}{1 + (k'_{i1} u_1)^{h_1}} \frac{1}{1 + (k_{i2} x_{\text{CreA}})^{h_2}} \frac{1}{1 + (k_{i3} x_{\text{UTF}})^{h_3}} - k_{id} x_i \end{cases} \quad (6.3)$$

Here x_{CreA} represents the amount of CreA protein, and x_{UTF} is the concentration of the UTF. These variables are all functions of time. $k_\ell : \ell = 1, \dots, 4$ are nonnegative parameters. h_2 and h_3 are Hill coefficients of CreA and UTF, respectively.

Compared to (6.2), repression of transcription is now controlled by transcription factors, CreA and UTF. In this model transcription starts at time zero and in the initial phases repression is absent. After 2 hours the protein concentration in the pools for CreA and UTF reaches a repressive level and the transcription of the target gene lowers. This pattern is a typical class C_2 pattern. Small values of the parameters (k_{i2} and k_{i3}) in the repressing functions are associated to weak repression by the transcription factors. With these low values (6.3) allows for the target gene responses exhibited in class C_1 .

The CreA and UTF proteins are repressive. Contrary to what might seem at first sight, the interactive effect of these transcription factors contribute to explaining the resurgence in transcription after 2 hours (see Appendix, Figure 6.8). With the modification given in (6.3), all the three transcription classifications: C_1 , C_2 and C_3 can now be explained by the models. The classification C_3 shows the effect of reactivation of transcription in time, which indicates a de-repression effect as a result of the independent interaction between the contribution of the CreA term and the UTF term in the model (6.3).

Ninfa and Mayo [158] established a positive association between high Hill coefficient values and the transcription dynamics. High $h_{l=1,2,3}$ values are associated with multiple stabilities e.g. bi-stability, and give a steeper shaping of the Hill functions. Binding configurations for monomers ($h_l = 1$), dimers ($h_l = 2$), tetramers ($h_l = 4$), hexamers ($h_l = 6$), and octamers ($h_l = 8$) for CreA, and UTF were investigated. Table 6.1 in the subsequent subsection on "model evaluation" shows the model goodness of fit R^2 - statistic. It satisfies to use dimeric binding for the D-xylose; while for the CreA and UTF higher order binding configurations need to be applied. For instance model fits to data with Hill coefficient configurations of: $h_1 = 1$, and $h_2, h_3 = 2, 4, 6$ yielded averaged R^2 values of about 0.67 for all the genes for both the 1 mM and 50 mM data sets. Therefore, the Hill function coefficient configuration in Table 6.1 was to be preferred.

In (6.3) a high D-xylose concentration favors the pathways that activate XlnR, CreA and UTF. At a low D-xylose induction concentration, the path in which the inactive form of XlnR is activated is favored over the path through CreA or any other UTF. While modeling using the experimental data, the effect of the feedback depicted in Figure 6.5 which was first thought to be needed to

explain the behavior in class C_3 was assessed and found to be non-significant on the transcription dynamics. Therefore, the feedback loop was left out of the final model used in (6.3).

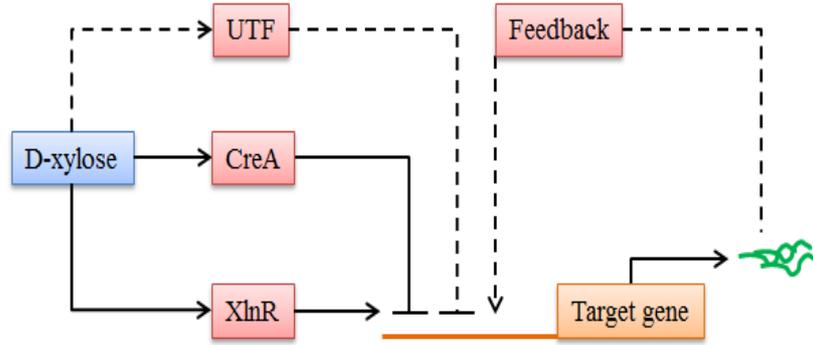


Figure 6.5: **Regulation mechanisms in the XlnR regulon.** The term UTF represents the unknown transcription factor. An indication of how the metabolites XlnR, CreA, and UTF control transcription of target genes in the XlnR regulon.

6.4.1 Procedure for model fit to experimental data

Nonlinear mechanistic models may be powerful in capturing network dynamics but they also come with their down sides. Their downside is that they usually contain more parameters than the number of measured variables and data points. For instance the models in equation (6.3) have a total of 9 unknown parameters. One of the main challenges is obtaining good parameter estimates with high precision levels using the experimental data. The problem with having many parameters in a system of equations is that there is a high likelihood that some of the parameter estimates are correlated. The existence of correlation biases some of the parameter values. This problem is also likely to result into getting stuck at a local minimum during the optimization. In such a case the parameter uncertainty bounds are wider and there may be many parameter combinations that portray a good model fit to data.

The model was simultaneously fitted to the 1 and 50 mM the experimental data. A minimization of the sum of squared errors between the measured data and estimated data values was used in the optimization. The combined goal function consists of a sum of two goal functions, one for the model fit to the 1 mM data and the other model fit to the 50 mM data. The models specified in (6.3) were fitted to the data and this fit resulted into a single set of parameter estimates. The model fits were performed one gene at a time (see Figure 6.6 and 6.7).

The optimization goal functions that were used are specified as follows. Let $\mathcal{J}_{i,1}(\theta_i)$ and $\mathcal{J}_{i,50}(\theta_i)$ be goal functions for the 1 and 50 mM set up, respectively; $\theta_i = [k_1, k_2, k_3, k_4, k_{is}, k_{i1}, k_{i2}, k_{i3}, k_{id}]^T$ is the parameter vector to be estimated,

Chapter 6. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

and i is the target gene index. The individual goal functions are given by:

$$\mathcal{J}_{i,1}(\theta_i) = \min \sum_{j=0}^N (y_{ij} - \hat{y}_{ij})^2, \text{ and } \mathcal{J}_{i,50}(\theta_i) = \min \sum_{j=0}^N (y_{ij} - \hat{y}_{ij})^2$$

The combined goal function is given by: $\mathcal{J}_{\text{combined}}(\theta_i) = \mathcal{J}_{i,1}(\theta_i) + \mathcal{J}_{i,50}(\theta_i)$. Optimizing this goal function results in the optimal parameter vector estimate $\hat{\theta}_i$ (see Table 6.3, in the Appendix). The optimization was performed using the *lsqnonlin* routine in MATLAB.

In the process of the model fit to data, first all the parameters in $\hat{\theta}_i$ were estimated for genes, then the average values of the estimates \hat{k}_1 and \hat{k}_2 computed. This is to ensure that the CreA dynamics is the same for all the target genes as it should be. The computed average values $\bar{\hat{k}}_1$ and $\bar{\hat{k}}_2$ were fixed and the remaining parameters re-estimated. Using the models in (6.3), two considerations were made for the UTF. In the first case the parameters k_3 and k_4 were considered to be constant for all genes (a single UTF acting on all relevant genes) and in another case they were considered to be varying between the genes (separate UTFs per target gene). Of the two cases, the latter resulted into better model fits to data.

6.4.2 Model evaluation

The model evaluation was performed by fitting (6.3) through the data points obtained from the combined data for the 1 and 50 mM D-xylose trigger experiments as described in the model fitting procedure above. The results are given in Figure 6.6 and 6.7. For the 1 mM experiments overall good model fit to data is obtained for most of the 15 target genes considered in the evaluation (Figure 6.6). Of all the genes, we observe that the transcription profile for *bglA* deviates from the general pattern around the 2nd hour. The bimodal transcription pattern for this gene is not covered by the three transcription classes depicted in Figure 6.4.

For the 50 mM experiment the models in (6.3) satisfies nearly all of the 15 target gene transcription profiles. The fit for the expression of *xlnD* seeks to compromise between the peak value around 2 hours and the constant low expression level from 3 hours. The measured expressions for *axeA* and *xyrA* exhibit a slight systematic error over time. The measured variables for the fit concerned the transcription values and the actual D-xylose input. The models in equation (6.3) also comprise the state variables CreA, and UTF. It is still a challenge to accurately measure these protein levels. In our experimental situation we did not manage to quantify them.

The model for the transcription dynamics described in (6.3) upon evaluation on the data revealed that under the 1 mM D-xylose conditions, the effect of CreA and UTF is minimal i.e. $\psi_{\text{CreA}} \approx 1$ and $\psi_{\text{UTF}} \approx 1$ (see Appendix section for interpretation of functions). Under low induction conditions with D-xylose, the activation by D-xylose described by the Hill function $\psi_{\text{D-xylose}}$ and the mRNA

6.4. Mathematical modeling and discussion

degradation $k_{id}x_i$ term determine the systems dynamics. However, for the high D-xylose induction, the Hill functions $\psi_{D\text{-xylose}}$, ψ_{CreA} and ψ_{UTF} collectively regulates transcription (see Figure 6.8, Appendix). Another challenge that arises with the formulation in (6.3) is that it neither accounts for nor distinguishes the transcription dynamics when a mutant strain is used from when a wild type strain is used.

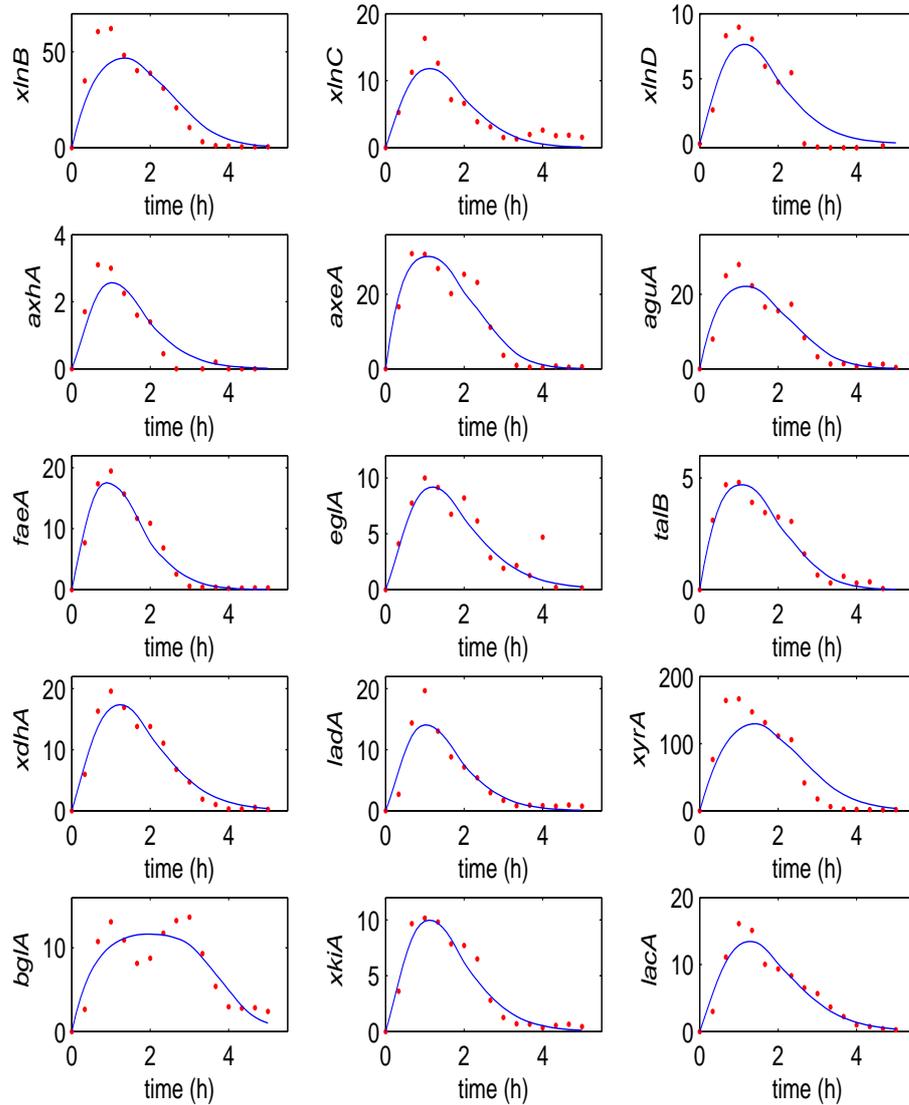


Figure 6.6: **Model fit to transcription data from the 1 mM D-xylose inducer experiment.** Nearly all the D-xylose was metabolized 5 h after induction. The figures are plotted on different scales to aid visibility. The models were fitted using Hill coefficients $h_1 = 2$, and $h_2, h_3 = 8$. The production and degradation parameters corresponding to CreA were fixed at $\hat{k}_1 = 0.3098$ and $\hat{k}_2 = 1.4131$, respectively.

Chapter 6. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

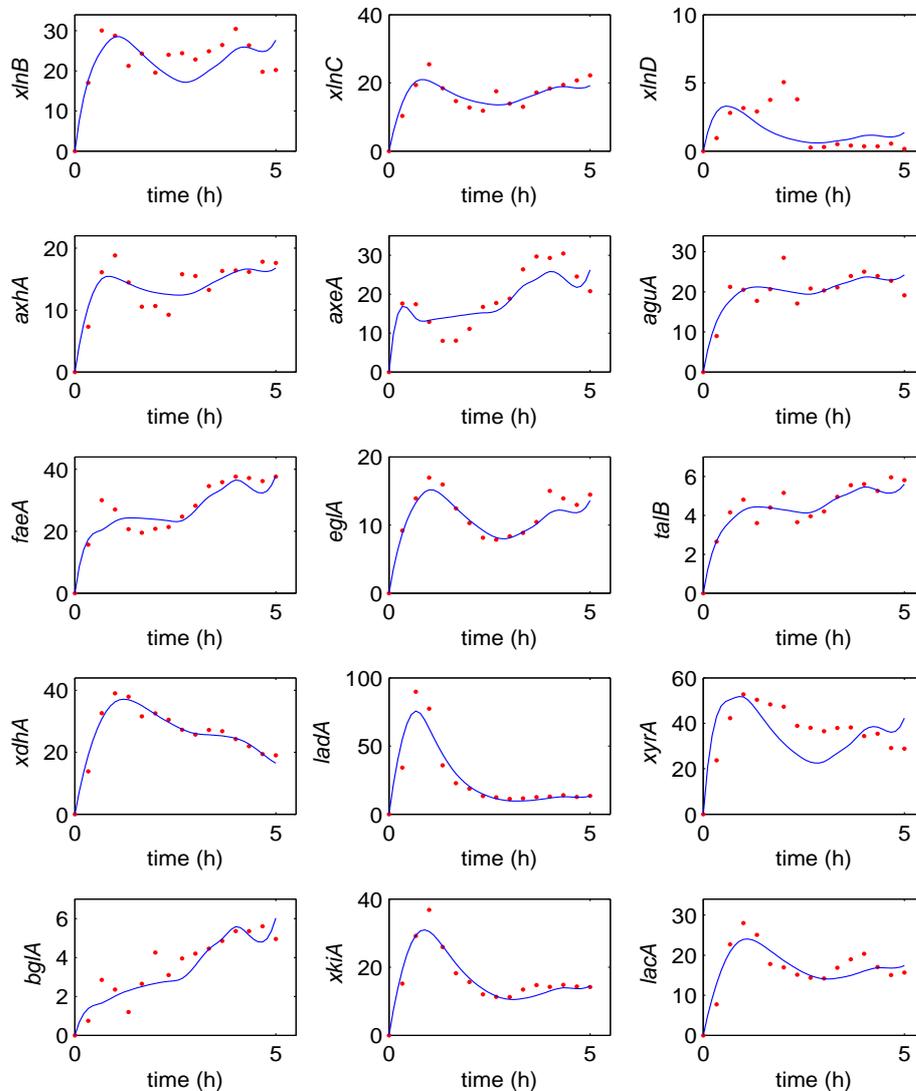


Figure 6.7: **Model fit to transcription data from the 50 mM D-xylose inducer experiment.** On induction the genes show varying transcription profiles. The figures are plotted on different scales to aid visibility. The models were fitted using Hill coefficients $h_1 = 2$, and $h_2, h_3 = 8$. The production and degradation parameters corresponding to CreA were fixed at $\hat{k}_1 = 0.3098$ and $\hat{k}_2 = 1.4131$, respectively.

To gain further insight into the network inference, the derived parameter estimates $\hat{\theta}_i$ were used to display the Hill function components from (6.3). The results show that the interplay of roles between the repressing functions ψ_{CreA} and ψ_{UTF} for CreA and UTF, respectively, varies between the genes (as seen from Figure 6.8, Appendix). For instance with *xlnC*, *axeA*, *axhA*, *ladA*, *bgIA*, *xkiA*, and *lacA* shows that only a single transcription factor played a role in

regulating transcription. For the other genes, both transcription factors CreA and UTF contributed to regulating transcription.

Gene	R^2 -statistic		
	R_1^2	R_{50}^2	R_{combined}^2
<i>xlnB</i>	0.8611	0.4374	0.6493
<i>xlnC</i>	0.8620	0.8541	0.8580
<i>xlnD</i>	0.8819	0.0399	0.4609
<i>axhA</i>	0.9154	0.8108	0.8631
<i>axeA</i>	0.9301	0.7926	0.8614
<i>aguA</i>	0.9034	0.7676	0.8355
<i>faeA</i>	0.9525	0.8935	0.9230
<i>eglA</i>	0.8719	0.9240	0.8979
<i>talB</i>	0.9611	0.9212	0.9411
<i>xdhA</i>	0.9637	0.9561	0.9599
<i>ladA</i>	0.8895	0.8967	0.8931
<i>xyrA</i>	0.8401	0.0026	0.4214
<i>bglA</i>	0.7949	0.7989	0.7969
<i>xkiA</i>	0.9529	0.9126	0.9327
<i>lacA</i>	0.9283	0.8001	0.8642

Table 6.1: **Assessing model goodness of fit.** The Hill coefficients used for the model fit were $h_1 = 2$, and $h_2, h_3 = 8$. The goodness of fit to data is given by the R^2 - statistic for the two data sets. Here R_1^2 and R_{50}^2 are the R^2 - statistics for the 1 mM and 50 mM data sets; R_{combined}^2 is the R^2 statistic for the fit on the combined data sets.

6.5 Conclusion

Understanding regulatory mechanisms in biological networks involves a modeling cycle with consolidation of prior knowledge, experimental design, data collection, hypothesis testing, model evaluation and model adjustment. For the XlnR regulon, studying the transcription network dynamics has provided further insight into its regulation mechanism. Description of the regulatory mechanisms were up till now based on one time point, or with one hour time intervals. These experimental methods do not reveal all the characteristics of the transcription dynamics. In this work we used sample interval times of 20 minutes, which revealed patterns in the response of the target genes in the XlnR regulon that were not yet described in previous studies.

For low D-xylose concentrations induction of the target genes is followed by a decline of transcription due depletion of D-xylose. This pattern corresponds to qualitative descriptions given in previous studies. High concentrations result in three classes of the transcription profiles: 1) a monotonic increasing function, 2) a function with a maximum and a lower steady state, and 3) a function that decreases and then again increases after reaching a maximum. These patterns were not yet reported in literature and could only be observed because of the

Chapter 6. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

chosen low sample time interval. In this paper we worked at the minimal possible interval time because of all (or partly manual) handling procedures. Decreasing the sample interval time has the potential to reveal the transcription response even more accurately.

Based on experimental data for transcription of the XlnR regulon, we proposed three regulatory mechanisms for the transcription dynamics: 1) activation by *xlnR*, 2) regulation by CreA and UTF protein, which become effective when their concentrations have been built up in time (1 – 2 hours), and 3) activation by the interactive effect of the transcription factors CreA and UTF (time scale 3 – 4 hours). The dynamics of nearly all target genes observed in the data sets for 1 and 50 mM D-xylose induction fit to the proposed model for these regulatory mechanisms.

From the experimental and modeling results we conclude that the transcription of the target genes in the XlnR regulon is more complex than derived from the initial studies for single data points or data points with a large interval time. The dynamics of the response was overlooked in previous studies. Just as we have experienced in this work, we expect that analysis of the dynamic responses of genes will provide better understanding of the mechanisms involved in gene transcription of other organisms.

Acknowledgement

This work is supported by the graduate school VLAG and the IPOP program of Wageningen University. The authors would like to thank Birgit Jovanovic of the Institute of Chemical Engineering at Vienna University of Technology for the assistance with the HPLC analyses.

Authors' contributions

ARMA performed the experiments and data processing. JO wrote the manuscript, performed the data analysis and modeling. LHdG provided the biological knowledge that was used for the model formulations. GvS and AJBvB also contributed in calculations and critical review of the methods used in the analysis. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

6.6 Appendix

6.6.1 Nomenclature - variables and parameters

i	- target gene index.
k_{is}	- mRNA synthesis parameter.
k'_{i1}	- effective affinity constant for gene 1 activating gene i (mM^{-1}).
k_{i2}	- inverse of Hill constant for CreA.
k_{i3}	- inverse of Hill constant for an UTF.
k_{id}	- mRNA degradation parameter.
$N = 16$	- maximum number of samples taken in the 5 h time-frame.
t	- time in hours (h).
u_1	- D-xylose concentration from two duplicate experiments.
x_1	- activity state of $xlnR$.
x_i	- activity state of target gene.
\dot{x}_i	- transcription rate.
y_{ij}	- measured data value at time instant j .
\hat{y}_{ij}	- estimated data value at time instant j .
UTF	- unknown transcription factor.
$\psi_{\text{D-xylose}} := \frac{(k'_{i1} u_1)^{h_1}}{1 + (k'_{i1} u_1)^{h_1}}$	- regulation Hill function by D-xylose.
$\psi_{\text{CreA}} := \frac{1}{1 + (k_{i2} x_{\text{CreA}})^{h_2}}$	- repressing Hill function by CreA.
$\psi_{\text{UTF}} := \frac{1}{1 + (k_{i3} x_{\text{UTF}})^{h_3}}$	- repressing Hill function by UTF.

Chapter 6. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers

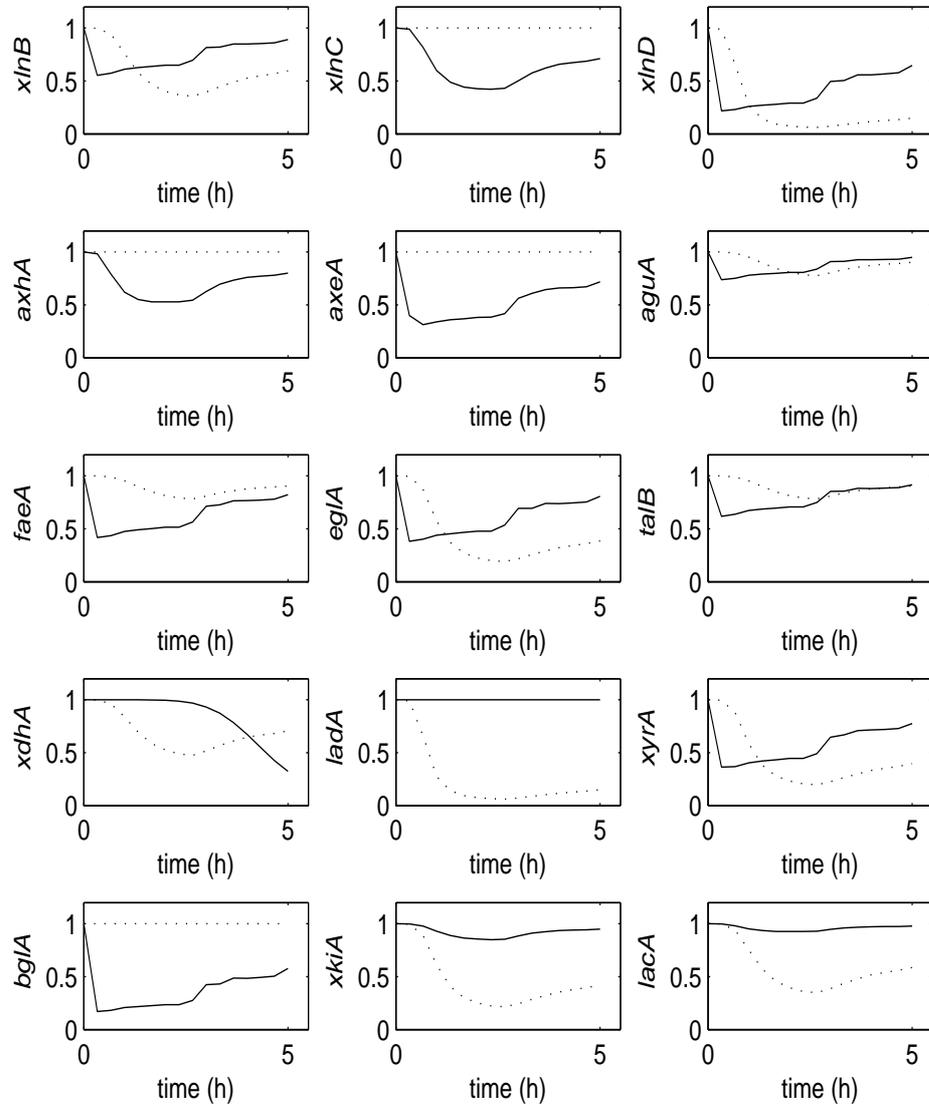


Figure 6.8: Hill function plots in time. Plot of regulation Hill functions for 15 target genes, each plot contains combined functions for the 50 mM D-xylose data. The solid lines and dotted lines correspond to the CreA Hill function ψ_{CreA} and UTF Hill function ψ_{UTF} for the 50 mM D-xylose data.

Gene	Parameter estimates and standard deviations						
	$\hat{k}_3 \pm \sigma_{\hat{k}_3}$	$\hat{k}_4 \pm \sigma_{\hat{k}_4}$	$\hat{k}_{is} \pm \sigma_{\hat{k}_{is}}$	$\hat{k}'_{i1} \pm \sigma_{\hat{k}'_{i1}}$	$\hat{k}_{i2} \pm \sigma_{\hat{k}_{i2}}$	$\hat{k}_{i3} \pm \sigma_{\hat{k}_{i3}}$	$\hat{k}_{id} \pm \sigma_{\hat{k}_{id}}$
<i>xlnB</i>	0.6334 ± 12057	30.000 ± 78.310	120.00 ± 17.179	2.1800 ± 1.0366	0.1068 ± 6.0396e-3	0.9312 ± 17728	1.8531 ± 0.7098
<i>xlnC</i>	0.3105 ± 32841	2.1232 ± 0.5323	56.832 ± 6.0763	0.9287 ± 0.1226	0.0108 ± 43540	0.1515 ± 16026	1.8469 ± 0.3532
<i>xlnD</i>	1.6170 ± 27434	30.000 ± 598.60	37.930 ± 53.291	0.8666 ± 1.3038	0.1394 ± 0.0284	0.4402 ± 7477.1	1.7216 ± 0.9940
<i>axhA</i>	0.3260 ± 1.5286e+5	2.5700 ± 1.4836	42.702 ± 6.9113	0.3971 ± 0.0603	0.0100 ± 77424	0.1648 ± 77272	1.8479 ± 0.6196
<i>axeA</i>	0.4597 ± 919.48	8.5735 ± 2.7451	122.52 ± 19.615	1.9933 ± 0.4140	0.0100 ± 22964	0.4199 ± 839.71	3.0552 ± 0.7352
<i>aguA</i>	1.0538 ± 4417.9	40.000 ± 181.40	76.200 ± 12.432	2.0000 ± 0.4768	0.0849 ± 7.6019e-3	0.6742 ± 2826.8	2.5294 ± 0.6601
<i>faeA</i>	0.9326 ± 17690	20.855 ± 9.1660	159.00 ± 27.077	0.7194 ± 0.0642	0.0845 ± 4.8561e-3	0.4710 ± 8934.8	2.8356 ± 0.4822
<i>eglA</i>	1.4056 ± 6095.6	50.000 ± 52.868	79.548 ± 20.567	0.5051 ± 0.0993	0.1188 ± 4.2754e-3	0.7635 ± 3311.2	1.2463 ± 0.1499
<i>talB</i>	1.2082 ± 307.16	37.932 ± 1.2143e-3	19.630 ± 1.6640	1.5446 ± 0.1752	0.0845 ± 4.9785e-3	0.5972 ± 151.83	2.7027 ± 0.3294
<i>xdhA</i>	0.0465 ± 34153	0.0314 ± 0.7648	71.672 ± 3.1823	0.9039 ± 0.0421	0.1004 ± 2.5648e-3	0.1142 ± 84035	1.4112 ± 0.1241
<i>ladA</i>	1.6435 ± 6.5026e+6	42.350 ± 8.4870e+5	222.84 ± 89.940	0.4095 ± 0.1326	0.1392 ± 0.0108	0.0365 ± 3.7044e+6	1.8606 ± 0.4936
<i>xyrA</i>	0.4132 ± 938.02	15.000 ± 16.215	300.00 ± 40.705	2.0001 ± 1.2234	0.1181 ± 7.6292e-3	0.7911 ± 1796.0	1.5398 ± 0.6525
<i>bgIA</i>	0.8935 ± 107.30	32.788 ± 72.552	22.149 ± 4.1419	9.2345 ± 2.7767	0.0101 ± 56154	0.9034 ± 108.72	1.7809 ± 0.4453
<i>xkiA</i>	0.4241 ± 2.5584e+5	1.9861 ± 1.7135	74.492 ± 5.1531	0.6156 ± 0.0496	0.1165 ± 5.3205e-3	0.0804 ± 48504	1.6027 ± 0.2758
<i>lacA</i>	0.4102 ± 37334	2.4163 ± 7.3644	47.813 ± 4.6073	0.9870 ± 0.1408	0.1073 ± 9.0374e-3	0.0912 ± 8298.4	1.3001 ± 0.2686

Table 6.3: **Parameter estimates.** Results showing the parameter estimates and their standard deviations. The production and degradation parameters corresponding to CreA were fixed at $\bar{k}_1 = 0.3098$ and $\bar{k}_2 = 1.4131$, respectively.

Chapter 7

Retrospectives, perspectives and contributions

7.1 Retrospectives and perspectives

7.1.1 Introduction

At the beginning of this research we set out to find answers to a number of research questions on network inference as specified in Chapter 1. This quest led to a number of interesting findings as well as new challenges. In the beginning most of the work was focused on network modeling and simulation, using linear dynamic models and regression on synthetic data. In that stage a regression model fitting was used at unraveling the underlying network connectivity structure (contents not given in this thesis). As is often the case, efforts to unravel network structures are hindered by noise in the data. Similarly, in our case the attempt turned out to be very difficult. In particular, depending on the signal to noise ratio, links were identified that did not exist in the test system. Though not very successful, these attempts were valuable to decide on other pathways for network inference. The linear interconnection models were abandoned, and the focus was shifted to using nonlinear mechanistic models with differential equations, which were considered to be more promising for addressing the research objectives. Differential equations were thus used to study different aspects of the sample case, the XlnR regulon - yielding interesting results. These model formulations were partly based on prior knowledge of the XlnR regulon. The achievements, challenges and recommendations are given in the subsections below.

7.1.2 Modeling and simulation of the XlnR regulon dynamics

Throughout literature, mathematical modeling has been very instrumental for studying biological networks. In Chapter 2, modeling and simulation of the XlnR regulon dynamics was performed based on the available prior knowledge of the regulatory mechanisms of the XlnR regulon. We started by gathering qualitative and quantitative information on target gene regulation from the work of van Peij [34, 109], de Vries et al. [19], van der Veen et al. [35], and Hasper [57]. Consolidating the qualitative information and converting them to a modeling framework to describe the network reaction mechanisms were no easy task. This was attributed to the absence of previously available models describing the XlnR regulon dynamics and the lack of suitable experimental data from previous studies.

Interesting properties of biological networks can be understood through studies on: i) system steady states, ii) the influence of network motifs associated with negative or, iii) positive feedback on transcription, and iv) necessary conditions for oscillation in transcription. In these simulations, the effects of time delay, feedback, and promoter site activity were investigated. Modeling of the XlnR regulon is an achievement in itself since prior to the work reported in this thesis, no publications on the theoretical modeling framework and experimental validation existed. The modeling frameworks used in this thesis are powerful enough to describe the interaction between the network components.

Generally, feedback mechanisms play an important role in the complexity of cellular processes. Feedbacks help organisms adapt to their environment in response to external perturbations. Feedback loops are also important for regulating intracellular processes and maintaining the system's state of equilibrium. In literature, feedback mechanisms and network dynamics are well studied for many simple network structures. However, for complex network structures, the influence of feedback loops is difficult to assess. Fortunately, the XlnR regulon is a sparse network and the influence of the feedback loop was successfully assessed. It must be noted that the soundness of the modeling and analysis is based on how much prior knowledge one has accumulated. The kinetic models provided a good description of the regulation mechanism. The mechanistic differential equations with nonlinear kinetics used in Chapter 2 provide an excellent platform for understanding the dynamics observed from experimental data.

7.1.3 On strategic design of time course experiments

For a long time biological and medical sciences have relied on observational data. The use of perturbation experiments, as is often practiced in engineering sciences, e.g. control theory and physics, has greatly changed the way biological networks are studied. In the past, several studies have demonstrated the benefits of network perturbation experiments (see e.g. Ideker et al. [159]; Friedman [160]; Tegnér et al. [161]). The use of (optimal) experimental designs of time course experiments has not yet been widely exploited in Systems Biology. For these reasons, we focused on devising strategies to enrich experimental data through network perturbation.

By starting with a known model structure and some a priori knowledge on the parameters obtained from previous experiments, we successfully simulated and evaluated various network perturbations with D-xylose as an inducing molecule (Chapter 3). If the parameters are unknown a pilot experiment with less frequent sampling should be performed for a preliminary model fit. Using the estimated parameters, the model can then be used in *in silico* simulations to design a fine-tuned experiment with more frequent data sampling. This iterative process enables acquisition of highly informative data, thus improving the parameter estimation and network overall inference.

Other important principles in the study of biological networks are perturbation and sensitivity analysis. Cells are always exposed to an influx of signals from their immediate environment. These signals control the way the genes in the cell are transcribed. The noise in the transcription and translation mechanisms may lead to changes in gene expression, protein and metabolite concentrations. This results in a need for re-establishing the required levels of mRNA, proteins and metabolites, therefore, maintaining the cellular function. The sensitivity analysis performed in our work in Chapter 3 enabled classification of those parameters that are easily identifiable. In this chapter we also showed that the use of the second pulse instead of just a single pulse of D-xylose im-

proved the experimental design. The second inducer pulse should be given at an effective moment in time; in this case when the inducer concentration became low. The results also indicate that for time course experiments, data sampling intervals can be optimized with respect to the accuracy of the parameters, available resources and labor costs.

The idea of using multiple pulses in time to perturb biological networks with the goal of improving the data information content has not been exploited in literature. The success of network reconstruction also relies on data resolution, number of data points sampled and the data sampling frequency among others. According to Tchangang et al. [162] time series data can be broadly divided into two classes: i) the short-time series with few sampled time points (typically between 3 to 8), and ii) the long-time series with more than 10 time points sampled. Such a division can be misleading depending on what aspects of network inference one is interested in, and therefore should be with respect to what research question one is interested on answering. In essence, for simple comparative studies of transcription between experimental settings, it might suffice to use the short-time series data, while for detailed modeling of network dynamics and parameter inference more data points are required.

Time course data are typically: i) sparse in terms of the number of replicates per sample and the number of time points per replicate, and ii) irregularly spaced in time [163]. According to Ernst et al. [164] more than 80% of all time series data sets surveyed contained less than 9 points. The rarity of long time-series data sets is attributed to financial costs, and the difficulty to obtain biological materials in large quantities. Given that most time course experiments involve less than 10 data points, it is crucial to carefully plan the data sampling strategy. This is where much can be gained from studies involving both wet-lab and synthetic data, such as those used in Chapter 3. In essence, studies aimed at the development of computational methods that get the most out of limited data are still required in Systems Biology.

7.1.4 On the qualitative and quantitative effects of CreA on transcription

It is known that CreA represses transcription, but at the onset of the work in this thesis, the qualitative and quantitative effects of CreA were unknown and even less the dynamics in time. We therefore set out to provide insight on the regulatory properties of CreA, and quantified its effect on transcription. The study of dynamics is essential for investigating the performance of the network components. In Chapter 4 the experiments were performed to check for the existence of dynamics in the transcription profiles of the genes in the XlnR regulon. The results confirmed the presence of dynamics. High resolution qPCR data with duplicates experiments at each time instant were used for the study.

Differences in transcription were observed between the 1 and 50 mM D-xylose induced settings, and also between the Wt and Mt strains. These experiments formed a basis for further experiments with higher sampling frequencies

7.2. My take on biological network reconstruction

(i.e. every 20 minutes compared to hourly sampling). The results from Chapter 4 lead to the conclusion that there are other regulation mechanisms involved in the functional regulation of XlnR activity and that they might be present at the post-transcriptional level. From this finding, it can be concluded that there is need for further experiments aimed at identifying these post-transcriptional regulation mechanisms.

The study this Chapter 5 is a quantitative follow-up to the qualitative assessment of the effect of CreA in Chapter 4. To successfully model the transcription dynamics of the genes using time course data, there was need to understand the effect of CreA on the target gene regulation. Experiments aimed at comparing transcription in Wt and Mt strains were performed, from which the partial repression effect of CreA and the effect of another unknown transcription factor (UTF) were inferred. Studies like this are very important since they elucidate on the changes that need to be made on the models for the network dynamics. With the findings, inferences, and conclusions from Chapter 5, the model structures proposed earlier were then revisited to make the necessary modifications based on observations from experimental data. The new model structure is given in Chapter 6.

7.1.5 Modeling and validation of transcription dynamics

In retrospect to the models used in Chapters 2 and 3 and the availability of the qPCR experimental data sets, a new model structure was proposed for the work in Chapter 6. The need for this new model structure stems from the experimental data. The amendments in the models accounted for the partial effect of CreA on transcription, and the introduction of extra differential equations to account for the dynamics and build up in concentration of CreA and UTF. Sometimes observations from experimental data do not fully conform to proposed model structures. Following the 1 and 50 mM experiments in Chapter 6, the transcription profiles were observed to vary between genes as well as in time. For this reason, the earlier proposed model structures used in Chapters 2 and 3 are insufficient to explain the dynamics observed from particularly the 50 mM D-xylose experimental data (see Figures 6.6 and 6.7). In our case, a few genes showed transcription profiles that significantly deviated from the other genes. These data profiles could not be fully explained with the models used in Chapter 2 and 3, hence the need for the modifications used in Chapter 6. The main difference is on how the regulation mechanism for *xlnR* is modeled.

7.2 My take on biological network reconstruction

The study of biological networks is a tedious, painstaking process. More often than not prior knowledge on a particular network is not available, except for extensively studied networks such as the SOS DNA repair network of *E. coli* [165], and the *S. cerevisiae* [38]. Unfortunately many biological networks have not yet been studied to that level; therefore, there is limited prior knowledge on

Chapter 7. Retrospectives, perspectives and contributions

their regulatory pathways and functionality of the metabolites that constitute it. Especially for networks that are not yet extensively modeled, like the XlnR regulon, much can still be achieved by starting from whichever prior knowledge is available. The reliance on measured data for the modeling can be limiting in case of missing data, and/or some variables cannot be easily quantified. This is a problem that is not limited to a particular inference method.

It is vital to approach every network with an open mind, since even the smallest of details can turn out to be crucial for its dynamics. Sometimes one has to expect the unexpected; for instance the highly nonlinear transcription dynamics following induction with 50 mM D-xylose. The differential equations approach is a popular formalism for modeling networks using transcription time course data. However, while using mechanistic models, e.g. the Hill functions and/or Michaelis-Menten functions in the transcription-translation models, some variables may not be easily experimentally quantifiable and/or may be correlated. This is particularly problematic when there is need for detailed modeling of the network dynamics and only few variables are quantified. In the event that some variables cannot be measured, not all is lost. A lot can still be achieved by inferring from the measured variables.

With recent advancements in modeling techniques and the invention of high through-put data quantification equipment, some of the above challenges will be overcome. The process of network inference should be accorded the same effort as the planning of the experiments. This is because network inference starts right from gathering of prior knowledge, planning for and carrying out experiments, and progresses through to the modeling, and parameter identification.

Biological networks have different levels of complexity. The complexities in large networks are difficult to unravel, but the smaller networks are not easier to deal with (see e.g. Widder et al. [166]). Less might be more in modeling biological networks as stated by Bornholdt [167], but there is a catch. Too much simplicity in a network can lead to loss of details in its regulation dynamics, while a very complex network is confusing. Hence, there is need for a balance between complexity and simplicity. Some suggestions on how to deal with this issue can be found in literature, e.g. identification of essential hubs (nodes with higher connectivity), and functional modules in a larger network as a preliminary step to the analysis as for sparser network structures, it poses less of an identification problem. This begs the question how sparse is sparse? Here sparseness refers to the network connectivity which describes the number of regulatory interactions between the genes in a network. The answer to this question is not definitive. As long as the essential biologically functional modules are preserved, then the network can be considered to be sufficiently sparse. The need for sparseness in a network is often motivated by insight from earlier studies on genetic network reconstruction, particularly work that demonstrates that networks in most biological systems are sparse (see e.g. [168, 169, 170, 171]).

Following a detailed look at the state-of-the-art in network inference, we see that a lot of effort has been devoted to methodology development as well

7.2. My take on biological network reconstruction

as to application to real life problems. However, while much emphasis has been given to methodology development on synthetic or believed-to-be-real networks, the application of these methods to addressing real-life problems in life sciences as a whole still lags behind. In the last decade significant attention was given to applications in various organisms. There is optimism basing on the contributions realized in this thesis that this trend will climb steadily, and not stagnate - ultimately providing solutions to many real-life problems.

State-of-the-art in network reconstruction

A.1 Introduction

At the start of the research that constitutes this thesis a search for an appropriate model formalism to reconstruct the XlnR was performed. This search led to a number of candidate inference methods, of these methods the differential equation approach was considered the most appropriate. An overview of the most popular network inference methods is given below.

A.2 The (Probabilistic) Bayesian Network formalism

The use of Bayesian networks was first introduced by Kauffman [29]. It is an approach based on conditional dependencies between sets of variables (see [33] for a review). Applications of *dynamic Bayesian networks* (DBNs) in network reconstruction can be found in the work of Kim et al. [172], and Zou and Conzen [173]. Unlike the simple Bayesian network (BN), the DBN can model cyclic regulation¹ in genetic networks [174]. The DBN as an extension of the BN incorporates time dynamics into the network reconstruction². According to Murphy and Mian [178], Bayesian networks are highly stochastic; thus, suitable for modeling with noisy transcription data [179]. The Bayesian formalism is efficient but requires many working assumptions and good network structural prior knowledge. It can also readily handle missing microarray data sets [180].

Rogers and Girolami [181] used Bayesian regression to infer sparse genetic networks and noted that the likelihood for observing false edges remains high, especially for noise levels higher than 10%. They observed that typically precision levels drop to about 10% and that for each true connection, 9 false connections were discovered. This points to complications that arise from parameter

¹True biological networks exhibit cyclic regulations.

²Application of DBNs in reverse engineering networks can be found in the works of Werhli et al. [175], Spirtes et al. [176] and Pearl [177].

variations in modeling genetic networks. Identifying which approach is best suited for a given data set is nontrivial [32] since the efficiencies may vary for different data sets.

A.3 Regression-based methods for network identification

Nonlinear regression involves techniques such as polynomial regression, Spline regression, Gauss-Newton and other iterative numerical techniques. Data clustering methods alone are insufficient to determine the kinetic parameters required in such models, and more sophisticated mathematical tools are needed [182]. The effectiveness of partial least squares regression was shown by Pihur et al. [183]. The combination of BN and nonlinear regression is also promising for network reconstruction.

Gardner et al. [184] developed the Network Identification by multiple Regression (NIR) algorithm; this algorithm uses steady state RNA measurements from transcriptional perturbation experiments. The NIR was shown to be effective for small scale microbial gene networks but has the disadvantage that it requires network prior knowledge. The algorithm was tested on simulated and real data for nine genes in *E. coli* and about 50% of the network edges were correctly recovered. The nine gene network was part of the *Son of Sevenless* (SOS) system in the larger *E. coli* network. A different algorithm (Time-Series Network Identification, TSNI) yielded a similar result as the NIR [165].

Time course AutoRegression models coupled with the Granger causality in genetic network reconstruction have been used for network reconstruction. An extension of this model is the AutoRegressive Integrated Moving Average (ARIMA). The Granger causality measure enables determination of causal relation between two signals. It also determines direct or indirect causality [185, 186, 187, 188, 189]. However, little has been done to study biological networks using AR-eXogenous models (ARX) and the more extended ARIMA models. It was shown that successful network inference can be achieved by the use of Granger causality [190] and partial correlation analysis based methods e.g. [191, 192, 193]. These methods do not infer causality between nodes. Another approach is to use Relevance networks - which is an association based inference method [194].

A.4 The Boolean Network formalism

The use of discrete models in biology dates back a long time [29, 67]. In discrete mathematics and computer science, discrete time models are viewed as computing machines [195]. The first computational methods for genetic network inference were the Boolean and random Boolean network approaches [196, 197, 198, 199]. Boolean functions map the state variables at time t to $t + 1$ and performs best for a small number of genes. Boolean networks work on the

A.5. The Ordinary Differential Equation (ODE) formalism

assumption that transcript production and mRNA degradation are controlled by switch-like processes. This approach uses discretized data; hence, at a risk of information loss. However, according to Rocke and Durbin [200] the use of Boolean formalism should be treated with caution given the relatively large noise levels in microarray data.

Huang et al. [201] considered the binary approximation of transcription an over-simplification. Many biological phenomena are portrayed as continuous, never-the-less numerous studies have demonstrated that binary (and ternary) discretization or up-regulated and down-regulated occasionally yields reliable results. Huang and co-workers also showed with the use of a binary approximation that the Boolean formalism has the ability to yield biologically meaningful results. Using ternary and higher order discretization levels may reveal more characteristics of a biological network but such discretization levels are computationally costly [202].

Liang et al. [27] proposed the REVerse Engineering ALgorithm (REVEAL) to construct large-scale Boolean networks from biological data. Using high through-put microarray data, Boolean networks are capable of capturing the dynamic behavior in complex systems. An even more powerful extension of the Boolean network is the Probabilistic Boolean Network approach. The random Boolean network can also realistically capture the essential characteristics of genetic regulatory networks [29, 159, 201, 203]. According to Shmulevich and Zhang [204], this justifies using the Boolean formalism for network inferences. Boolean networks only allow for qualitative rather than quantitative inferences. Steggles et al. [205] showed that Boolean networks fail to capture important network dynamics. It is also inconsistent in its qualitative dynamic behavior [206, 207].

After a careful scrutiny of the methods in subsections A.2 to A.4, none of them was best suited for modeling the XlnR regulon. The preferred formalism is the ordinary differential equations (overview in subsection A.5).

A.5 The Ordinary Differential Equation (ODE) formalism

A.5.1 Variants of the ODE formalism

ODEs are efficient for modeling small-sized networks [208, 209], but face the problem of computational time complexity for large dimensional networks [167, 210]. De Hoon et al. [211] illustrated the efficiency of ODEs with real data from *B. Subtilis*. By using ODEs network dynamics can be studied prior to identification [212] - usually through *in silico* studies and using synthetic data.

One of the popular differential equations formalism for network inference is S-systems (see [213, 214, 215, 216, 217, 218, 219]). Structurally, S-systems are rich enough to capture relevant biological dynamics [220]. They are advantageous in terms of system analysis and control design since they allow the customizing

of analytical and computational methods. Steady-state evaluation, control analysis, and sensitivity analysis of a given system can be established mathematically using S-system parameters [221, 222]. However, S-systems have a major disadvantage that all of its large number of parameters must be estimated. This number is $2\kappa(\kappa+1)$, κ being the number of state variables. Another bottleneck is the complicated parameter estimation, and model fitting to experimental data.

Kabir et al. [223] used Linear Time Invariant models to infer biological network structures and parameter estimation. Using synthetic data, Kabir and coworkers demonstrated the potential of *Self-Adaptive* Differential Evolutionary algorithms in biological network identification. Zhan and Yeung [154] proposed a method that combines spline theory with Linear Programming and Nonlinear Programming. They used enzyme kinetics models to describe the network dynamics and study systems parameter sensitivity.

A.5.2 Mathematical representation

A popular representation for modeling biological networks is the transcription-translation model:

$$\begin{cases} \dot{x}_i = \varphi(f_i(z_1; k_{i1}, h_1), \dots, f_i(z_m; k_{im}, h_m)) - k_{id}x_i, \\ \dot{z}_i = \psi(x_i; r_i) - \eta_i z_i, \quad \text{given } x_i(0), z_i(0) \end{cases} \quad (\text{A.1})$$

The nonnegative constants k_{id} and η_i represent the mRNA and protein degradation parameters, respectively; the terms h_l where $l = 1, \dots, i, \dots, m$ are Hill coefficients, m - number of transcription factors for gene i , and n - number of genes. The vector-valued functions $f_i \in \{f_i^-, f_i^+\} : \mathfrak{R} \rightarrow \mathfrak{R}$ describe the gene regulation in time. f_i^- and f_i^+ are repressing and activating Hill functions, respectively. These functions describe the dependence of the mRNA concentration on the protein concentration z_l . The mRNA synthesis function φ consists of sums or products of f_i .

The use of the formulation (A.1) is illustrated for instance in Chapter 2. Often the translation function ψ of mRNA x_i to protein z_i is linear. When the protein z_l has no effect on the mRNA levels x_l , then the corresponding term in the model is set to zero, i.e. $f_i = 0$. The most common forms of f_i are Hill functions and Michaelis-Menten functions [50, 224, 225]. The parameters $[k_{i1}, \dots, r_i, \eta_i]^T$ are often estimated using the *Maximum Likelihood* approach [226, 227, 228]. The article by Polynikis et al. [56] details the applications of differential equations in the study of biological networks.

Modeling cycle, abbreviations, acronyms and definitions

B.1 The modeling cycle in network reconstruction

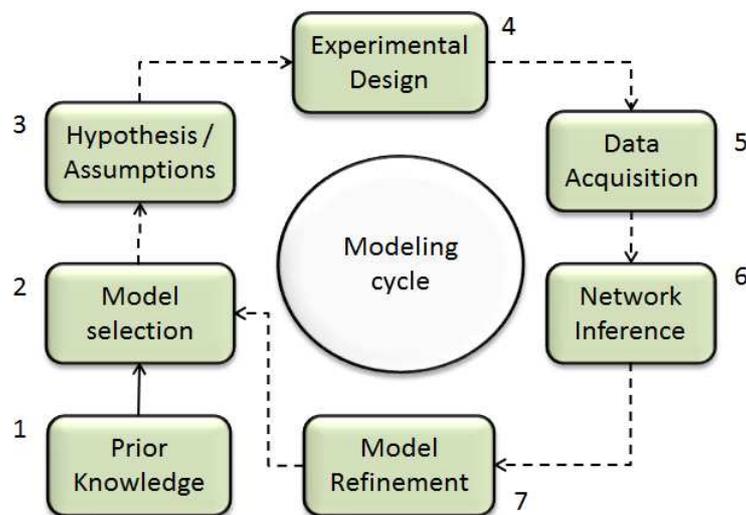


Figure B.1: **The modeling cycle in network reconstruction.** The conventional network inference starts from step 1 and evolves through step 7. The starting points for this cycle may vary depending on the situation at hand. Good handling of each step determines to a great extent the “quality” of the inferred network structure, identified parameters and associated deductions.

To stress the challenge of network reconstruction, d’Alché-Buc and Schachter [229] stated that: “As this field of research evolves, it is becoming clear that there is no one-size-fits-all solution, but rather a range of frameworks and methods, each with its specific trade-off between abstraction and tractability, the ultimate test being the ability to answer relevant biological questions.” The modeling cycle depicted in Figure B.1 is an illustration of the above statement. We have seen that reverse engineering involves numerous steps, ranging from

Chapter B. Modeling cycle, abbreviations, acronyms and definitions

gathering information, planning and designing the experiment, and data collection to the actual network identification.

The Prior knowledge at step 1 comes from literature, databases, and information from experts. Once enough information has been gathered, it is important to start considering which formalism or analytic techniques to use for modeling the network (step 2). Once a model formalism is chosen, it is vital to clearly state the research hypothesis to be tested (step 3). This is closely followed by the experimental design process (step 4). Getting the experimental design right ensures that high quality data is collected (step 5). The network inference at step 6 is mathematically involving and computationally expensive. The identification results from step 6 are then matched to expectations in the model refinement at step 7. Between the seven steps, there are other small significant steps which are important for the overall network reconstruction.

B.2 List of abbreviations and acronyms

Symbol	Description
A	interaction (connectivity) matrix.
i	gene (transcript) index.
x_i	expression intensity (state variable).
\dot{x}_i	transcription rate.
y_i	measured mRNA level.
u	input signal vector.
t and t_0	continuous time and initial time point, respectively.
J	Jacobian matrix.
\mathcal{J}	Objective (goal) function.
n	number of genes.
N	maximum discrete time index.
X	matrix of time series data.
\Re and \mathbb{R}	Real line and matrix space, respectively.
p	statistical significance value.
Ω	parameter space.
τ	infinitesimal time element.
\mathbb{E}	expectation operator.
Λ_i	diagonal matrix of variances.
λ_i	eigenvalue.
\mathbb{C}	complex number space.
η_i	protein degradation parameter.
δ	small infinitesimal change.
$\ \cdot\ _1$	Manhattan or city block distance.
$\ \cdot\ _2$	Fröbenius or Euclidian norm.
a.u	arbitrary unit.
qPCR	quantitative Polymerase Chain Reaction.

B.3 Some useful definitions

Definition B.3.1. Gene regulatory network. A gene regulatory network is a graphical or visual representation of DNA segments in a cell [230].

Definition B.3.2. Static network structure. A regulatory network is static if it has fixed point-to-point connections between nodes, which are independent of time. Unlike static networks, dynamic networks are time-dependent.

Definition B.3.3. Feedback loop. A feedback loop refers to a cyclic chain of links in a regulatory network [44].

Definition B.3.4. Regulon. A regulon is a group of genes regulated by the same regulatory molecule. The genes in a regulon share a common regulatory element binding site or promoter.

Bibliography

- [1] A. J. Levine, W. Hu, and Z. Feng, "The p53 pathway: what questions remain to be explored?," *Cell Death Differ.*, vol. 13, no. 6, pp. 1027–1036, 2006.
- [2] E. Batchelor, A. Loewer, and G. Lahav, "The ups and downs of p53: understanding protein dynamics in single cells," *Nat. Rev. Cancer*, vol. 9, no. 5, pp. 371–377, 2009.
- [3] S. Volinia, M. Galasso, S. Costinean, L. Tagliavini, G. Gamberoni, A. Drusco, J. Marchesini, N. Mascellani, M. E. Sana, R. A. Jarour, C. Despots, M. Teitell, R. Baffa, R. Aqeilan, M. V. Iorio, C. Taccioli, R. Garzon, G. D. Leva, M. Fabbrì, M. Catozzi, M. Previati, S. Ambs, T. Palumbo, M. Garofalo, A. Veronese, A. Bottoni, P. Gasparini, C. C. Harris, R. Visone, Y. Pekarsky, A. de la Chapelle, M. Bloomston, M. Dillhoff, L. Z. Rassenti, T. J. Kipps, K. Huebner, F. Pichiorri, D. Lenze, S. Cairo, M.-A. Buendia, P. Pineau, A. Dejean, N. Zanesi, S. Rossi, G. A. Calin, C.-G. Liu, J. Palatini, M. Negrini, A. Vecchione, A. Rosenberg, and C. M. Croce, "Reprogramming of miRNA networks in cancer and leukemia," *Genome Research*, vol. 20, no. 5, pp. 589–599, 2010.
- [4] M. D'Antonio, V. Pendino, S. Sinha, and F. D. Ciccarelli, "Network of Cancer Genes (NCG 3.0): integration and analysis of genetic and network properties of cancer genes," *Nucleic Acids Res.*, vol. 40, no. 12, pp. 1–6, 2011.
- [5] S. Ma, Q. Gong, and H. J. Bohnert, "An *Arabidopsis* gene network based on the graphical Gaussian model," *Genome Res.*, vol. 17, no. 11, pp. 1614–1625, 2007.
- [6] J. J. B. Keurentjes, J. Fu, I. R. Terpstra, J. M. Garcia, G. van den Ackerveken, L. B. Snoek, A. J. M. Peeters, D. Vreugdenhil, M. Koornneef, and R. C. Jansen, "Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci," *PNAS*, vol. 108, no. 20, pp. 1708–1713, 2007.
- [7] L. Mao, J. L. van Hemert, S. Dash, and J. A. Dickerson, "*Arabidopsis* gene co-expression network and its functional modules," *BMC Bioinformatic*, vol. 10, p. 346, 2009.
- [8] I. A. Maraziotis, A. Dragomir, and A. Bezerianos, "Gene networks reconstruction and time-series prediction from microarray data using recurrent neural fuzzy networks," *IET Syst. Biol.*, vol. 1, no. 1, pp. 41–50, 2007.
- [9] H. Zare, D. Sangurdekar, P. Srivastava, M. Kaveh, and A. Khodursky, "Reconstruction of *Escherichia coli* transcriptional regulatory networks via regulon-based associations," *BMC Syst. Biol.*, vol. 3, p. 39, 2009.

Bibliography

- [10] K. Tsuda, M. Sato, T. Stoddard, J. Glazebrook, and F. Katagiri, "Network Properties of Robust Immunity in Plants," *PLoS Genetics*, vol. 5, no. 12, p. e1000772, 2009.
- [11] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood, "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network," *Science*, vol. 292, no. 5518, pp. 929–934, 2001.
- [12] A. L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nat. Rev. Genet.*, vol. 5, no. 2, pp. 101–113, 2004.
- [13] R. L. Farrell and P. S. Skerker, "Chlorine-free bleaching with catalase treatment," In: J. Visser, G. Beldman, M.A. Kusters van Someren, and A.G.J. Voragen (ed.) *Xylans and xylanases*, Elsevier, Amsterdam, pp. 315–324, 1992.
- [14] C. Grassin and P. Fauquemberque, "Applications of pectinases in beverages," In: J. Visser and A.G.J. Voragen, (eds.), *Pectins and pectinases, Progress in Biotechnology 14*, Elsevier, Amsterdam, pp. 453–462, 1996.
- [15] J. Maat, M. Roza, J. Verbakel, H. Stam, M. J. Santos da Silva, M. Bosse, M. R. Egmond, M. L. D. Hagemans, R. F. M. Gorcom, J. G. M. Hessing, C. A. M. J. J. van den Hondel, and C. van Rotterdam, "Xylanases and Their Applications in Bakery," In: *Xylan and Xylanases, Progress in Biotechnology No. 7*, J. Visser, M.A. Kusters van Someren, G. Beldman and A.G.J. Voragen, (Eds.), Elsevier, Amsterdam, pp. 349–360, 1992.
- [16] A. M. Nissen, L. Anker, N. Munk, and N. K. Lange, "Xylanases for the pulp and paper industry," In: *Xylans and xylanases (Eds: J. Visser, G. Beldman, M.A. Kusters van Someren and A.G.J. Voragen)*, Elsevier, Amsterdam., pp. 325–338.
- [17] J. C. Duarte, M. Costa-Ferreira, and G. Sena-Martins, "Cellobiose dehydrogenase. Possible roles and importance for pulp and paper biotechnology," *Biore-source Technol.*, vol. 68, no. 1, pp. 43–48, 1999.
- [18] A. R. Stricker, R. L. Mach, and L. H. de Graaff, "Regulation of transcription of cellulases- and hemicellulases- encoding genes in *Aspergillus niger* and *Hypocrea jecorina* (*Trichoderma reesei*)," *Appl. Microbiol. Biotechnol.*, vol. 78, no. 2, pp. 211–220, 2008.
- [19] R. P. de Vries, J. Visser, and L. H. de Graaff, "CreA modulates the XlnR-induced expression on xylose of *Aspergillus niger* genes involved in xylan degradation," *Res. Microbiol.*, vol. 150, no. 4, pp. 281–285, 1999.
- [20] W. Prathumpai, M. McIntyre, and J. Nielsen, "The effect of CreA in the glucose and xylose catabolism in *Aspergillus nidulans*," *Appl. Microbiol. Biotechnol.*, vol. 63, no. 6, pp. 748–753, 2004.
- [21] J. Strauss, H. K. Horvath, B. M. Abdallah, J. Kindermann, R. L. Mach, and C. P. Kubicek, "The function of CreA, the carbon catabolite repressor of *Aspergillus nidulans*, is regulated at the transcriptional and post-transcriptional level," *Mol. Microbiol.*, vol. 32, no. 1, pp. 169–178, 1999.
- [22] M. R. Drysdale, S. E. Kolze, and J. M. Kelly, "The *Aspergillus niger* carbon catabolite repressor encoding gene, *creA*," *Gene*, vol. 130, no. 2, pp. 241–5, 1993.
- [23] P. Erdős and A. Rényi, "On random graphs," *Publicationes Mathematicae*, vol. 6, no. 1959, pp. 290–297, 1959.

- [24] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo, "How to infer gene networks from expression profiles," *Mol. Syst. Biol.*, vol. 3, p. 78, 2007.
- [25] E. J. Chikofsky and J. H. Cross II, "Reverse engineering and design recovery: A taxonomy," *IEEE Software*, vol. 7, no. 1, pp. 13–17, 1990.
- [26] T. Chen, H. L. He, and G. M. Church, "Modeling gene expression with differential equations," in *Proc. Pac. Symp. Biocomput.*, vol. 4, pp. 29–40, 1999.
- [27] S. Liang, S. Fuhrman, and R. Somogyi, "REVEAL, A General Reverse Engineering Algorithm For Inference Of Genetic Network Architectures," in *Pacific Symposium on Biocomputing*. World Scientific Publishing Co., vol. 3, pp. 18–29, 1998.
- [28] P. D'haeseleer, X. Wen, S. Fuhrman, and R. Somogyi, "Linear modeling of mRNA expression levels during CNS development and injury," in *Proc. Pac. Symp. Biocomput.*, pp. 41–52, 1999.
- [29] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *J. Theor. Biol.*, vol. 22, no. 3, pp. 437–467, 1969.
- [30] R. Breitling, "What is systems biology?," *Frontiers in Physiology*, vol. 1, p. 9, 2010.
- [31] H. Kitano, "Systems Biology: Toward System-level Understanding of Biological Systems," *MIT Press*, pp. 1–29, 2001.
- [32] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitzky, "Revealing strengths and weaknesses of methods for gene network inference," *PNAS*, vol. 107, no. 14, pp. 6286–6291, 2010.
- [33] H. de Jong, "Modeling and simulation of genetic regulatory systems: a literature review," *J. Comput. Biol.*, vol. 9, no. 1, pp. 67–103, 2002.
- [34] N. N. M. E. van Peij, M. M. C. Gielkens, R. P. de Vries, J. Visser, and L. H. de Graaff, "The Transcriptional Activator XlnR Regulates Both Xylanolytic and Endoglucanase Gene Expression in *Aspergillus niger*," *Appl. Environ. Microbiol.*, vol. 64, no. 10, pp. 3615–3619, 1998.
- [35] D. van der Veen, J. M. Oliveira, W. A. M. van den Berg, and L. H. de Graaff, "Variance components analysis reveals contribution of sample processing to transcript variation," *Appl. Environ. Microbiol.*, vol. 75, no. 8, pp. 2414–2422, 2009.
- [36] E. Sakamoto and H. Iba, "Inferring a System of Differential Equations for a Gene Regulatory Network by using Genetic Programming," in *Proc. Congress on Evolutionary Computation*, pp. 720–726, 2001.
- [37] J. J. Tyson and H. G. Othmer, "The dynamics of feedback control circuits in biochemical pathways," *Prog. Theor. Biol.*, vol. 5, pp. 1–62, 1978.
- [38] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. Benjamin Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young, "Transcriptional regulatory networks in *Saccharomyces cerevisiae*," *Science*, vol. 298, no. 5594, pp. 799–804, 2002.
- [39] A. Becskei and L. Serrano, "Engineering stability in gene networks by autoregulation," *Nature*, vol. 405, no. 6786, pp. 590–593, 2000.

Bibliography

- [40] T. B. Kepler and T. C. Elston, "Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations," *Biophys. J.*, vol. 81, no. 6, pp. 3116–3136, 2001.
- [41] M. L. Simpson, C. D. Cox, and G. S. Saylor, "Frequency domain analysis of noise in autoregulated gene circuits," *PNAS*, vol. 100, no. 8, pp. 4551–4556, 2003.
- [42] Y. Tao, "Intrinsic and external noise in an auto-regulatory genetic network," *J. Theor. Biol.*, vol. 229, no. 2, pp. 147–156, 2004.
- [43] J. Goutsias and S. Kim, "Stochastic Transcriptional Regulatory Systems with Time Delays: A Mean-Field Approximation," *Phys. Biol.*, vol. 13, no. 5, pp. 1049–1076, 2006.
- [44] K. Sneppen, S. Krishna, and S. Semsey, "Simplified Models of Biological Networks," *Annu. Rev. Biophys.*, vol. 39, pp. 43–59, 2010.
- [45] H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada, K. Yoshikawa, and R. Kageyama, "Oscillatory expression of the bHLH factor *hes1* regulated by a negative feedback loop," *Science*, vol. 298, no. 5594, pp. 840–843, 2002.
- [46] M. H. Jensen, G. Tiana, and K. Sneppen, "Sustained oscillations and time delays in gene expression of protein *Hes1*," *FEBS Lett.*, vol. 541, no. 1-3, pp. 176–177, 2003.
- [47] G. Tiana, M. H. Jensen, and K. Sneppen, "Time delay as a key to apoptosis induction in the p53 network," *Eur. Phys. J.*, vol. 29, no. 1, pp. 135–140, 2002.
- [48] G. Tiana, S. Krishna, S. Pigolotti, M. H. Jensen, and K. Sneppen, "Oscillations and temporal signalling in cells," *Phys. Biol.*, vol. 4, no. 2, pp. R1–R17, 2007.
- [49] R. D. Bliss, P. R. Painter, and A. G. Marr, "Role of feedback inhibition in stabilizing the classical operon," *J. Theor. Biol.*, vol. 97, no. 2, pp. 177–193, 1982.
- [50] P. Mendes, W. Sha, and K. Ye, "Artificial gene networks for objective comparison of analysis algorithms," *Bioinformatics*, vol. 19, no. Suppl. 2, pp. 122–129, 2003.
- [51] F. J. Isaacs, J. Hasty, C. R. Cantor, and J. J. Collins, "Prediction and measurement of an autoregulatory genetic module," *PNAS*, vol. 100, no. 13, pp. 7714–7719, 2003.
- [52] A. Sayyed-Ahmad, K. Tuncay, and P. J. Ortoleva, "Transcriptional regulatory network refinement and quantification through kinetic modeling, gene expression microarray data and information theory," *BMC Bioinformatics*, vol. 8, p. 20, 2007.
- [53] S. Agrawal, C. Archer, and D. V. Schaffer, "Computational Models of the Notch Network Elucidate Mechanisms of Context-dependent Signaling," *PLoS Comput. Biol.*, vol. 5, p. e1000390, 2009.
- [54] G. Yagil and E. Yagil, "On the relation between effector concentration and the rate of induced enzyme synthesis," *Biophys. J.*, vol. 11, no. 1, pp. 11–27, 1971.
- [55] A. V. Hill, "The possible effect of the aggregation of the molecules of haemoglobin on its dissociation curves," *J. Physiol.*, vol. 40, no. 4, pp. 4–7, 1910.
- [56] A. Polynikis, S. J. Hogan, and M. di Bernardo, "Comparing different ODE modeling approaches of gene regulatory networks," *J. Theor. Biol.*, vol. 261, no. 4, pp. 511–530, 2009.

- [57] L. Hasper, *Function and mode of regulation of the transcriptional activator XlnR from Aspergillus*. PhD thesis, Wageningen University, Microbiology Department, 2003.
- [58] J. M. Mahaffy, "Genetic control models with diffusion and delays," *Math. Biosci.*, vol. 90, pp. 519–533, 1988.
- [59] J. M. Mahaffy, D. A. Jorgensen, and R. L. Vanderheyden, "Oscillations in a model of repression with external control," *J. Math. Biol.*, vol. 30, no. 7, pp. 669–691, 1992.
- [60] W. T. Mocek, R. Rudbicki, and E. O. Voit, "Approximation of delays in biochemical systems," *Math. Biosci.*, vol. 198, no. 2, pp. 190–216, 2005.
- [61] A. Verdugo and R. Rand, "Hopf bifurcation in a DDE model of gene expression," *Science Direct*, vol. 13, no. 2, pp. 235–242, 2008.
- [62] N. Aro, T. Pakula, and M. Penttilä, "Transcriptional regulation of plant cell wall degradation by filamentous fungi," *FEMS Microbiol. Rev.*, vol. 29, no. 4, pp. 719–739, 2005.
- [63] A. A. Hasper, E. Dekkers, M. van Mil, P. J. I. van de Vondervoort, and L. H. de Graaff, "EglC, A New Endoglucanase from *Aspergillus niger* with major activity towards Xyloglucan," *Appl. Environ. Microbiol.*, vol. 68, no. 4, pp. 1556–1560, 2002.
- [64] I. M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, and P. D. Karp, "EcoCyc: a comprehensive database resource for *Escherichia coli*," *Nucleic Acids Res.*, vol. 33, pp. D334–337, 2005.
- [65] R. Maithreye, R. R. Sarkar, V. K. Parnaik, and S. Sinha, "Delay-Induced Transient Increase and Heterogeneity in Gene Expression in Negatively Auto-Regulated Gene Circuits," *PLoS One*, vol. 3, p. e2972, 2008.
- [66] R. Thomas, D. Thieffry, and M. Kauffman, "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.*, vol. 57, pp. 247–276, 1995.
- [67] R. Thomas and R. d'Ari, "Biological feedback," CRC, Press, Boca Raton, 1990.
- [68] H. Liu, J.-A. Lu, J. Lü, and D. J. Hill, "Structure identification of uncertain general complex dynamical networks with time delay," *Automatica*, vol. 45, no. 8, pp. 1799–1807, 2009.
- [69] W. Yu, J. Lu, G. Chen, Z. Duan, and Q. Zhou, "Estimating Uncertain Delayed Genetic Regulatory Networks: An Adaptive Filtering Approach," *IEEE Trans. Automat. Contr.*, vol. 54, pp. 792–897, 2009.
- [70] E. Balsa-Canto, A. A. Alonso, and J. R. Banga, "Computational procedures for optimal experimental design in biological systems," *IET Syst. Biol.*, vol. 2, no. 4, pp. 163–172, 2008.
- [71] X. J. Feng and H. Rabitz, "Optimal identification of biochemical reaction networks," *Biophys. J.*, vol. 86, no. 3, pp. 1270–1281, 2004.
- [72] H. Kitano, "Systems biology: a brief overview," *Science*, vol. 295, no. 5560, pp. 1662–1664, 2002.

Bibliography

- [73] Z. Kutalik, K. H. Cho, and O. Wolkenhauer, "Optimal sampling time selection for parameter estimation in dynamic pathway modelling," *BioSystems*, vol. 75, no. 1-3, pp. 43–55, 2004.
- [74] P. F. O. Lindner and B. Hitzmann, "Experimental design for optimal parameter estimation of an enzyme kinetic process based on the analysis of the Fisher information matrix," *J. Theor. Biol.*, vol. 238, no. 1, pp. 111–123, 2006.
- [75] J. Omony, L. H. de Graaff, G. van Straten, and A. J. B. van Boxtel, "Modeling and analysis of the dynamics behavior of the XlnR regulon in *Aspergillus niger*," *BMC Syst. Biol.*, vol. 5, no. Suppl 1, p. S14, 2011.
- [76] P. Vanrolleghem, M. van Daele, P. van Overschee, and G. Vansteenkiste, "Model structure characterization of nonlinear water treatment systems," in *Proc. 10th IFAC Conf. Syst. Ident.*, vol. 1, pp. 279–284, 1994.
- [77] C. Michalik, M. Stuckert, and W. Marquardt, "Optimal Experimental Design for Discriminating Numerous Model Candidates: The AWDC Criterion," *Ind. Eng. Chem. Res.*, vol. 49, no. 2, pp. 913–919, 2009.
- [78] M. Baltes, R. Schneider, C. Sturm, and M. Reus, "Optimal Experimental Design for Parameter Estimation in Unstructured Growth Models," *Biotechnol.*, vol. 10, no. 5, pp. 480–488, 1994.
- [79] K. J. Keesman and J. D. Stigter, "Optimal parametric sensitivity control for the estimation of kinetic parameters in bioreactors," *Math. Biosci.*, vol. 179, no. 1, pp. 95–111, 2002.
- [80] J. D. Stigter and K. J. Keesman, "Optimal parametric sensitivity control of a fed-batch reactor," *Automatica*, vol. 40, no. 8, pp. 1459–1464, 2004.
- [81] D. Faller, U. Klingmüller, and J. Timmer, "Simulation methods for optimal experimental design in systems biology," *Simulation-Transactions of the Society for Modeling and Simulation International*, vol. 79, no. 12, pp. 717–725, 2003.
- [82] K. Selvarajoo and M. Tsuchiya, "Systematic Determination of Biological Network Topology: Nonintegral Connectivity Method (NICM)," *Introduction to systems biology*, vol. Part IV, pp. 449–471, 2007.
- [83] D. Clausznitzer, O. Oleksiuk, L. Løvdok, V. Sourjik, and R. G. Endres, "Chemotactic response and adaptation dynamics in *Escherichia coli*," *PLoS Comput. Biol.*, vol. 6, p. e1000784, 2010.
- [84] C. Kreutz and J. Timmer, "Systems biology: experimental design," *FEBS J.*, vol. 276, no. 4, pp. 923–942, 2009.
- [85] A. R. Mach-Aigner, J. Omony, B. Jovanovic, A. J. B. van Boxtel, and L. H. de Graaff, "D-xylose concentration-dependent Hydrolase Expression Profiles and the according role of CreA and XlnR in *A. niger*," *Appl. Environ. Microbiol.*, vol. 78, no. 9, pp. 3145–3155, 2012.
- [86] Z. Zi, K. H. Cho, M. H. Sung, X. Xia, J. Zheng, and Z. Sun, "In silico identification of the key components and steps in IFN- γ induced JAK-STAT signaling pathway," *FEBS Lett.*, vol. 579, no. 5, pp. 1101–1108, 2005.
- [87] Y. Zhang and A. Rundell, "Comparative study of parameter sensitivity analyses of the TCR-activated Erk-MAPK signalling pathway," *IEE P. Syst. Biol.*, vol. 153, no. 4, pp. 201–211, 2006.

- [88] H. Yue, M. Brown, J. Knowles, H. Wang, D. S. Broomhead, and D. B. Kell, "Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NF- κ B signalling pathway," *Mol. Biosyst.*, vol. 2, no. 12, pp. 640–649, 2006.
- [89] R. Gunawan, Y. Cao, L. Petzold, and F. J. Doyle III, "Sensitivity Analysis of Discrete Stochastic Systems," *Biophys. J.*, vol. 88, no. 4, pp. 2530–2540, 2005.
- [90] K. Godfrey and J. di Stefano III, "Identifiability of model parameters," *In Identification and System Parameter Estimation*, vol. 174, no. 2, pp. 89–114, 1985.
- [91] K. Bernaerts, K. P. M. Gysemans, T. N. Minh, and J. F. van Impe, "Optimal experiment design for cardinal values estimation: guidelines for data collection," *Int. J. Food Microbiol.*, vol. 100, no. 1-3, pp. 153–165, 2005.
- [92] P. A. Vanrolleghem, M. van Daele, and D. Dochain, "Practical identifiability of a biokinetic model of activated sludge respiration," *Wat. Res.*, vol. 29, no. 11, pp. 2561–2570, 1995.
- [93] L. Pronzato, "Optimal experimental design and some related control problems," *IFAC: Automatica*, vol. 44, no. 2, pp. 303–325, 2008.
- [94] G. Goodwin, "Identification: Experiment design," *Systems and Control Encyclopedia*, vol. 4, pp. 2257–2264, 1987.
- [95] M. Bentele, I. Lavrik, M. Ulrich, S. Stöber, D. W. Heermann, H. Kalthoff, P. H. Kramer, and R. Eils, "Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis," *J. Cell Biol.*, vol. 166, no. 6, pp. 839–851, 2004.
- [96] F. Geier, J. Timmer, and C. Fleck, "Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge," *BMC Syst. Biol.*, vol. 1, no. Suppl 1, p. 11, 2007.
- [97] M. K. S. Yeung, J. Tegnér, and J. J. Collins, "Reverse engineering gene networks using singular value decomposition and robust regression," *PNAS*, vol. 99, no. 9, pp. 6163–6168, 2002.
- [98] H. J. D. Bussink, J. P. T. W. van den Hombergh, P. R. L. A. van den Ijssel, and J. Visser, "Characterization of polygalacturonase-overproducing *Aspergillus niger* transformants," *Appl. Microbiol. Biotechnol.*, vol. 37, no. 3, pp. 324–329, 1992.
- [99] L. Delgado, B. A. Trejo, C. Huitrón, and G. Aguilar, "Pectin lyase from *Aspergillus* sp. CH-Y-1043," *Appl. Microbiol. Biotechnol.*, vol. 39, no. 1992, pp. 515–519, 1992.
- [100] L. Parenicová, J. A. Benen, H. C. Kester, and J. Visser, "*pgaA* and *pgaB* encode two constitutively expressed endopolygalacturonases of *Aspergillus niger*," *Biochem. J.*, vol. 345 Pt, no. 3, pp. 637–644, 2000.
- [101] M. M. C. Gielkens, J. Visser, and L. H. de Graaff, "Arabinoxylan degradation by fungi: characterization of the arabinoxylan-arabinofuranohydrolase encoding genes from *Aspergillus niger* and *Aspergillus tubingensis*," *Curr. Genet.*, vol. 31, no. 1, pp. 22–29, 1997.
- [102] N. N. M. E. van Peij, J. Brinkmann, M. Vrsanská, J. Visser, and L. H. de Graaff, " β -xylosidase activity, encoded by *xlnD*, is essential for complete hydrolysis of xylan by *Aspergillus niger* but not for induction of the xylanolytic enzyme spectrum," *Eur. J. Biochem.*, vol. 245, no. 1, pp. 164–173, 1997.

Bibliography

- [103] M. R. Andersen, W. Vongsangnak, G. Panagiotou, M. P. Salazar, L. Lehmann, and J. Nielsen, "A trispecies *Aspergillus* microarray: comparative transcriptomics of three *Aspergillus* species," *PNAS, USA*, vol. 105, no. 11, pp. 4387–4392, 2008.
- [104] M. J. L. de Groot, C. van den Dool, H. A. B. Wösten, M. Levisson, P. A. vanKuyk, G. J. G. Ruijter, and R. P. de Vries, "Regulation of Pentose Catabolic Pathway Genes of *Aspergillus niger*," *Food Technol. Biotechnol.*, vol. 45, no. 2, pp. 134–138, 2007.
- [105] R. P. de Vries and J. Visser, "Regulation of the Feruloyl Esterase (*faeA*) Gene from *Aspergillus niger*," *Appl. Environ. Microbiol.*, vol. 65, no. 12, pp. 5500–5503, 1999.
- [106] R. P. de Vries and J. Visser, "*Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides," *Microbiol. Mol. Biol. Rev.*, vol. 65, no. 4, pp. 497–522, 2001.
- [107] M. M. C. Gielkens, E. Dekkers, J. Visser, and L. H. de Graaff, "Two cellobiohydrolase-encoding genes from *Aspergillus niger* require D-xylose and the xylanolytic transcriptional activator XlnR for their expression," *Appl. Environ. Microbiol.*, vol. 65, no. 10, pp. 4340–4345, 1999.
- [108] A. A. Hasper, J. Visser, and L. H. de Graaff, "The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D-xylose reductase gene expression," *Mol. Microbiol.*, vol. 36, no. 1, pp. 193–200, 2000.
- [109] N. N. M. E. van Peij, J. Visser, and L. H. de Graaff, "Isolation and analysis of *xlnR* encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*," *Mol. Microbiol.*, vol. 27, no. 1, pp. 131–142, 1998.
- [110] R. D. Finn, J. Mistry, B. Schuster-Böckler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. Sonnhammer, and A. Bateman, "Pfam: clans, web tools and services," *Nucleic Acids Res.*, vol. 34, no. Suppl. 1, pp. D247–51, 2006.
- [111] R. P. de Vries, H. C. van den Broeck, E. Dekkers, P. Manzanares, L. H. de Graaff, and J. Visser, "Differential expression of three α -galactosidase genes and a single β -galactosidase gene from *Aspergillus niger*," *Appl. Environ. Microbiol.*, vol. 65, no. 6, pp. 2453–2460, 1999.
- [112] P. A. vanKuyk, M. J. de Groot, G. J. Ruijter, R. P. de Vries, and J. Visser, "The *Aspergillus niger* D-xylulose kinase gene is co-expressed with genes encoding arabinan degrading enzymes, and is essential for growth on D-xylose and L-arabinose," *Eur. J. Biochem.*, vol. 268, no. 20, pp. 5414–5423, 2001.
- [113] H. N. Arst Jr and C. R. Bailey, "The regulation of carbon metabolism in *Aspergillus nidulans*," *Academic Press, London/New York/San Francisco*, 1977.
- [114] G. J. G. Ruijter, S. A. Vanhanen, M. M. C. Gielkens, P. J. I. van de Vondervoort, and J. Visser, "Isolation of *Aspergillus niger creA* mutants and effects of the mutations on expression of arabinases and L-arabinose catabolic enzymes," *Microbiology*, vol. 143, no. Pt 9, pp. 2991–2998, 1997.
- [115] C. Scazzocchio, V. Gavrias, B. Cubero, C. Panozzo, M. Mathieu, and B. Felenbok, "Carbon catabolite repression in *Aspergillus nidulans*: a review," *Can. J. Bot.*, vol. 73, no. Suppl. 1, pp. S160–S166, 1995.

- [116] L. H. de Graaff, H. C. van den Broek, A. J. J. van Ooijen, and J. Visser, "Regulation of the xylanase-encoding *xlnA* gene of *Aspergillus tubingensis*," *Mol. Microbiol.*, vol. 12, no. 3, pp. 479–490, 1994.
- [117] K. Kinoshita, M. Takano, T. Koseki, K. Ito, and K. Iwano, "Cloning of the *xynNB* gene encoding xylanase B from *Aspergillus niger* and its expression in *Aspergillus kawachii*," *J. Ferment. Bioeng.*, vol. 79, no. 5, pp. 422–428, 1995.
- [118] M. J. A. Flipphi, M. van Heuvel, P. van der Veen, J. Visser, and L. H. de Graaff, "Cloning and characterization of the *abfB* gene coding for the major α -L-arabinofuranosidase (ABF B) of *Aspergillus niger*," *Curr. Genet.*, vol. 24, no. 6, pp. 525–532, 1993.
- [119] R. P. de Vries, C. H. Poulsen, S. Madrid, and J. Visser, "*AguA*, the gene encoding an extracellular α -glucuronidase from *Aspergillus tubingensis*, is specifically induced on xylose and not on glucuronic acid," *J. Bacteriol.*, vol. 180, no. 2, pp. 243–249, 1998.
- [120] L. H. de Graaff, J. Visser, H. C. van den Broeck, F. Strozyk, F. J. M. Kormelink, and J. C. Boonman, "Cloning, expression and use of acetyl xylan esterases from fungal origin," *European patent application*, pp. 0507369–A/7, 1992.
- [121] R. P. de Vries, B. Michelsen, C. H. Poulsen, P. A. Kroon, R. H. van den Heuvel, C. B. Faulds, G. Williamson, J. P. van den Hombergh, and J. Visser, "The *faeA* genes from *Aspergillus niger* and *Aspergillus tubingensis* encode ferulic acid esterases involved in degradation of complex cell wall polysaccharides," *Appl. Environ. Microbiol.*, vol. 63, no. 12, pp. 4638–4644, 1997.
- [122] L. Rutten, C. Ribot, B. Trejo-Aguilar, H. A. B. Wösten, and R. P. de Vries, "A single amino acid change (Y318F) in the L-arabitol dehydrogenase (LadA) from *Aspergillus niger* results in a significant increase in affinity for D-sorbitol," *BMC Microbiol.*, vol. 9, p. 166, 2009.
- [123] E. Battaglia, S. F. Hansen, A. Leendertse, S. Madrid, H. Mulder, I. Nikolaev, and R. P. de Vries, "Regulation of pentose utilisation by AraR, but not XlnR, differs in *Aspergillus nidulans* and *Aspergillus niger*," *Appl. Microbiol. Biotechnol.*, vol. 91, no. 2, pp. 387–397, 2011.
- [124] G. Pontecorvo, J. A. Roper, L. M. Hemmons, K. D. Macdonald, and A. W. Bufton, "The genetics of *Aspergillus nidulans*," *Adv. Genet.*, vol. 5, pp. 141–238, 1953.
- [125] G. J. G. Ruijter, M. Bax, H. Patel, S. J. Flitter, P. J. I. van de Vondervoort, R. P. de Vries, P. A. vanKuyk, and J. Visser, "Mannitol is required for stress tolerance in *Aspergillus niger* conidiospores," *Eukaryot. Cell*, vol. 2, no. 4, pp. 690–698, 2003.
- [126] P. A. vanKuyk, J. A. Diderich, A. P. MacCabe, O. Hererro, G. J. Ruijter, and J. Visser, "*Aspergillus niger mstA* encodes a high-affinity sugar/H⁺ symporter which is regulated in response to extracellular pH," *Biochem. J.*, vol. 379, no. Pt, pp. 375–383, 2004.
- [127] W. Vishniac and M. Santer, "The thiobacilli," *Bacteriol. Rev.*, vol. 21, pp. 195–213, 1957.
- [128] J. Huggett, K. Dheda, S. Bustin, and A. Zumla, "Real-time RT-PCR normalisation; strategies and considerations," *Genes Immun.*, vol. 6, no. 4, pp. 279–284, 2005.
- [129] M. W. Pfaffl, "A new mathematical model for relative quantification in real-time RT-PCR," *Nucleic Acids Res.*, vol. 29, no. 9, p. e45, 2001.

Bibliography

- [130] J. A. Pérez-González, N. N. van Peij, A. Bezoen, A. P. MacCabe, D. Ramón, and L. H. de Graaff, "Molecular cloning and transcriptional regulation of the *Aspergillus nidulans xlnD* gene encoding a β -xylosidase," *Appl. Environ. Microbiol.*, vol. 64, no. 4, pp. 1412–1419, 1998.
- [131] J. M. F. de Oliveira, M. W. van Passel, P. J. Schaap, and L. H. de Graaff, "Proteomic Analysis of the Secretory Response of *Aspergillus niger* to D-maltose and D-xylose," *PLoS One*, vol. 6, p. e20865, 2011.
- [132] A. R. Mach-Aigner, M. E. Pucher, and R. L. Mach, "D-xylose as a repressor or inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*)," *Appl. Environ. Microbiol.*, vol. 76, no. 6, pp. 1770–1776, 2010.
- [133] A. R. Stricker, K. Grosstessner-Hain, E. Würleitner, and R. L. Mach, "Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*," *Eukaryot. Cell*, vol. 5, no. 12, pp. 2128–2137, 2006.
- [134] A. R. Mach-Aigner, L. Gudynaite-Savitch, and R. L. Mach, "L-arabitol is the actual inducer of xylanase expression in *Hypocrea jecorina* (*Trichoderma reesei*)," *Appl. Environ. Microbiol.*, vol. 77, no. 7, pp. 5988–5994, 2011.
- [135] A. R. Mach-Aigner, M. E. Pucher, M. G. Steiger, G. E. Bauer, S. J. Preis, and R. L. Mach, "Transcriptional regulation of *xyr1*, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *hypocrea jecorina*," *Appl. Environ. Microbiol.*, vol. 74, no. 21, pp. 6554–6562, 2008.
- [136] Y. Noguchi, H. Tanaka, K. Kanamaru, M. Kato, and T. Kobayashi, "Xylose Triggers Reversible Phosphorylation of XlnR, the Fungal Transcriptional Activator of Xylanolytic and Cellulolytic Genes in *Aspergillus oryzae*," *Biosci. Biotechnol. Biochem.*, vol. 75, no. 5, pp. 953–959, 2011.
- [137] M. Orejas, A. P. MacCabe, J. A. Pérez-González, S. Kumar, and D. Ramón, "Carbon catabolite repression of the *Aspergillus nidulans xlnA* gene," *Mol. Microbiol.*, vol. 31, no. 1, pp. 177–184, 1999.
- [138] M. Orejas, A. P. MacCabe, J. A. Pérez-González, S. Kumar, and D. Ramón, "The wide-domain carbon catabolite repressor CreA indirectly controls expression of the *Aspergillus nidulans xlnB* gene, encoding the acidic endo- β -(1,4)-xylanase x_{24} ," *J. Bacteriol.*, vol. 183, no. 5, pp. 1517–1523, 2001.
- [139] E. N. Tamayo, A. Villanueva, A. A. Hasper, L. H. de Graaff, D. Ramón, and M. Orejas, "CreA mediates repression of the regulatory gene *xlnR* which controls the production of xylanolytic enzymes in *Aspergillus nidulans*," *Fungal Genet. Biol.*, vol. 45, no. 6, pp. 984–993, 2008.
- [140] H. N. Arst Jr and D. J. Cove, "Nitrogen metabolite repression in *Aspergillus nidulans*," *Mol. & Gen. Genet.*, vol. 126, no. 2, pp. 111–141, 1973.
- [141] H. N. Arst Jr and D. W. MacDonald, "A gene cluster in *Aspergillus nidulans* with an internally located *cis*-acting regulatory region," *Nature*, vol. 254, no. 5495, pp. 26–31, 1975.
- [142] M. A. Newton, C. M. Kendzierski, C. S. Richmond, F. R. Blattner, and K. W. Tsui, "On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data," *J. Comput. Biol.*, vol. 8, no. 1, pp. 37–52, 2001.

- [143] K. M. Kerr, M. Martin, and G. A. Churchill, "Analysis of variance for gene expression microarray data," *J. Comput. Biol.*, vol. 7, no. 6, pp. 819–837, 2000.
- [144] T. Ideker, V. Thorsson, A. F. Siehel, and L. E. Hood, "Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data," *J. Comput. Biol.*, vol. 7, no. 6, pp. 805–817, 2000.
- [145] M. Mathieu and B. Felenbok, "The *Aspergillus nidulans* CREA protein mediates glucose repression of the ethanol regulon at various levels through competition with the ALCR-specific transactivator," *EMBO J.*, vol. 13, no. 17, pp. 4022–4027, 1994.
- [146] J. Omony, A. R. Mach-Aigner, L. H. de Graaff, G. van Straten, and A. J. B. van Boxtel, "Evaluation of design strategies for time course experiments: the XlnR regulon in *Aspergillus niger*," in *The 9th International Conference on Computational Methods in Systems Biology, (CMSB2011), Paris, France, ACM Digital Library*, pp. 25–33, 2011.
- [147] C. Panozzo, E. Cornillot, and B. Felenbok, "The CreA repressor is the sole DNA-binding protein responsible for carbon catabolite repression of the *alcA* gene in *Aspergillus nidulans* via its binding to a couple of specific sites," *J. Biol. Chem.*, vol. 273, no. 11, pp. 6367–6372, 1998.
- [148] M. Mathieu, S. Fillinger, and B. Felenbok, "In vivo studies of upstream regulatory cis-acting elements of the *alcR* gene encoding the transactivator of the ethanol regulon in *Aspergillus nidulans*," *Mol. Microbiol.*, vol. 36, no. 1, pp. 123–131, 2000.
- [149] P. Kulmburg, M. Mathieu, C. Dowzer, J. Kelly, and B. Felenbok, "Specific binding sites in the *alcR* and *alcA* promoters of the ethanol regulon for the CREA repressor mediating carbon catabolite repression in *Aspergillus nidulans*," *Mol. Microbiol.*, vol. 7, no. 6, pp. 847–857, 1993.
- [150] J. Omony, A. R. Mach-Aigner, L. H. de Graaff, G. van Straten, and A. J. B. van Boxtel, "Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*," *IEEE/ACM TCBB*, vol. 9, no. 5, pp. 1316–1325, 2012.
- [151] E. Heinzle, A. P. Biwer, and C. L. Cooney, "Development of Sustainable Bioprocess: Modeling and Assessment," *John Wiley and Sons, England*, 2007.
- [152] D. Bewer, M. Barenco, R. Callard, M. Hubank, and J. Stark, "Fitting ordinary differential equations to short time course data," *Philos. Transact. A Math. Phys. Eng. Sci.*, vol. 366, no. 1865, pp. 519–544, 2008.
- [153] B. C. Goodwin, "Temporal Organization In Cells; A Dynamic Theory Of Cellular Control Processes," *Academic Press, New York*, 1963.
- [154] C. Zhan and L. F. Yeung, "Parameter estimation in systems biology models using spline approximation," *BMC Syst. Biol.*, vol. 5, p. 14, 2011.
- [155] G. Lillacci and M. Khammash, "Parameter Estimation and Model Selection in Computational Biology," *PLoS Comput. Biol.*, vol. 6, p. e1000696, 2010.
- [156] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, "Systems Biology in Practice: Concepts, Implementation and Application," *Wiley-VCH, Weinheim*, 2005.

Bibliography

- [157] J. Omony, A. R. Mach-Aigner, L. H. de Graaff, G. van Straten, and A. J. B. van Boxtel, "Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*," *Submitted for publication*, 2012.
- [158] A. J. Ninfa and A. E. Mayo, "Hysteresis vs. Graded Responses: The Connections Make All the Difference," *Sci. STKE*, vol. 2004, no. 232, p. pe20, 2004.
- [159] T. E. Ideker, V. Thorsson, and R. M. Karp, "Discovery of regulatory interactions through perturbation: Inference and experimental design," in *Proc. Pac. Symp. Biocomput.*, vol. 5, pp. 302–313, 2000.
- [160] N. Friedman, "Inferring cellular networks using probabilistic graphical models," *Science*, vol. 303, no. 5659, pp. 799–805, 2004.
- [161] J. Tegnér, M. K. S. Yeung, J. Hastay, and J. J. Collins, "Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling," *PNAS, USA*, vol. 100, no. 10, pp. 5944–5949, 2003.
- [162] A. B. Tchagang, K. V. Bui, T. McGinnis, and P. V. Benos, "Extracting biologically significant patterns from short time series gene expression data," *BMC Bioinformatics*, vol. 10, p. 255, 2009.
- [163] A. Sinha and M. Markatou, "A Platform for Processing Expression of Short Time Series (PESTS)," *BMC Bioinformatics*, vol. 12, p. 13, 2011.
- [164] J. Ernst, G. Nau, and Z. Bar-Joseph, "Clustering short time-series gene expression data," *Bioinformatics*, vol. 21, no. Suppl. 1, pp. i159–i168, 2005.
- [165] M. Bansal, G. D. Gatta, and D. di Bernardo, "Inference of gene regulatory networks and compound mode of action from time series gene expression profiles," *Bioinformatics*, vol. 22, no. 7, pp. 815–822, 2006.
- [166] S. Widder, J. Schicho, and P. Schuster, "Dynamic patterns of gene regulation I: Simple two-gene systems," *J. Theor. Biol.*, vol. 246, no. 3, pp. 395–419, 2007.
- [167] S. Bornholdt, "Less Is More In Modeling Large Genetic Networks," *Science*, vol. 310, no. 5747, pp. 449–451, 2005.
- [168] M. I. Arnone and E. H. Davidson, "The hardwiring of development: organization and function of genomic regulatory systems," *Development*, vol. 124, no. 10, pp. 1851–1864, 1997.
- [169] D. Thieffry, A. M. Huerta, E. Pérez-Reuda, and J. Collado-Vides, "From Specific Gene Regulation to Genomic Networks: a Global Analysis of Transcriptional Regulation in *Escherichia coli*," *BioEssays*, vol. 20, no. 5, pp. 433–440, 1998.
- [170] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, no. 407, pp. 651–654, 2000.
- [171] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 3, pp. 41–42, 2001.
- [172] S. Y. Kim, S. Imoto, and S. Miyano, "Inferring gene networks from time series microarray data using dynamic Bayesian networks," *Bioinformatics*, vol. 4, no. 3, pp. 228–235, 2003.
- [173] M. Zou and S. D. Conzen, "A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data," *Bioinformatics*, vol. 21, no. 1, pp. 71–79, 2005.

- [174] Y. Zhang, Z. Deng, H. Jiang, and P. Jia, "Dynamic Bayesian Network (DBN) with Structure Expectation Maximization (SEM) for Modeling of Gene Network from Time Series Gene Expression Data," in *BIOCOMP*, pp. 41–47, 2006.
- [175] A. V. Werhli, M. Grzegorzczak, and D. Husmeier, "Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks," *Bioinformatics*, vol. 22, no. 20, pp. 2523–2531, 2006.
- [176] P. Spirtes, G. Glymour, S. Kauffman, V. Aimalie, and F. Wimberly, "Constructing Bayesian network models of gene expression networks from microarray data," in *Proc. Atlantic Symp. Comp. Biol., Genome Information Systems & Technology*, 2000.
- [177] J. Pearl, "Causality: Models, Reasoning, and Inference," *Cambridge University Press*, Mar. 2000.
- [178] K. Murphy and S. Mian, "Modelling Gene Expression Data Using Dynamic Bayesian Networks," *Technical Report*, 1999.
- [179] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian networks to analyze expression data," *J. Comp. Biol.*, vol. 7, no. 3-4, pp. 601–620, 2000.
- [180] D. Heckerman, "Bayesian Networks for Data Mining," *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 79–119, 1997.
- [181] S. Rogers and M. Girolami, "A Bayesian regression approach to the inference of regulatory networks from gene expression data," *Bioinformatics*, vol. 21, no. 14, pp. 3131–3137, 2005.
- [182] P. Brazhnik, A. de la Fuente, and P. Mendes, "Gene networks: how to put the function in genomics," *Trends Biotechnol.*, vol. 20, no. 11, pp. 467–472, 2002.
- [183] V. Pihur, S. Datta, and S. Datta, "Reconstruction of genetic association networks from microarray data: A partial least squares approach," *Bioinformatics*, vol. 24, no. 4, pp. 561–568, 2008.
- [184] T. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins, "Inferring genetic networks and identifying compound mode of action via expression profiling," *Science*, vol. 301, no. 5629, pp. 102–105, 2003.
- [185] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.
- [186] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, "Analyzing multiple nonlinear time series with extended Granger causality," *Science direct, Elsevier, Phys. Lett. A*, vol. 324, no. 3, pp. 26–35, 2004.
- [187] N. D. Mukhopadhyay and S. Chatterjee, "Causality and pathway search in microarray time series experiment," *Bioinformatics*, vol. 23, no. 4, pp. 442–449, 2007.
- [188] R. Nagarajan, "A Note on Inferring Acyclic Network Structures Using Granger Causality Tests," *IJB*, vol. 5, no. 1, p. Article 10, 2009.
- [189] A. Fujita, P. Severino, J. R. Sato, and S. Miyano, "Granger causality in systems biology: modeling gene networks in time series microarray data using vector autoregressive models," in *Proceeding BSB2010 Proceedings of the Advances in bioinformatics and computational biology, and 5th Brazilian conference on Bioinformatics*, 2010.

Bibliography

- [190] C. Zou, C. Ladroue, S. Guo, and J. Feng, "Identifying interactions in the time and frequency domains in local and global networks - A Granger Causality Approach," *BMC Bioinformatics*, vol. 11, p. 337, 2010.
- [191] A. de la Fuente, N. Bing, I. Hoeschele, and P. Mendes, "Discovery of meaningful associations in genomic data using partial correlation coefficients," *Bioinformatics*, vol. 20, no. 18, pp. 3565–3574, 2004.
- [192] A. Reverter and E. K. F. Chan, "Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks," *Bioinformatics*, vol. 24, no. 21, pp. 2491–2497, 2008.
- [193] P. M. Magwene and J. Kim, "Estimating genomic coexpression networks using first-order conditional independence," *Genome Biol.*, vol. 5, no. 12, p. R100, 2004.
- [194] A. S. Butte and I. S. Kohane, "Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements," in *Proc. Pac. Symp. Biocomput.*, pp. 418–429, 2000.
- [195] J. E. Hopcroft, R. Motwani, and J. D. Ullman, "Introduction to Automata Theory, Languages, and Computation," *Addison Wesley, Boston, MA, USA*, 2006.
- [196] E. R. Dougherty and I. Shmulevich, "Mappings Between Probabilistic Boolean Networks," *Signal Processing*, vol. 83, no. 4, pp. 799–809, 2003.
- [197] I. Shmulevich, E. R. Dougherty, and W. Zhang, "From Boolean to probabilistic Boolean networks as models of genetic regulatory networks," *Proceedings of the IEEE*, vol. 90, no. 11, pp. 1778–1792, 2002.
- [198] I. Shmulevich and S. A. Kauffman, "Activities and Sensitivities in Boolean network models," *Phys. Rev. Lett.*, vol. 93, no. 4, pp. 048701(1–4), 2004.
- [199] R. F. Hashimoto, S. Kim, I. Shmulevich, W. Zhang, M. L. Bittner, and E. R. Dougherty, "Growing genetic regulatory networks from seed genes," *Bioinformatics*, vol. 20, no. 8, pp. 1241–1247, 2004.
- [200] D. M. Rocke and B. Durbin, "A model for measurement error for gene expression arrays," *J. Comput. Biol.*, vol. 8, no. 6, pp. 557–569, 2001.
- [201] S. Huang, "Gene expression profiling, genetic networks and cellular states: an integrating concept for tumorigenesis and drug discovery," *J. Mol. Med.*, vol. 77, no. 6, pp. 469–480, 1999.
- [202] D. Repsilber, H. Liljenström, and S. G. Andersson, "Reverse engineering of regulatory networks: simulation studies on a genetic algorithm approach for ranking hypotheses," *Biosystems*, vol. 66, no. 1-2, pp. 31–41, 2002.
- [203] E. Dimitrova, L. D. Garcia-Puente, F. Hinkelmann, A. S. Jarrah, R. Laubenbacher, B. Stigler, and P. Vera-Licona, "Parameter estimation for Boolean models of biological networks," *TCS*, vol. 412, no. 26, pp. 2816–2826, 2010.
- [204] I. Shmulevich and W. Zhang, "Binary Analysis and Optimization-Based Normalization of Gene Expression Data," *Bioinformatics*, vol. 18, no. 4, pp. 555–565, 2002.
- [205] L. J. Steggles, R. Banks, O. Shaw, and A. Wipat, "Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach," *Bioinformatics*, vol. 23, no. 3, pp. 336–343, 2006.

- [206] L. Glass and S. A. Kauffman, "The logical analysis of continuous, non-linear biochemical control networks," *J. Theor. Biol.*, vol. 39, no. 1, pp. 103–129, 1973.
- [207] A. Goldbeter, "Computational approaches to cellular rhythms," *Nature*, vol. 420, pp. 238–245, 2002.
- [208] T. Akutsu, S. Miyano, and S. Kuhara, "Identification of genetic networks from a small number of gene expression patterns under the Boolean network model," in *Proc. Pacific Symp. Biocomput.*, vol. 4, pp. 17–28, 1999.
- [209] E. August and A. Papachristodoulou, "Efficient, sparse biological network determination," *BMC Syst. Biol.*, vol. 3, p. 25, 2009.
- [210] Y. Shi, T. Mitchell, and Z. Bar-Joseph, "Inferring pairwise regulatory relationships from multiple time series datasets," *Bioinformatics*, vol. 23, no. 6, pp. 755–763, 2007.
- [211] M. J. L. de Hoon, Y. Makita, K. Nakai, and S. Miyano, "Prediction of Transcriptional Terminators in *Bacillus subtilis* and Related Species," *PLoS Comput. Biol.*, vol. 1, no. 3, p. e25, 2006.
- [212] J. Mazur, D. Ritter, G. Reinelt, and L. Kaderali, "Reconstructing nonlinear dynamic models of gene regulation using stochastic sampling," *BMC Bioinformatics*, vol. 10, p. 448, 2009.
- [213] J. Almeida and E. Voit, "Neural-Network-Based Parameter Estimation in S-System Models of Biological Networks," *Genome Informatics*, vol. 14, pp. 114–123, 2003.
- [214] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi, and M. Tomita, "Dynamic modeling of genetic networks using genetic algorithm and S-system," *Bioinformatics*, vol. 19, no. 5, pp. 643–650, 2003.
- [215] S. Kimura, K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya, "Inference of S-system models of genetic networks from noisy time-series data," *Chem-Bio. Informatics Journal*, vol. 4, no. 1, pp. 1–14, 2004.
- [216] S. Kimura, K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya, "Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm," *Bioinformatics*, vol. 21, no. 7, pp. 1154–1163, 2005.
- [217] N. Norman and H. Iba, "Inference of gene regulatory networks using S-system and differential evolution," in *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 439–446, 2005.
- [218] C. Spieth, F. Streichert, N. Speer, and A. Zell, "A memetic inference method for gene regulatory networks based on S-systems," in *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 152–157, 2004.
- [219] R. Thomas, S. Mehrotra, E. T. Papoutsakis, and V. Hatzimanikatis, "A model-based optimization framework for the inference on gene regulatory networks from DNA array data," *Bioinformatics*, vol. 20, no. 17, pp. 3221–3235, 2004.
- [220] X. Yang, J. E. Dent, and C. Nardini, "An S-System Parameter Estimation Method (SPEM) for biological networks," *J. Comp. Biol.*, vol. 19, no. 2, pp. 175–187, 2012.

Bibliography

- [221] D.-Y. Cho, K.-H. Cho, and B.-T. Zhang, "Identification of biochemical networks by S-tree based genetic programming," *Bioinformatics*, vol. 22, no. 13, pp. 1631–1640, 2006.
- [222] E. O. Voit, "Computational Analysis of Biochemical Systems," *Cambridge University Press*, 2000.
- [223] M. Kabir, N. Norman, and H. Iba, "Reverse engineering gene regulatory network from microarray data using linear time-variant model," *BMC Bioinformatics*, vol. 11, no. Suppl. I, p. S56, 2010.
- [224] A. Fersht, "Enzyme structure and mechanism," *W. H. Freeman and Company, New York*, vol. 2, 1985.
- [225] J. H. Hofmeyr and A. Cornish-Bowden, "The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models," *Comput. Appl. Biosci.*, vol. 13, no. 4, pp. 377–385, 1997.
- [226] K. Schittkowski, "Parameter estimation in systems of nonlinear equations," *Numerische Mathematik*, vol. 68, no. 1994, pp. 129–142, 1994.
- [227] R. A. Fisher, "On an absolute criterion for fitting frequency curves," *Messenger of Mathematics*, vol. 41, pp. 155–160, 1912.
- [228] L. Ljung, "Systems identification," *Theory for the user, Prentice Hall PRT.*, p. 21, 1999.
- [229] F. d'Alché Buc and V. Schachter, "Modeling and identification of biological networks," in *International Symposium on Applied Stochastic Models and Data Analysis, Brest (France)*, 2005.
- [230] Z. R. Yang, "Machine Learning Approaches to Bioinformatics (Science, Engineering, and Biology Informatics)," *World scientific*, vol. 4, 2009.

Summary

In the past few decades there has been tremendous progress in the fields of Bioinformatics and Systems Biology, particularly in studying biological networks. The development of high through-put data acquisition techniques has been vital in achieving this progress. Generally, the underlying mechanisms of gene regulation constitute an important process in living cells. This process controls transmission of information encoded in the DNA sequence to proteins through the central dogma of molecular biology.

The study of biological networks helps improve our understanding of complex molecular processes. The xylanolytic XlnR regulon of *A. niger* encodes several enzymes, which are responsible for various processes in the cell. For instance, some of these enzymes are involved in the degradation of the polysaccharide xylan and cellulose. This thesis provides a detailed look into modeling regulon dynamics and target gene regulation mechanisms. Here, the XlnR regulon of *A. niger* is used as a model sample system to address challenges, which are associated with the process of network dynamics and experimental design. The XlnR regulon dynamics are studied in Chapter 2, and the models are evaluated on real experimental data in Chapter 6. Primarily the models were based on qualitative prior information of the regulation mechanisms for the target genes. This thesis also highlights the need to take strategic experimental designs into consideration when studying biological networks by advocating the deployment of time course experiments. It is through integration of theory and experiments that interesting challenges on networks such as the XlnR regulon can be addressed.

The application of optimal and strategic experimental design in time course experiments to study biological networks is still at its infancy. Only a few studies have paid attention to this issue in an attempt to improve the information content in their data sets. The benefits of using induction signals for studying biological networks are illustrated in this thesis. To obtain meaningful results, various aspects of experimental design were studied. There is much to gain from the use of sensitivity functions, the Fisher Information Matrix (FIM), and the E-modified criterion in the design of time course experiments using mathematical models (Chapter 3). For instance, the extent to which the individual

kinetic parameters influence the measured mRNA levels (gene expression) can be assessed by using sensitivity functions. This underlines the need for smart network perturbation techniques and data sampling strategies, both of which are crucial for improving the confidence intervals on the parameters.

A comparison of transcription in strains of *A. niger* was performed in Chapters 4 and 5. These studies provided information on both the qualitative and quantitative roles of the carbon catabolite repressor protein (CreA). The response of xylan backbone-degrading enzymes to different D-xylose induction concentrations was investigated. The expression of accessory enzyme-encoding genes was found to be favored by using a high D-xylose concentration. The extent to which transcription varies between the wild type and mutant strains after D-xylose induction was assessed. The potential of D-xylose to induce expression of cellulase-encoding genes (target genes) was assessed. High D-xylose concentrations are preferable for inducing the gene expression of enzymes in the pentose metabolic pathway. The absence of a functional CreA positively influences expression of genes encoding xylan degrading enzymes, independent of D-xylose concentration. The *xlnR* gene was found to be constitutively expressed. Without performing time course experiments with frequent sampling and high data resolution, there is a risk of missing out on the time dynamics of the expression of the target genes. The role of CreA in the regulation of transcription in the *A. nidulans*, *A. oryzae* and *A. niger* species of *Aspergillus* has been widely studied. It mediates repression of the target genes and it was also found that mutation of CreA significantly increases transcription in the target genes of the XlnR regulon following low D-xylose induction concentrations.

Basing on the work in Chapter 6, we see that mechanistic modeling with differential equations is a powerful way to study networks such as the XlnR regulon. The XlnR regulon was perturbed to generate transcription data from both *in silico* and wet-lab qPCR experiments. Network induction is a good way to study the information content of genomic data sets. Compared to the low D-xylose induction, transcription was found to be lower following high D-xylose induction. Exceptions were found in a very few number of genes, which did not exhibit significant differences in transcription under the two conditions. Using nonlinear differential equations, the regulatory mechanisms of the XlnR regulon were modeled and validated on qPCR data. The strategic experimental designs led to new insights on how to maximize the transcription data information content. It was shown that differences exist both between the responses of the different target genes over time.

In Chapter 6, the quantitative information of CreA proved very useful for modifying the earlier-on proposed models in Chapters 2 and 3, and to evaluate the new models. Induction with D-xylose, particularly at high concentrations, yields a diversity of transcription profiles. To understand such transcription dynamics, it is crucial to understand the roles of the individual network components e.g. genes, and transcription factors or any other signaling molecules. The explanatory power of the models was significantly improved by incorporating these components into the modeling.

Getting reliable inferences from experimental data sets is essential for the study of biological networks. However, the validity of results depends on a number of processes, such as the proper planning and execution of the experimental procedure of data collection and analysis. Fortunately, numerous analytical tools are available to facilitate such studies. The work in this thesis provides a basis for future research with regards to network. This includes issues such as: i) optimal network perturbation and design of time course experiments, ii) analysis of time constants and transcription dynamic responses, iii) data sampling strategies, and iv) evaluation of candidate model structures and their appropriateness in describing the transcription dynamics from measured mRNA time course data.

Samenvatting

Op het gebied van de bioinformatica en de systeembioïogie is de afgelopen decennia een geweldige vooruitgang geboekt, en zeker bij de studie van biologische netwerken. Het ontwikkelen van data acquisitie technieken met een hoge doorvoersnelheid zijn hierbij van vitaal belang geweest. In levende cellen vormt het gen regulatiemechanisme een belangrijk proces. Het reguleert de transmissie van de in de DNA sequentie vastgelegde informatie naar eiwitten conform het centrale dogma van de moleculaire biologie.

De studie van biologische netwerken draagt bij aan ons begrip van complexe moleculaire processen. Het xylanolytische regulon XlnR van *A. niger* codeert verschillende enzymen, die verantwoordelijk zijn voor diverse processen in de cel. Sommige van deze enzymen zijn bijvoorbeeld betrokken bij de afbraak van het polysacharide xylan en van cellulose. Dit proefschrift biedt een gedetailleerd zicht op het modelleren van regulon dynamica en regulatiemechanismen van doel genen. Het XlnR regulon van *A. niger* wordt hier gebruikt als een voorbeeld systeem om uitdagingen die verbonden zijn met het process van netwerk dynamica en experimentontwerp aan te pakken. De dynamica van het XlnR regulon wordt bestudeerd in Hoofdstuk 2, en de modellen worden geëvalueerd op werkelijke experimentele gegevens in Hoofdstuk 6. Aanvankelijk waren de modellen gebaseerd op tevoren bekende kwalitatieve informatie over het regulatiemechanisme van de doelgenen. Dit proefschrift belicht ook de wenselijkheid van strategisch experiment ontwerp bij het bestuderen van biologische netwerken, door een pleidooi te houden voor het verrichten van tijdreeks experimenten. Juist door integratie van theorie en experiment kan het hoofd worden geboden aan de interessante uitdagingen met betrekking tot netwerken als het XlnR regulon.

De toepassing van optimaal en strategisch experimentontwerp voor tijdreeks experimenten ter bestudering van biologische netwerken staat nog in de kinderschoenen. Er zijn slechts een paar studies die hier aandacht aan hebben besteed in een poging om de informatieinhoud van de verzamelde gegevens te verhogen. De voordelen van het gebruik van inductiesignalen om biologische netwerken te bestuderen worden geïllustreerd in dit proefschrift. Om tot zinvolle resultaten te komen zijn diverse aspecten van experimentontwerp bestud-

eerd. Veel kan worden gewonnen door het gebruik van gevoeligheidsfuncties, de Fisher Informatie Matrix (FIM), en het E-gemodificeerd criterium bij het ontwerpen van tijdreeks experimenten op basis van wiskundige modellen. Zo kan bijvoorbeeld de mate waarin de individuele kinetische parameters de gemeten mRNA niveaus (gen expressies) beïnvloeden worden vastgesteld door middel van gevoeligheidsfuncties. Dit onderstreept de behoefte aan slimme netwerk perturbatie technieken en data bemonsterings strategieën, welke beide cruciaal zijn voor het verbeteren van de betrouwbaarheidsintervallen van de parameterwaarden.

In de hoofdstukken 4 en 5 werd een vergelijking uitgevoerd van de transcriptie in verschillende stammen van *A. niger*. Deze studies leverden zowel in kwalitatieve als kwantitatieve informatie op over rol van het koolstof katabole repressor eiwit CreA. De respons van xylanketen afbrekende enzymen in reactie op verschillende D-xylose inductie concentraties werd bestudeerd. Gevonden werd dat de expressie van de modifierende enzym-coderende genen bevordert wordt door een hoge D-xylose concentratie. Nagegaan werd in hoeverre transcriptie tussen de wild type en de gemuteerde stammen na inductie door D-xylose verschilt. Het vermogen van D-xylose om cellulase coderende genen (de "doel genen") tot expressie te brengen werd bepaald. Hoge D-xylose concentraties zijn bevorderlijk om gen expressie te induceren van enzymen in de pentose metabole route. De afwezigheid van functioneel CreA heeft een positieve invloed op de expressie van genen die coderen voor xylan-afbrekende enzymen, ongeacht de hoogte van de D-xylose concentratie. Gevonden werd dat het *xlnR* gen altijd tot expressie komt. Zonder dynamische experimenten te doen met frequente bemonstering en hoge data resolutie loopt men het risico informatie te missen in het tijdsverloop van de response van de doelgenen. De rol van CreA bij de regulatie van de transcriptie in de *Aspergillus* soorten *A. nidulans*, *A. oryzae* en *A. niger* is ruim bestudeerd. Het veroorzaakt de repressie van de doel genen en gebleken is ook dat mutatie van CreA de transcriptie in de doel genen van het XlnR regulon bij lage inducerende concentraties van D-xylose significant verhoogt.

Het werk in hoofdstuk 6 laat zien dat mechanistisch modelleren met differentiaalvergelijkingen een krachtige manier is om netwerken zoals het XlnR regulon te bestuderen. Het XlnR regulon werd geïnduceerd om transcriptie data te genereren van zowel *in silico* als natte laboratorium qPCR experimenten. Netwerk inductie is een goede manier om de informatie inhoud van genoom data te bestuderen. Gevonden werd dat de transcriptie bij hoge D-xylose inductie lager is dan bij inductie met lage concentraties. Slechts bij een klein aantal genen was er geen significant verschil in transcriptie tussen de twee situaties.

Met behulp van niet-lineaire differentiaalvergelijkingen werden de regulatorische mechanismen van het XlnR regulon gemodelleerd en gevalideerd op qPCR data. Het methoden voor strategisch experiment ontwerp hebben tot nieuw inzicht geleid hoe het informatieniveau van transcriptie gegevens kan worden gemaximaliseerd. Aangetoond werd dat er verschillen bestaan in response tussen verschillende doelgenen.

In hoofdstuk 6 bleek de kwantitatieve informatie van CreA buitengewoon bruikbaar om de eerder voorgestelde modellen uit de hoofdstukken 2 en 3 aan te passen, en om de nieuwe modellen te evalueren. Inductie met D-xylose, in het bijzonder bij hoge concentraties, geeft een diversiteit aan transcriptieprofielen. Om deze typen transcriptieprofielen te begrijpen is het cruciaal om de rollen te begrijpen van de individuele netwerk componenten, bijv. genen, en transcriptiefactoren of welke andere signaalmoleculen dan ook. Het verklarend vermogen van de modellen werd significant verbeterd door deze componenten in de modellen op te nemen.

Het verkrijgen van betrouwbare gevolgtrekkingen uit experimentele data sets is essentieel bij het bestuderen van biologische netwerken. De geldigheid van de resultaten hangt echter af van een aantal processen, zoals de juiste planning en uitvoering van de experimentele procedures waarmee gegevens worden verzameld en geanalyseerd. Gelukkig bestaan er talrijke analytische gereedschappen om dergelijke studies te ondersteunen. Het werk beschreven in dit proefschrift biedt een fundering voor toekomstig onderzoek met betrekking tot netwerk reconstructie. Hieronder begrepen zijn: i) optimale netwerk perturbatie en optimaal ontwerp van tijdreeks experimenten, ii) analyse van tijdconstantes en dynamische transcriptie responsies, iii) bemonsterings strategieën, en iv) evaluatie van kandidaat modelstructuren en van hun geschiktheid om de uit gemeten mRNA tijdreeksen afgeleide transcriptie dynamica te beschrijven.

Acknowledgements

It has been a wonderful period of research and lots of experiences in this great institution, Wageningen University. My words of thanks first and foremost go to each and every staff member of the university that has contributed in my well-being and research. I am grateful to the Systems and Control Group (SCO) now called Biomass Refinery and Process Dynamics (BRD), and the Laboratory of Systems and Synthetic Biology (SSB) for offering me a conducive working environment during my research. With help from the ever welcoming staff members, I feel a much better scientist today than I was yesterday.

Many thanks to my supervisors who were very committed in helping me throughout this research work. Many thanks to the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences) and the "Investeringsplan / Ondernemingplan" (IPOP) for all the financial support that enabled this project work to be carried out. I also extend special words of thanks to K.J. Keesman, J.D. Stigter, R.J.C. van Ooteghem, C.J. van Asselt and L.G. van Willigenburg of the Systems and Control Group for assisting me comprehend the fundamental concepts of systems theory and systems identification. This I found very helpful. I thank A.R. Mach-Aigner and Birgit Jovanovic of the Institute of Chemical Engineering, Vienna University of Technology (TU Vienna) for all the experimental work they performed during my research.

I am very grateful the wonderful time shared with the other PhD students (James, Xin, Ellen, Emmanuel, Johannes, Nurul, ...) of the Systems and Control group. Mrs. M. Portengen has particularly been very helpful. I extend special words of thanks to Prof.dr.ir. G. van Straten for his critical but objective comments as my thesis supervisor. To L.H. de Graaff and A.J.B. van Boxtel, I am blessed for having you as my thesis co-supervisors. Your commitment helped see me through the last four years as a PhD student. During this time I learnt a lot of things and shared wonderful experiences with fellow students and senior researchers in the numerous groups namely: The Systems Biology Colloquia (at Biometris), mathematics and statistics discussion group.

I would like to thank both Prof. L.S. Luboobi and Dr. M. Nabasirye of Makerere University Kampala (MUK) for all the assistance they extended to me

while I was a student at MUK and University of Nairobi. There are of course many people who have been instrumental in facilitating my studies, offering advise and guiding me in my career choices. I would like to particularly extend my sincere gratitude to Prof.dr. G.R. Renardel de Lavalette of the University of Groningen. Without you, my dream of pursuing postgraduate studies and furthering my career as a scientist might never have materialized.

For a long time I have been studying far away from my loved ones in Uganda. I thank them for their for social, financial and material support through the years. My sincere gratitude and appreciation goes to my brothers, sisters, father (Nimrod Oloo, RIP), and mother (Jesker Aber, RIP). To my Ugandan friends in Wageningen: Thomas, Agnes, Cyrus, Amos, Harriet, Bihoctavia, ... ; I would like to thank you all for the times we shared together.

Anything I did not really enjoy in the last four years? Of course, memories of the blistering cold winter-winds (*elfstedentocht*), short winter days, and long hours of darkness - stressing isn't it? For the future, I wish for even better days. Above all, I wholeheartedly thank God for all the blessings.

JIMMY OMONY

Email: jimmy.omony@gmail.com

About the author - English



Jimmy Omony was born on 5th, Jun. 1979 in Entebbe, Uganda. He pursued a PhD research in Systems Biology at Wageningen University (WUR), The Netherlands (Jun. 2008 to Aug. 2012). During which time he also served as a PhD student representative for the Systems and Control group (SCO) to the VLAG PhD council. He holds a Masters of Science in Mathematics (with specialty in Bioinformatics) from University of Groningen (2005 – 07), (The Netherlands), **Thesis title:** "Reconstructing a Gene Regulatory Network (GRN) of the bacterium *Bacillus subtilis*"; a Masters of Science in Biometry from University of Nairobi - Kenya, (2003 – 04), **Thesis title:** "Modeling choice of breeding services by Smallholder dairy farmers in the Kenyan Highlands"; a Bachelor of Science in Mathematics (major) and Psychology (minor) from Makerere University, Uganda (1999 – 2002); The Uganda Advanced Certificate of Education, U.A.C.E (1997 – 98); The Uganda Ordinary Certificate of Education, U.C.E (1993 – 96); and the Uganda Primary Leaving Examination, P.L.E (–1992). He was also a former member of the international Biometric Society Kenya (2003 – 04). He worked as the Monitoring and Evaluation specialist (2005) with World Vision Uganda (KURET) and doubled as a part time lecturer at Gulu University (Uganda, 2005). In Apr-Nov, 2004 he under-went a student fellowship training at the International Livestock Research Institute (ILRI)-Kenya. In 2006 – 07 he was elected and served as the President of the African Student Community (ASC) at the University of Groningen. He also taught Mathematics oriented courses at undergraduate level at university of Nairobi (Kenya).

Since November 2012, the author works as a postdoctoral researcher at the Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen.

Over de auteur - Nederlands

*J*immy Omony werd geboren op 5 Juni 1979 in Entebbe, Oeganda. Hij deed zijn PhD onderzoek in Systeembioogie aan de Wageningen Universiteit (WUR). Hij was vertegenwoordiger voor de PhD-studenten van de Systems and Control Group (SCO) bij de VLAG PhD-raad. Hij heeft een Masters of Science in Wiskunde (met een specialisatie in Bioinformatica) aan de RijksUniversiteit Groningen (RuG) met succes gevolgd van 2005 tot 2007. Zijn Masterthesis was getiteld: "Reconstructing a Gene Regulatory Network (GRN) of the bacterium *Bacillus subtilis*". Daarnaast heeft hij een Masters of Science in Biometrie behaald aan de Universiteit van Nairobi (2003 – 2004) met de thesis: "Modeling choice of breeding services by Smallholder dairy farmers in the Kenyan Highlands". Hij heeft een Bachelor of Science in wiskunde met als minor Psychologie van de Makerere Universiteit in Uganda (1999 – 2002). Daarvoor behaalde hij de Uganda Advanced Certificate of Education, U.A.C.E (1997 – 98); Uganda Ordinary Certificate of Education, U.C.E (1993 – 96); en de Uganda Primary Leaving Examination, P.L.E (–1992). Jimmy was daarnaast lid van de international biometrievereniging van Kenia van 2003 tot 2004. Hij werkte in 2005 als Monitoring en Evaluatie specialist bij World Vision Uganda, terwijl hij part-time les gaf aan de Universiteit van Gulu (Uganda). Van April tot November 2004 volgde hij een training aan het International Livestock Research Institute (ILRI) in Kenia. In 2006 – 2007 was hij voorzitter van de African Student Community (ASC) in Groningen. Hij gaf ook wiskundecursussen op "undergraduate level" aan de Universiteit van Nairobi (Kenia).

Vanaf November 2012 werkt de auteur als postdoctoraal onderzoeker bij het Groningen Biomolecular Sciences and Biotechnology Institute (GBB) aan de Rijksuniversiteit Groningen.

List of publications

- **J. Omony**, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel. Modeling and analysis of the dynamic behavior of the XlnR regulon in *Aspergillus niger*. *BMC Systems Biology*, 5(Suppl 1): S14, 2011.
- A.R. Mach-Aigner, **J. Omony**, B. Jovanovic, A.J.B. van Boxtel, L.H. de Graaff. D-xylose concentration-dependent Hydrolase Expression Profiles and the role of CreA and XlnR in *A. niger*. *Applied and Environmental Microbiology*, 78(9), pp. 3145 – 3155, 2012.
- **J. Omony**, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel. Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(5), pp. 1316 – 1325, 2012.
- **J. Omony**, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel. Quantification of the effect of the carbon catabolite repressor CreA on transcription in the XlnR regulon of *Aspergillus niger*. 2012. *Submitted for publication*.
- **J. Omony**, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel. On the transcription dynamics in the XlnR regulon of *A. niger* to D-xylose triggers. 2012. *In preparation*.

(Peer-reviewed) conference proceedings

- **J. Omony**, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel (2012). The modeling cycle in genetic network reconstruction: the XlnR regulon of *Aspergillus niger*. In *The 31-st Benelux Meeting on Systems and Control. Book of abstracts*. Heijderbos, Heijen/Nijmegen, The Netherlands.
- **J. Omony**, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel (2011). Evaluation of design strategies for time course experiments: the XlnR regulon in *Aspergillus niger*. In *The 9th International Conference on Computational Methods in Systems Biology (CMSB'11)*, ACM Digital Library. pg. 25 – 33. Paris, France. IEEE Technical Committee on Simulation (TCSIM): **Best Student Paper Award**.
- **J. Omony**, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel (2011). Parameter estimation in genetic networks using a constrained stochastic space search method. In *The 30th Benelux Meeting on Systems and Control. Book of abstracts*. Lommel, Belgium.

-
- **J. Omony**, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel (2010). Modeling the dynamics of the XlnR regulon network in *Aspergillus niger*. In *The 4th International Conference on Computational Systems Biology (ISB2010)*, pg. 128 – 138, Suzhou, China.
 - **J. Omony**, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel (2010). Unraveling the dynamics in genetic networks from gene expression data. In *The Netherlands Biotechnology Congress. Book of abstracts*, NBC-13. Ede, The Netherlands.
 - **J. Omony**, G. van Straten, A.J.B. van Boxtel, L.H. de Graaff (2009). Reverse-engineering genetic networks without prior knowledge. In *The 28th Benelux Meeting on Systems and Control. Book of abstracts*. Spa, Belgium.

Posters

- **J. Omony**, A.R. Mach-Aigner, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel (2011). Evaluation of design strategies for time course experiments: the XlnR regulon in *Aspergillus niger*. PhD trip: Beijing and Shanghai, China.
- **J. Omony**, L.H. de Graaff, G. van Straten, A.J.B. van Boxtel (2011). Dynamic properties of the XlnR regulon in *Aspergillus niger*. PhD trip: Beijing and Shanghai, China.



Contest Winner

This award is presented to

Jimmy Omony

For the Student Best Paper Award at CMSB 2011

The 9th International Conference on Computational Methods in Systems Biology

21 september 2011

François Fages
Program Committee Chair

Dave Cavalcanti
IEEE TCSIM representative

Prize: 500\$ grant

Figure B.2: Student Best Paper Award Certificate: IEEE Technical Committee on Simulation (TCSIM).

Completed training activities

Discipline specific activities

DISC: 31-st Benelux Meeting on Systems and Control. Heijderbos, Heijen/Nijmegen, The Netherlands. 2012.

CMSB2011: 9-th International Conference on Computational Methods in Systems Biology. Paris, France.

DISC: 30-th Benelux Meeting on Systems and Control. Lommel, Belgium. 2011.

ISB2010: 4-th International Conference on Computational Systems Biology. Suzhou, China.

DISC: 28-th Benelux Meeting on Systems and Control. Spa, Belgium. 2009.

EECI: The use of Poisson Processes in Modeling and Nonlinear Control, Gif-sur-Yvette, (France) Module-M15, 2009.

SCO: Systems and Control Theory - Part I, 2009.

DISC: Summer school on cells and systems, Driebergen-Zeist, The Netherlands, 2008.

SENSE: Parameter Estimation and Model structure Identification, 2008.

DISC: Modeling and control of hybrid systems, 2008.

General courses

WGS: Course Career perspectives (CCP), 2011.

WGS: Philosophy and Ethics of Food Science and Technology (PEFST), 2011.

CENTA: Scientific writing, SW, 2009.

CENTA: Workshop presentation skills (PS), 2009.

Patent workshop: IOP Genomics / Netherlands Patent Office, Utrecht, 2009.

WGS: Project and time management, 2009.

WGS: Introduction to Information Literacy, EndNote, 2008.

Optionals

Laboratory of Microbiology PhD trip to China and Japan, 2011.

NBC-13: Netherlands Biotechnology Congress, 2010.

WUR: Systems Biology Colloquia (Biometris), 2009 – 2012.

WUR: Systems Biology Day, 2009, 2011.

WUR: Systems Biology Afternoon, 2009.

WUR: Training Performance and Development Interview, 2009.

WUR: Biotechnology: Cell Physiology and Genetics, 2009.

Dutch I (DU I, Level A1), CENTA, (2009).

Dutch II (DU II, Level 1B), Volksuniversiteit - Wageningen (2009).

SCO: Writing project proposal, 2008 - 2009.

VLAG: VLAG student week, 2008.

The research described in this thesis was financially supported by the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences) and the "Investeringsplan / Ondernemingplan" (IPOP) program of Wageningen University.

Financial support from the Systems and Control Group (SCO) now called Biomass Refinery and Process Dynamics (BRD) for printing this thesis is gratefully acknowledged.