Inaugural-Dissertation

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich-Mathematischen Gesamtfakultät

der

Ruprecht - Karls - Universität

Heidelberg

vorgelegt von

M.Sc. Sarah Kaspar

geboren in Lindenfels

Tag der mündlichen Prüfung:

The role of promoter architecture in transduction, integration and multiplexing of environmental signals in bacteria

Gutachter: Prof. Dr. Ursula Kummer

                 Dr. Jürgen Pahle

Abstract

For the processing of environmental signals, bacteria often use two-component systems (TCSs), in which a transcription factor (TF) is activated upon phosphorylation. There is a high diversity both in the way in which phosphorylation activates the TF (phospho-mode) and in the architecture of the TCS-driven promoters. In this work, I model gene expression in TCSs in order to study the interplay of promoter architecture and phospho-mode with respect to the distinct processing tasks of signal transduction, integration and multiplexing.

First a comprehensive study implements biologically relevant phospho-modes and investigates their influence on the promoter architectures required for effective signal transduction. This study is also the modeling basis for the whole thesis. I find that effective signal transduction is accomplished by different promoter architectures, depending on the mode of phosphorylation. Further, the phospho-mode also determines to which extent the response shape is tunable through modulations in the promoter architecture.

In the quorum sensing (QS) signaling pathway in *Bacillus subtilis*, the integration of two signals on the gene expression level is investigated. When only the pathway architecture is considered, both signals seem to converge, since they activate the same TF. However, my model identifies promoter architectural features which allow to control the weighting of the two signals. The meaning of these results is discussed with respect to the evolution of QS-dependent promoters. Finally, based on model-derived promoter design criteria, I suggest a synthetic multiplexing system for experimental use. The purpose of this system is to allow decoding of the two QS signals on the expression level and thus inference on their natural distribution.

The results of this thesis can support the interpretation of promoter sequence data and inform design questions in synthetic biology.

Zusammenfassung

Um Signale aus ihrer Umgebung zu verarbeiten, nutzen Bakterien häufig Zweikomponentensysteme (two-component systems, TCSs), bei denen ein Transkriptionsfaktor (TF) durch Phosphorylierung aktiviert wird. Dabei weisen sowohl die Art der Aktivierung durch Phosphorylierung (Phosphomodus) als auch die Architekturen der angesteuerten Promotoren eine hohe Diversität auf. In dieser Arbeit wird mit Hilfe von Modellierung der Genexpression in TCSs das Zusammenspiel von Promoterarchitektur und Phosphomodus im Hinblick auf die verschiedenen "Signalverarbeitungsaufgaben" der Signaltransduktion, Integration und des Multiplexings untersucht.

Eine umfassende Studie beleuchtet zunächst biologisch relevante Phosphomodi und deren Einfluss auf die für die Effektivität von Signaltransduktion erforderliche Promoterarchitektur. Sie legt außerdem die Modellierungsgrundlage für die gesamte Arbeit. Es zeigt sich, dass möglichst effektive Signaltransduktion je nach Phosphomodus durch unterschiedliche Promoterarchitekturen erreicht wird. Ebenso hat der Phosphomodus Einfluss darauf, wie gut sich die Form des Input-Output-Verhältnisses durch Verändern der Promoterarchitektur einstellen lässt.

Des Weiteren wird am Beispiel des Quorum Sensing (QS) Signalpfads in *Bacillus subtilis* die Verarbeitung von zwei Signalen auf Genexpressionsebene untersucht. Betrachtet man nur die Architektur des Signalpfads, so konvergieren beide Signale, da sie den gleichen TF aktivieren. Mein Modell identifiziert jedoch Charakteristiken der Promoterarchitektur, die es erlauben, die Gewichtung der Signale zu steuern. Darauf aufbauend, wird die Bedeutung dieser Erkenntnis im Hinblick auf die Evolution der Promotoren wird diskutiert. Schließlich schlage ich auf Basis meiner Modellierungsergebnisse ein synthetisches "Multiplexing-System" zur Anwendung in der Experimentalbiologie vor, das mittels gezieltem Promoterdesign die Dekodierung der beiden QS-Signale erlauben und so Aufschluss auf deren natürliches Vorkommen geben soll.

Die Ergebnisse dieser Arbeit können die Interpretation von Sequenzdaten sowie Designfragen in der synthetischen Biologie unterstützen.

Contents

List of abbreviations

AA	amino acid
AI	autoinducer
B. subtilis	*Bacillus subtilis*
BS	binding site
DR	direct repeat
E. coli	*Escherichia coli*
GRF	gene regulatory function
HK	histidine kinase
ICE	integrative and conjugative element
IR	inverted repeat
EMSA	electrophoretic mobility shift assay
nt	nucleotide
PDR	perfect direct repeat
PIR	perfect inverted repeat
QS	quorum sensing
RE	recognition element
RNAP	RNA polymerase
RR	response regulator
TCS	two-component system
TF	transcription factor
V. harveyi	*Vibrio harveyi*
wt	wild type

Chapter 1

Introduction

1.1 Gene expression in bacteria

General questions on bacterial signal transduction

Bacteria adapt to their environment by rapidly changing the combination of genes they express: They respond to changes in nutrient availability by shifting their metabolism (Alberts et al., 2008); some species can sporulate upon starvation (Driks, 2002); depending on osmotic or pH conditions, they adapt their cell wall/membrane composition (Beales, 2004) and at high bacterial density, they can engage into cooperative behavior like swarming or virulence (Miller and Bassler, 2001). All these processes have in common that they are regulated by gene expression. However, the way in which environmental cues are processed depends on the task to be performed: When fine-tuning the intracellular composition of solutes in response to osmotic stress, accurate transduction is required (Csonka, 1989), while life style changes – for example entering a competent or sporulation state – require decisions (Veening et al., 2005; Leisner et al., 2008). For cooperative behavior, multiple signals from surrounding bacteria are integrated to make a decision (Waters and Bassler, 2006). Some "programs", for example competence and dormancy, also exclude each other, meaning that a logical operation is performed at some point. Since, compared to eukaryotes, bacterial signal transduction displays relatively low complexity, i.e. comparably low numbers of genes and often one-to-one mappings from gene to function, they are perfect model organisms to study some basic questions on biological signal processing: (1) What "processing task" can a system perform? Or, if treated in a more quantitative fashion, how good is a system at the respective task? And (2) what are the requirements for that? These will be the general questions of this thesis.

Basics of bacterial gene expression

The basic requirement for bacterial gene expression (review: Browning and Busby (2004)) is binding of RNA polymerase (RNAP) to an RNAP binding site (BS) in the promoter region of the

respective gene (Fig. 1.1 A). Bacterial RNAP BSs have a -10 element and a -35 element, named after their position with respect to the transcription start site. These elements are specific for σ factors, which are exchangeable RNAP subunits. For each σ factor, there is a perfect -35 and -10 element to which it binds with high affinity. However, natural promoters do not typically show such perfect binding interfaces, so that transcription requires further activators, which have BSs in the operator region of promoters.
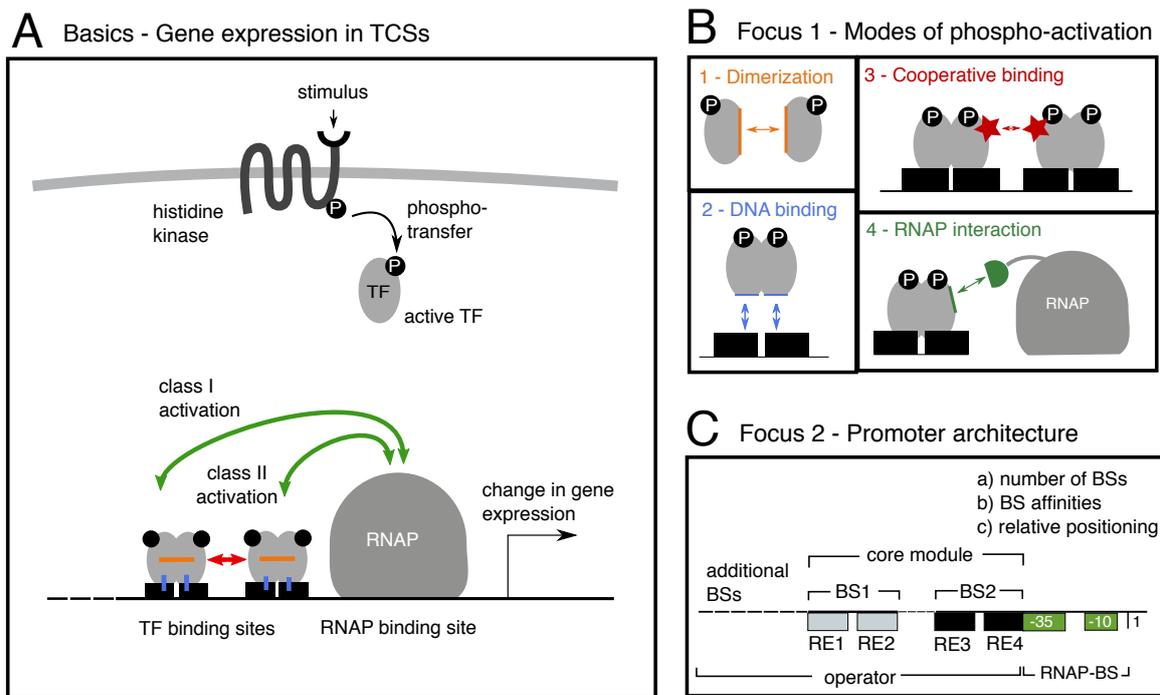
Figure 1.1: Overview: Gene expression in two-component-systems and the foci of this work. A: Signaling in two-component systems (TCSs) involves a histidine kinase (HK) which, upon environmental stimuli, transfers phosphate to a response regulator (RR). RRs are often transcription factors (TFs), which are activated by phosphorylation. As a consequence, the phosphorylation signal leads to changes in gene expression. **B:** Phospho-activation of a RR induces a conformational change, which makes an interaction interface accessible. This facilitates dimerization, cooperative binding, DNA binding or RNAP interaction. **C:** Promoters vary in their architecture. A core module of two TF BSs, each of which consists of two recognition elements (REs), is a frequent feature. Variations are found in the number of BSs, their affinities and their relative positioning.

Transcription factors (TFs) include transcriptional repressors and activators. Activation is necessary when the promoter is weak. Common types of transcriptional activation are (see also Fig. 1.1A, C):

- Class I: The activator binds upstream of the -35 element and interacts with the RNAP α subunit.

- Class II: The TF binds in a region that overlaps the -35 element and contacts the σ factor.

- Class III: Both class I and class II activation take place.

TF binding may also bend the DNA, a process called looping. By this mechanism, molecules that bind further upstream of the promoter can still establish contact with RNAP or other TFs and thus regulate transcription. Repressed genes usually display TF BSs within (or overlapping) the RNAP BS, so that they compete with RNAP for binding.

A further interaction is found between TFs: they frequently bind cooperatively to the DNA, which means that binding of one TF facilitates (or stabilizes) the binding of another TF to another site.

Signal transduction in two-component systems

A large proportion of a bacterium's genome is regulated by response regulators (RRs), which are part of two-component systems (TCSs) – the most common processing units for environmental signals including pH, temperature, nutrients, chemotactic signals and autoinducers (AIs) in bacteria. TCSs consist of a histidine kinase (HK), which is a transmembrane receptor for the respective signal and autophosphorylates upon activation, and a RR, which receives the phosphate group from its cognate HK and, in the activated phosphorylated form, triggers an intracellular response (Stock et al., 2000) (Fig. 1.1A). TCSs predominantly exist in bacteria, but also in archaea and eukaryotes, which makes the understanding of their signaling properties of universal importance. Within bacteria, they are (among) the major mediators of environmental signals of all kinds, making them a potential drug target (Gotoh et al., 2010). Due to their modular structure they further constitute a predestined target for recombination by synthetic biologists (Ninfa, 2010). Applications include coupling of HKs to non-cognate RRs, or engineering the promoters of target genes in the classic way (examples: Levskaya et al. (2005); Whitaker et al. (2012)). Because of the versatility of TCS, possibilities of combinations are numerous.

Even though RRs display various functionalities, including enzymatic, protein binding or RNA binding activity, most RRs are TFs (Gao et al., 2007). They comprise an N-terminal receiver domain which gets phosphorylated by the cognate HK and a C-terminal effector domain responsible for DNA binding (Stock et al., 2000). While RRs are classified into the OmpR, NarL and NtrC subclasses based on the sequence of their DNA binding domain (Albright et al., 1989; Mizuno, 1997), they reveal a high diversity with respect to functionality and structure within these families (Gao et al., 2007).

Modes of phospho-activation

I will argue in Ch. 3 that the interplay between promoter architecture (see below) and function of a gene cannot be studied independently of the way in which phosphorylation through HK activates the TF, which I call the mode of phospho-activation. Studies typically test whether a certain TCS activates a target gene by (1) *in vivo* experiments which show that the components

3

of the TCS are necessary for gene expression and (2) *in vitro* experiments which show that the RR binds to the promoter region of the gene and that the phosphorylated form does so more efficiently. Such observations can be due to different underlying mechanisms (Fig. 1.1B):

I) RRs appear to occupy BSs as dimers and evidence for phosphorylation-dependent dimerization has been found, for example, for NarL and OmpR in *Escherichia coli* (*E. coli*) (Katsir et al., 2015; Barbieri et al., 2013). Dimerization is thought to stabilize two TF monomers on neighboring parts of a BS (REs, explanation below) and to thereby support binding to the promoter region (Marianayagam et al., 2004).

II) Enhanced affinity of a TF could also be due to a conformational change that improves its DNA binding interface and thus protein-DNA interaction. For example, the DNA binding domain of NarL (*E. coli*) is precluded in the unphosphorylated state (Baikalov et al., 1996).

III) Some studies suggest phosphorylation-dependent cooperativity between TFs, presumably TF dimers (Sinha et al., 2008; Huang et al., 1997). In contrast to dimerization, which could occur in solution or on the DNA, cooperativity is generally thought to require DNA contact.

IV) Another option could be phosphorylation-dependent interaction between TF and RNAP (Boucher et al., 1997).

Despite the knowledge of various modes of phospho-activation it is still poorly understood how they affect gene regulation in the context of certain promoter architectures.

Promoter architecture

Interestingly, not only the RRs themselves, but also the promoter regions they bind to reveal a high degree of complexity. Promoter architectures can differ in a) the number of BSs they comprise, b) their affinities to TFs and c) their relative positioning (Fig. 1.1C).
Usually, the BSs identified by DNA protection assays comprise two short, symmetric repeats, often hexamers or heptamers, connected by a linker. The repeats are thought to be recognition elements (REs) to each of which a RR monomer can bind, so that a BS is bound by a RR dimer (Fig. 1.1C). In many regulons a common motif, which I call core region, for a particular RR is found to consist of two BSs. For example, the promoters regulated by PhoP and ComA regulons in *Bacillus subtilis* (*B. subtilis*) show a common core region of DR-DR (direct repeats) and IR-DR (inverted and direct repeat), respectively, while some of the genes have additional regulators upstream (Wolf et al., 2016; Liu et al., 1998; Qi and Hulett, 1998). But also one or three BSs are a frequent arrangement. For CpxR, which mediates the envelope stress response in *E. coli*, around 100 genes with putative promoters were found, of which most have one BS, but 25 have

two or (less frequently) three BSs (De Wulf et al., 2002). The promoters regulated by MprA in *Mycobacterium tuberculosis* have one to three BSs (He and Zahrt, 2005). A more diverse pattern is found for genes regulated by NarL (*E. coli*): This RR recognizes an IR called "7-2-7 site" with high affinity, but not all NarL-dependent promoters contain this BS (Darwin et al., 1997). Some, like P_{narG}, consist of clusters of heptamer sequences (putative REs) with different orientations (Dong et al., 1992; Li and Stewart, 1992; Tyson et al., 1993; Walker and DeMoss, 1994). Others, for instance P_{nrfA}, have both a 7-2-7 site and additional heptamer sequences (Tyson et al., 1994).

Promoters are – just as genes – subject to evolution (reviews: McAdams et al. (2004); Wray (2007)). BSs may accumulate or be lost through gene duplication or gene loss, while point mutations within BSs can fine-tune their affinity to TFs. Therefore, finding optimal promoter designs for certain "processing tasks" and comparing them with those found in a system of interest can give hints towards which task it has been selected for during evolution.

Nowadays, sequence data are available and facilitate the annotation of TF and RNAP BSs in the genome, revealing also the positioning among them, called promoter architecture (e.g. Collado-Vides et al. (1991)). Furthermore, BS strengths can be well estimated through binding assays or from sequence similarity within regulons (Stormo and Fields, 1998). Still, what remains unclear in many cases is the architecture's functional impact on gene regulation, meaning what behavior to expect from a promoter with a certain architecture (Guet et al., 2002). This question will be investigated both in the context of different phospho-modes (Ch. 3) and for a specific pathway in Chs. 4 and 5.

1.2 Signal convergence in bacterial quorum sensing

The signal processing system that I will consider in particular in Chs. 4 and 5 is that of quorum sensing (QS) in the soil bacterium *B. subtilis*. QS is a behavior in which bacteria estimate their density by producing, secreting and sensing of small molecules called autoinducers, or, in this work (for a more functional wording) QS signals (Fig. 1.2A). When the concentration of QS signals exceeds a certain threshold, bacteria react with changes in gene expression (Fig. 1.2B). These initiate behavioral programs which often involve the secretion of so-called exofactors – molecules that fulfill their function outside the cell. They include degradative enzymes for nutrient processing, virulence factors such as antibiotics, or surfactins. QS allows behaviors like nutrient digestion, virulence, collective swarming, genetic competence or biofilm formation (Fuqua et al., 1994; Miller and Bassler, 2001).

Many bacteria respond to more than one QS signal and these signals usually converge in a common bottleneck molecule which then regulates gene expression (Waters and Bassler, 2005) (Fig. 1.2C, D). In this thesis (Chs. 4 and 5) I will discuss the QS pathway in *B. subtilis*.

The QS master regulator ComA in *B. subtilis* is modified by two QS signals (Fig. 1.2D). They both eventually enhance gene expression: A species-specific signal releases ComA from a sequestering reaction and a strain-specific signal phosphorylates the TF.
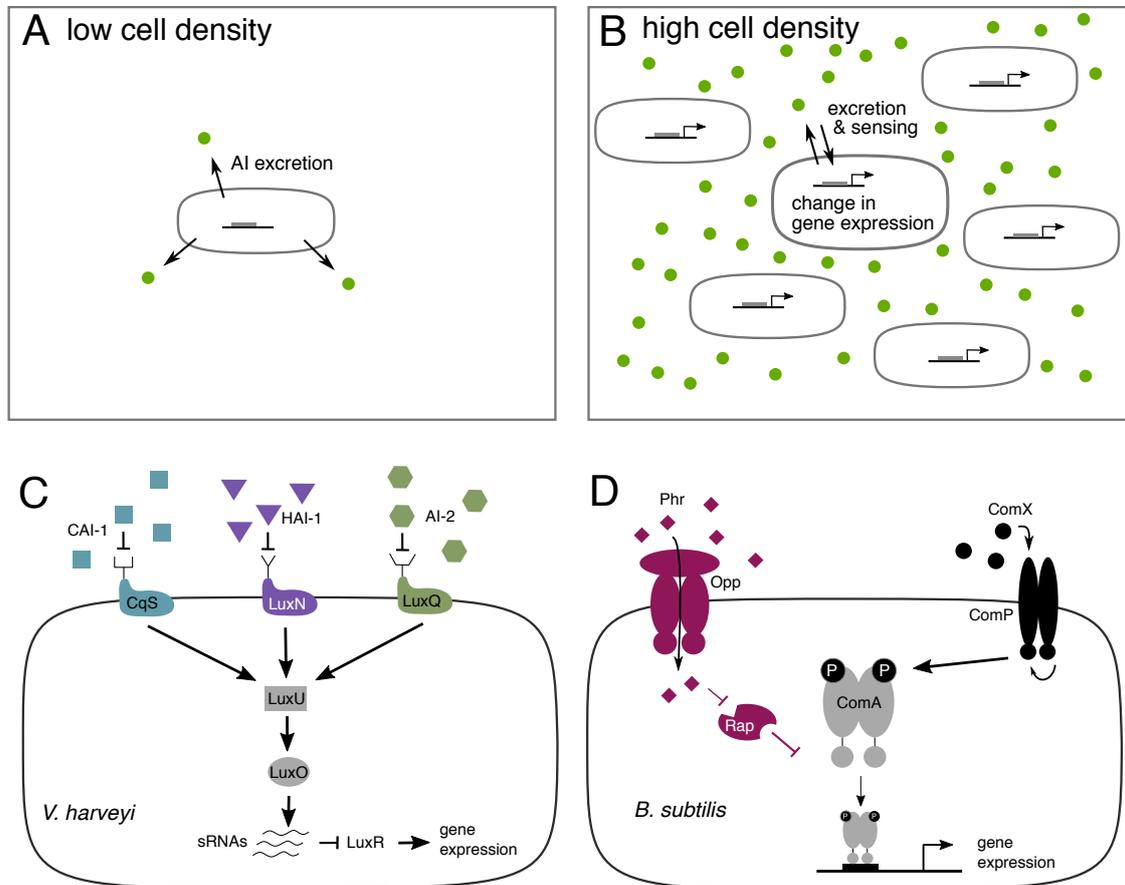

Figure 1.2: Signal convergence in bacterial quorum sensing. A-B: General concept of quorum sensing (QS). A: Bacteria produce, secrete and sense QS signals (AIs), which at low bacterial density are diluted through diffusion. B: At high bacterial densities, the signaling molecule accumulates, which is sensed and leads to changes in gene expression. **C-D**: Examples for QS systems in which different QS signals converge at a common molecule. C: In *V. harveyi* the signals CAI-1, HAI-1 and AI-2 all modulate the phosphorylation state of LuxU (Ng and Bassler, 2009). D: In *B. subtilis* the QS signals Phr and ComX regulate the sequestering and phosphorylation state of the RR ComA, respectively (details in the text).

Six Rap (short for RR aspartate phosphatases) proteins – RapC, D, F, H, G and K – inhibit the transcription of ComA-dependent genes *in vivo* (Solomon et al., 1996; Ogura and Fujita, 2007; Bongiorni et al., 2005; Hayashi et al., 2006; Smits et al., 2007; Auchtung et al., 2006). For at least 3 of them – RapC, F and H – the inhibition was already found *in vitro* to be due to a sequestering reaction of ComA (Core and Perego, 2003; Bongiorni et al., 2005; Smits et al., 2007). PhrC (better known as CSF), F G, H and K all counteract the negative effect of their cognate Rap protein on ComA (Solomon et al., 1996; Core and Perego, 2003; Bongiorni et al., 2005; Hayashi et al., 2006; Smits et al., 2007; Mirouze et al., 2011; Auchtung et al., 2006). Phrs (short for phosphatase regulators) are small peptides which are exported and cleaved by

signal peptidases to their effective size of mostly 5 amino acids (AAs) (reviews: Perego and Brannigan (2001); Pottathil and Lazazzera (2003)). They are imported by an ATP-dependent oligopeptide permease (Opp) (Solomon et al., 1996; Rudner et al., 1991; Perego et al., 1991). Intracellularly, they bind to a C-terminal tetratricopeptide repeat domain of Rap and prevent its binding to ComA by what seems to be an allosteric mechanism (Gallego del Sol and Marina, 2013).

The RR ComA and the sensor histidine kinase ComP form a TCS (Weinrauch et al., 1990). ComP is a transmembrane protein, which, upon activation by the QS molecule ComX, transfers a phospho-group to the N-terminal domain of ComA (Weinrauch et al., 1989, 1990; Piazza et al., 1999; Magnuson et al., 1994; Ansaldi et al., 2002). Phosphorylation of ComA is thought to activate its functionality as a TF by increasing its affinity to promoter regions (Weinrauch et al., 1990; Roggiani and Dubnau, 1993). However, the mechanism of this activation has not been resolved yet.

ComX is a strain-specific signal (Tran et al., 2000; Tortosa et al., 2001; Ansaldi et al., 2002), while the RapC/PhrC system is species-specific and highly conserved within *B. subtilis* strains (95% sequence identity) (Pottathil et al., 2008).

At first glance, it appears that in pathways with signal convergence, all QS signals should have the same effect and that the information about signal identity should be largely lost at the bottleneck. Indeed, experiments suggest that this is the case in QS in *Vibrio harveyi* (*V. harveyi*) (Long et al., 2009).

However, these networks still offer a few ways of increasing signal distinction and thus kin selection, on the level of pathway architecture and production rates (Mehta et al., 2009; Even-Tov et al., 2016) (see Sec. 1.3).

I will present in Chs. 4 and 5 another possible option which is easily overlooked in network diagrams as presented in Fig 1.2: the promoters of the QS-dependent genes. If QS-dependent genes have different promoters, then the signal processing might differ at this stage. In Ch. 4 I will explore features of promoter architecture that are likely to allow different weighting of the signals and in Ch. 5 I will present a theory for a multiplexing device that might be implemented experimentally in order to infer the distribution or dynamics of the natural QS signals from gene expression output.

1.3   Modeling approaches

Mathematical modeling can support experimental studies on gene expression, as it allows the formulation of hypotheses and the exploration of options in shorter time and without disturbance of factors like measuring noise, ambiguity in the set-up, or cross-talk from signaling pathways that are not subject of the study. Models can predict what processing task (a simplified version

of) the system of interest can perform in theory, and under which conditions. These predictions can then be compared with natural conditions or motivate experiments.

Modeling at thermodynamic equilibrium

Suitable methods for modeling gene expression depend on several factors, including the considered scale, the amount of knowledge on the system and the modeler's question (reviews: Hasty et al. (2001); De Jong (2002); Ay and Arnosti (2011)). Large regulatory networks spanning multiple TFs and genes with only coarse-grained knowledge on the regulation might be well captured with graph-based models, for example Boolean networks, where genes are simply described as ON or OFF (Shmulevich et al., 2002; Kauffman et al., 2003) or Bayesian networks Friedman et al. (2000). The Boolean approach is also found in the random telegraph model, where expression from only one gene is investigated, but complexity and computational cost arise from the stochastic nature of gene expression which the model aims to describe (Paulsson, 2005).

A common way to model smaller genetic circuits is by using differential equations, which describe the interdependencies of the involved molecules (e.g. TFs, DNA and mRNA) on the basis of chemical rate equations. Well-studied examples in bacteria include the *lac*-operon as well as the *trp*-operon (Mackey et al., 2004). One advantage of this technique is its ability to capture dynamics. Systems of differential equations can sometimes be solved analytically – usually following simplifications – and are solved numerically, otherwise.

For processes on the promoter-level, especially for describing the distribution of TFs on several BSs, so-called thermodynamic models can be a good choice. These models operate at thermodynamic equilibrium and estimate the probability of DNA states by using association constants (rather than rates) for describing interactions between DNA and proteins.

Pioneering approaches to this end come from Von Hippel et al. (1974) (*lac* operon) and Shea and Ackers (1985) (*lambda* phage operator). In order to compute the equilibrium probabilities of certain DNA configurations (e.g. states in which the DNA is occupied by a specific combination of TFs at specific sites), Shea and Ackers (1985) developed a method based on statistical mechanics. The method allows to assign weights to each state that depend on binding energies, which can be measured as association constants. The probability for a DNA state, also called fractional occupancy, is then calculated by dividing the state's weight by the sum of all weights. This method serves as the theoretical basis for many following modeling studies (Buchler et al., 2003; Bintu et al., 2005b; He et al., 2010; Ahsendorf et al., 2014), including this one. Apart from thermodynamic correctness, advantages of this approach are a comparably low number of parameters and the ability to choose the degree of model complexity (Ay and Arnosti, 2011): For instance, while Bintu et al. (2005b) presented a modeling framework that considers the dis-

tribution of TF along the genome, on both specific and unspecific sites, another framework by Ahsendorf et al. (2014) neglects this distribution (under certain assumptions) and thus allows the formulation of much simpler gene regulatory functions.

Promoter architecture

Statistical mechanics-based models are a good tool for comprehensive studies on promoter architecture. Buchler et al. (2003) applied the method from Shea and Ackers (1985) to implement hypothetical promoters capable of processing logical tasks. They varied BS affinities, interaction with RNAP and the ability of TFs to bind cooperatively, the latter two of which can be understood as features resulting from relative placement of TF and RNAP BSs. Setty et al. (2003) constructed a model of the *lac* operon and fitted it to data. They then made predictions on how certain logical operations on the inputs glucose and lactose could be improved by adjusting BS affinities. Estrada et al. (2016) investigated the relation between the number of TF BSs within a promoter and the steepness that the response curve can take.

TCS-dependent gene expression

For TCSs in particular, the phosphorylation of RRs has been well studied in theory and experiment, with respect to robustness of the concentration of phosphorylated RR to changes in HK or ATP (phosphate) concentrations (Batchelor and Goulian, 2003; Shinar et al., 2007), but also concerning the qualitative response behavior (response shape) of phosphorylated RR to environmental signals (Igoshin et al., 2008). However, modeling studies that resolve TCS-specific processes on the promoter level, especially regarding the different phospho-modes, are lacking.

Quorum sensing

The topic of converging QS signals has been addressed by both experimental and modeling studies.

In a modeling study on *V. harveyi* QS Mehta et al. (2009) suggested that certain constraints on the QS signal production rates (i.e. input signal distribution) could increase the amount of information on the signal contained in the regulator LuxO. A biologically plausible constraint could be, for example, that individual bacteria secrete all QS signals at the same rate. Since not all bacteria are capable of producing the highly specific signal (AI-1), the concentration of less specific signal (AI-2) in the environment would always be higher than that of the specific one. Therefore, bacteria could distinguish between low-AI-1/high-AI-2 and high-AI1/low-AI2: the former would not be possible and could thus be excluded.

Even-Tov et al. (2016) focused on the impact of signaling pathway architecture on kin selection. They pointed out an important difference between positive regulation of a QS effector by QS signals and release from inhibition. For instance, in *B. subtilis* ComA is activated by the QS

signal ComX via phosphorylation and released from a sequestering reaction by Phr peptides (Fig. 1.2A). As the authors show by both modeling and experiments, addition of more specific ComX/ComP pathways decreases the bacterium's fitness when co-cultured with wild type (wt). This is because the additional pathway makes the bacterium more responsive to QS signals, but not exclusively so for those of its kins. On the other hand, introducing an extra Rap/Phr system gives bacteria a competitive advantage in co-culture with the wt. In this case, the extra Rap serves as an additional safety lock that can only be unlocked by the extra cognate Phr. Bacteria that have the new, specific Rap/Phr pathway only respond (and produce common goods) when they are surrounded by clones that produce the extra Phr. When surrounded by wt bacteria, they produce signal but do not respond to wt Phr. This cheating mechanism therefore increases fitness by increasing kin selection.

Multiplexing

Multiplexing means a process in which several signals are pooled, transmitted over a common medium and then decoded. Decoding of input signals that are modulated on the same TF has been addressed in an information-theoretic analysis of experimental data in yeast (Hansen and O'Shea, 2015), where the discrimination is based on amplitude and frequency of the TF dynamics. Further, there is one modeling study on a hypothetical multiplexing system (de Ronde and ten Wolde, 2014), in which the two signals are also encoded via amplitude and frequency. However, this system does not include gene expression. In conclusion, I find the theoretical contemplation of multiplexing in biological systems as a largely unexplored field.

1.4   Scope of the thesis

In this thesis I use thermodynamic models to study steady state gene expression in TCSs in response to one or more environmental signals. The main question is how promoter architecture can help to fulfill a certain processing task. I will look at three such tasks:
In Ch. 3 I set up four comprehensive models for signal transduction in TCSs on the gene expression level. I use these models to compare four modes of phospho-activation with respect to their requirements on promoter architecture for effective signal transduction.
In Ch. 4 I apply one of the models from Ch. 3 to gene expression in *B. subtilis* QS in order to investigate the weighting between two QS signals.
In Ch. 5 I study the same QS pathway, but with respect to the question whether it can be also used to multiplex the two QS signals.
I will conclude by comparing the tasks, with respect to their requirements in promoter architecture, and discussing the findings regarding their application to sequence data analysis and synthetic biology.

1.5 Novelties of the thesis

- *Gene activation:* While many gene expression studies in bacteria deal with gene repression and negative feedback, the focus of this thesis will be on gene activation. I will consider the processing tasks of transduction of a single signal as well as weighting between and multiplexing of two signals.

- *Phospho-modes:* In this work on TCSs, I will explicitly distinguish between different possible ways in which phosphorylation of the RR can enhance gene expression (phospho-modes). So far, these have not been considered in modeling studies. Cooperativity, which is one of the phospho-modes I investigate, has been studied as a general feature of TF binding, but not as a phosphorylation-dependent one.

- *Promoter architecture in QS*: This work's focus on promoter architecture is new in its application to the processing of QS signals. Previous studies have focused on signaling pathway architecture (Even-Tov et al., 2016) or signal distribution (i.e. production rates) (Mehta et al., 2009).

- *Idea for novel device:* I propose a synthetic multiplexing system for experimental purposes based on the *B. subtilis* QS pathway. This system could constitute a novel device to learn about the natural distribution of QS signals.

- *Non-frequency multiplexing:* The proposed multiplexing device is different from previous work, as it is based on two independent modifications on the same TF, to be decoded employing knowledge on promoter architecture. Previous studies on multiplexing in living cells, of both practical and theoretical nature, investigate frequency multiplexing, where the system's dynamics are crucial for decoding.

- *Theory for multiplexing in bacteria:* Multiplexing has not yet been observed nor hypothesized in bacteria. The theoretical model for a multiplexing device could thus inform the search for such processing.

Chapter 2

Methods

2.1 Application of the Wegscheider condition

In Sec. 3.2, the Wegscheider condition (Wegscheider, 1901) is applied on reaction cycles in order to determine detailed balance conditions for model 1. In this model, a DNA state with a dimer bound can be reached either by TF dimerization in solution followed by dimer binding or by sequential monomer binding followed by dimerization on the DNA. The corresponding reaction cycle is shown in Fig. 3.3A. The application of the Wegscheider condition is explained in the following.

For detailed balance, the following equations must hold:

$$p_{\{x,x\}} \cdot k_{RE1}^+ \cdot [T^p] = p_{\{p,x\}} \cdot k_{RE1}^- \qquad \Leftrightarrow \frac{p_{\{x,x\}}}{p_{\{t,x\}}} = \frac{k_{RE1}^-}{k_{RE1}^+} \cdot \frac{1}{[T^p]} \qquad (2.1a)$$

$$p_{\{p,x\}} \cdot k_{RE2}^+ \cdot [T^p] = p_{\{p,p\}} \cdot k_{RE2}^- \qquad \Leftrightarrow \frac{p_{\{p,x\}}}{p_{\{p,p\}}} = \frac{k_{RE2}^-}{k_{RE2}^+} \cdot \frac{1}{[T^p]} \qquad (2.1b)$$

$$p_{\{p,p\}} \cdot v^+ = p_{\{p-p\}} \cdot v^- \qquad \Leftrightarrow \frac{p_{\{p,p\}}}{p_{\{p-p\}}} = \frac{v^-}{v^+} \qquad (2.1c)$$

$$p_{\{p-p\}} \cdot k_{BS}^- = p_{\{x,x\}} \cdot k_{BS}^+ \cdot [T_2^p] \qquad \Leftrightarrow \frac{p_{\{p-p\}}}{p_{\{x,x\}}} = \frac{k_{BS}^+}{k_{BS}^-} \cdot [T_2^p] \qquad (2.1d)$$

$$[T^p]^2 \cdot k_{dim}^+ = [T_2^p] \cdot k_{dim}^- \qquad \Leftrightarrow [T_2^p] = \frac{k_{dim}^+}{k_{dim}^-} \cdot [T^p]^2 \qquad (2.1e)$$

p_i indicates the probability of state i. The Wegscheider condition (Wegscheider, 1901) then states that

$$\frac{p_{\{x,x\}}}{p_{\{t,x\}}} \cdot \frac{p_{\{t,x\}}}{p_{\{t,t\}}} \cdot \frac{p_{\{t,t\}}}{p_{\{t-t\}}} \cdot \frac{p_{\{t-t\}}}{p_{\{x,x\}}} = 1. \qquad (2.2)$$

Substitutions:

$$K_{RE1} = \frac{k_{RE1}^+}{k_{RE1}^-}, \qquad K_{RE2} = \frac{k_{RE2}^+}{k_{RE2}^-}, \qquad K_{BS} = \frac{k_{BS}^+}{k_{BS}^-}, \qquad K_{dim} = \frac{k_{dim}^+}{k_{dim}^-}, \qquad v = \frac{v^+}{v^-} \quad (2.3)$$

By inserting Eqs. 2.1 into the cycle condition (Eq. 2.2) and applying above substitutions (Eq. 2.3), we get the thermodynamic constraint

$$K_{BS} = \frac{K_{RE1} \cdot K_{RE2} \cdot v}{K_{dim}}. \tag{2.4}$$

Such a thermodynamic constraint is given for each BS. More generally speaking, the K_A for BSj is given by

$$K_{BSj} = \frac{K_{REj} \cdot K_{REj+1} \cdot v_j}{K_{dim}}. \tag{2.5}$$

For deriving the need for cooperative dimerization (Fig. 3.3B), the conditions for detailed balance can be found using the same solution process as above.

2.2   Model implementation

All models were implemented in Copasi (Hoops et al., 2006) (version 4.22, build 184) using R, version 1.0.136 (R Core Team, 2017), and the package CoRC (Foerster and Pahle, 2018) (version 0.3.0) as an interface between the two softwares.

Time courses / steady state results

To calculate steady states, I used Copasi deterministic time course computation using the standard setting "LSODA", with a duration of 5000 and interval size 100. To make sure that the system had reached steady state, I then compared the last two values of the time course (in R) and implemented a stop criterion in case that these values differed by more than 1%. This happened in none of my simulations.

2.3   Parameters

Model standard parameters are given in Tab. 2.1. These parameters were used in all simulations, unless stated otherwise.

Parameter	Value
$[DNA]$	1 nM
$[T_2]$	100 nM
K_{dim}	$[T_2]$
$[T_2] \cdot K_{BS}$	1
ρ	10
μ	10
ν	10
α	1

2.4 Analytic derivation of gene regulatory functions

For the analytic derivation of gene regulatory functions (GRFs), a method by Ahsendorf et al. (2014) was applied and then the resulting terms were further simplified. The method includes representing the model as a graph, calculating steady state probabilities of the promoter states and summarizing these into a function that returns the average concentration of gene output. The underlying assumption is that DNA binding does not significantly alter the concentration of free TF, so that, effectively, reaction rates are linearized.

Graph representation of DNA binding module

The DNA binding reaction networks of the models in this thesis can be represented as graphs where the nodes are promoter states and the edge labels are reaction rates (Fig. 2.1).
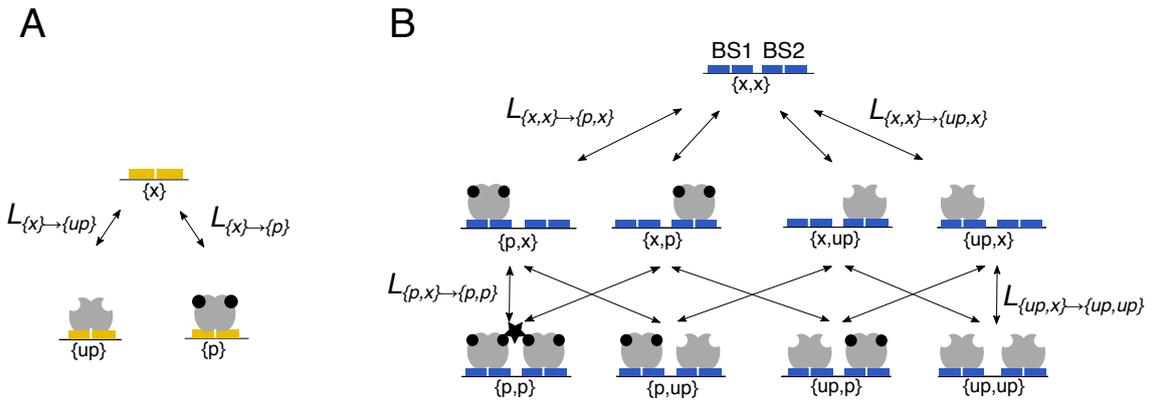

Figure 2.1: Graph representation. DNA binding reactions can be represented in a graph, shown for model 3a (**A**) and model 3 (**B**), which are introduced in Sec. 3.2.

The notation of DNA states is the same as in Sec. 3.2: The set of possible states is called S. The notation for a specific state is $i = \{a,b\}$ where $a \in \{up, p, x\}$ indicates whether BS1 is bound by

the TF species T_{us}^{up} ($a = up$) or T_{us}^{p} ($a = p$), or unbound ($a = x$). $b \in \{up, p, x\}$ indicates whether BS2 is bound.

The binding and unbinding reactions constitute the graph edges with edge labels l^+ and l^-:

$$l_{A \to B}^+ = k_{A \to B}^+ \cdot \rho_{A \to B}^+ \cdot [T_{us}^x] \tag{2.6}$$

$$l_{A \to B}^- = k_{A \to B}^- \cdot \rho_{A \to B}^- \tag{2.7}$$

In this notation $A \to B$ are binding reactions that lead from state A to state B with $A, B \in S$, k^+ and k^- are binding and unbinding rates, modulated by the cooperativity factors ρ^+ and ρ^-, and $[T_{us}^x] \in [T_{us}^{up}], [T_{us}^p]$ gives the concentration of the TF that binds during the binding reaction. The formulation of edge labels contains the above mentioned assumption of constant TF concentration.

Formalism for steady state

For calculations in steady state, the ratios between binding and unbinding reactions,

$$L_{A \to B} = \frac{l_{A \to B}^+}{l_{A \to B}^-}, \tag{2.8}$$

are needed, so that, effectively, only ratios between forward and reverse rates are relevant. These are the association constants and cooperativity factors

$$K_{A \to B} = \frac{k_{A \to B}^+}{k_{A \to B}^-}, \qquad \text{and} \qquad \rho_{A \to B} = \frac{\rho_{A \to B}^+}{\rho_{A \to B}^-}. \tag{2.9}$$

Label ratios in steady state can therefore be expressed as

$$L_{A \to B} = K_{A \to B} \cdot \rho_{A \to B} \cdot [T_{us}^x]. \tag{2.10}$$

Weights in steady state

According to Ahsendorf et al. (2014), for each promoter state i, a weight w_i is constructed. The reference state is the empty DNA which has the weight $w_1 = 1$. For every other state, a path from the empty state is chosen. To obtain the weight, the label ratios along the path are multiplied.

The steady state weights for the promoter states in gene regulatory systems C (2-BS promoter with the expression output C) and Y (1-BS promoter with the expression output Y), used in Chs. 4 and 5 were calculated from the graphs given in Fig. 2.1 and are listed in Tab. 2.2.

Table 2.2: Steady state weights. With empty promoter as reference state.

promoter C	
State	**Steady state weight**
$\{x,x\}$	1
$\{x,p\}$	$K_2 \cdot [T_{us}^p]$
$\{x,up\}$	$K_2 \cdot [T_{us}^{up}]$
$\{p,x\}$	$K_1 \cdot [T_{us}^p]$
$\{up,x\}$	$K_1 \cdot [T_{us}^{up}]$
$\{up,up\}$	$K_1 \cdot K_2 \cdot [T_{us}^{up}]^2$
$\{up,p\}$	$K_1 \cdot K_2 \cdot [T_{us}^{up}] \cdot [T_{us}^p]$
$\{p,up\}$	$K_1 \cdot K_2 \cdot [T_{us}^{up}] \cdot [T_{us}^p]$
$\{p,p\}$	$\rho \cdot K_1 K_2 \cdot [T_{us}^p]^2$

promoter Y	
State	**Weight**
$\{x\}$	1
$\{p\}$	$K_y \cdot [T_{us}^p]$
$\{up\}$	$K_y \cdot [T_{us}^{up}]$

Probabilities of promoter states

The steady state probability of any state i is given by normalizing the steady state weights (w) to the total amount $Z = \sum_i w_i$:

$$p(i) = \frac{w_i}{\sum_i w_i}. \tag{2.11}$$

For the models used here, these totals are:

$$
\begin{aligned}
Z_c =\ & 1 + K_1 \cdot ([T_{us}^{up}] + [T_{us}^p]) \\
& + K_2 \cdot ([T_{us}^{up}] + [T_{us}^p]) \\
& + K_1 K_2 \left([T_{us}^{up}]^2 + 2[T_{us}^{up}][T_{us}^p] + \rho[T_{us}^p]^2\right)
\end{aligned}
\tag{2.12}
$$

$$Z_y = 1 + K_y \left([T_{us}^{up}] + [T_{us}^p]\right)$$

Gene regulatory functions

GRFs can then generally be expressed as

$$GRF = \sum_i p(i) \cdot \alpha_i \tag{2.13}$$

The GRFs initially derived from applying Eq. 2.13 to the weights in Tab. 2.2 are too bulky to name them here, and hardly insightful. In order to express the GRFs in terms of the two signals, f_{us} and f^p, and to simplify the functions for good comprehensibility, the following substitutions

were applied:

$$[T_{us}^p] = f_{us} \cdot f^p \cdot T$$
$$[T_{us}^{up}] = f_{us} \cdot (1 - f^p) \cdot T$$
$$c_1 = T \cdot K_1$$
$$c_2 = [T_2] \cdot K_2$$
$$c_y = [T_2] \cdot K_y$$

This simplifies the GRFs to

$$[Y] = \frac{f_{us} \cdot c_y}{1 + f_{us} \cdot c_y} \alpha_y \qquad (2.14)$$

and

$$[C] = \frac{\alpha_{\{t,x\}} \cdot f_{us} \cdot c_1 + \alpha_{\{x,t\}} \cdot f_{us} \cdot c_2 + \alpha_{\{t,t\}} \cdot f_{us}^2 \cdot c_1 \cdot c_2 \cdot [1 + (\rho - 1)(f^p)^2]}{1 + f_{us}(c_1 + c_2) + f_{us}^2 \cdot c_1 \cdot c_2 \cdot [1 + (\rho - 1)(f^p)^2]}. \qquad (2.15)$$

2.5 Taylor expansion of GRFs

2D Taylor expansion of GRFs around 0 was performed in Mathematica (Wolfram Research, 2018), using the function "SeriesCoefficient". SeriesCoefficient was applied six times to Eq. 5.4, to cover all combinations of c_1 and c_2 up to second order.

2.6 Analytic solution of f^p and f_{us} with Mathematica

To derive f_{us} and f^p from GRFs (Ch. 5), the GRFs were solved for the two variables. GRF_Y can be directly solved for f_{us}. GRF_C can be solved for f^p after replacing f_{us} by its analytical solution. The latter calculation was performed using the "Solve" function in Mathematica (Wolfram Research, 2018).

Chapter 3

Modes of phospho-activation in bacterial two-component systems

This chapter is based on a draft (from 9 September 2018) for a yet unsubmitted manuscript, which was written by myself.

3.1 Introduction

For TCSs, various modes of phospho-activation, i.e. ways in which phosphorylation through HKs can activate TFs, are known. Tab. 3.1 gives an overview of modes of phosphorylation that have been proposed in the literature (see also Sec. 1.1).

Table 3.1: Suggested modes of phosphorylation-induced gene expression in two-component systems. The mode gives the presumably phosphorylation-dependent process or quantity.

Mode	Examples	References
Dimerization	NarL (*E. coli*)	Katsir et al. (2015)
	PhoP (*S. enterica*)	Lejona et al. (2004)
	OmpR (*E. coli*)	Gao et al. (2008); Barbieri et al. (2013)
Enhanced affinity	NarL (*E. coli*)	Baikalov et al. (1996)
	BvgA (*B. pertussis*)	Boucher et al. (1997)
Cooperativity	OmpR (*E. coli*)	Huang et al. (1997)
	PhoP (*M. tuberculosis*)	Sinha et al. (2008)
	PhoP (*B. subtilis*)	Liu et al. (1998)
RNAP interaction	BvgA (*B. pertussis*)	Boucher et al. (1997)

Here I apply simple, thermodynamically correct equilibrium models, which are common for studies on promoter architecture (see Sec. 1.3), to answer the basic question, which behavior to expect from gene activation systems depending on their phospho-mode, class of activation

(expression scheme) and promoter architecture. I find that (1) different phospho-modes require distinct promoter architectures and expression schemes for effective signal transduction and that (2) response shapes and their sensitivity to parameter changes also depend on the phospho-mode.

Finally, I will discuss how the findings of this chapter might support the interpretation of experimental results, but also of promoter sequence data, and how they can be of use in the field of synthetic biology where specific purposes often require finely tuned input-output relations.

3.2 Models representing four modes of phosphorylation

I use thermodynamically correct models in steady state, which comes at the advantage of low complexity, a small parameter set and general validity. In this chapter I consider gene activation at a generic 2-BS promoter, which displays the minimum number of BSs that allows for cooperativity and different gene expression logics. In principle, all described models from this chapter can be easily reduced to one BS or extended to more BSs (this is done in Chs. 4 and 5). The 2-BS promoter consists of four recognition elements (REs) upstream of the RNAP BS, termed RE1-4 (Fig. 3.1). These could be grouped into two neighboring BSs, but in theory, also one BS could be formed from the two middle REs, RE2 and RE3. I consider four possible features of expression, corresponding to distinct phospho-modes. For each of these, I implemented a model in which this feature is phosphorylation-dependent.
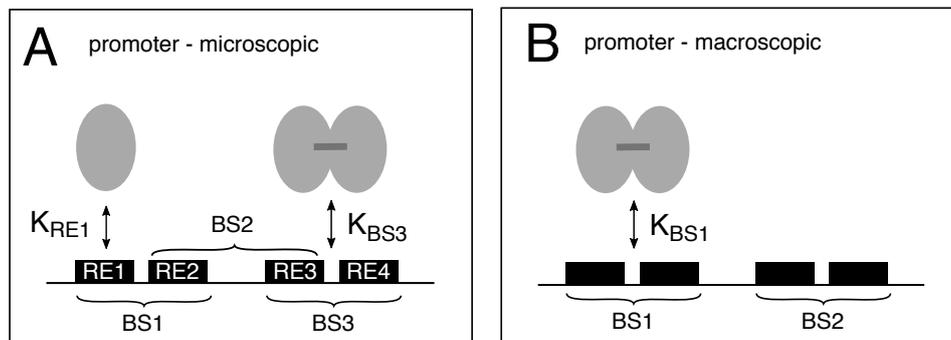

Figure 3.1: Promoter used in the models. A: Microscopic promoter representation as used in model 1, where REs are resolved. **B:** Macroscopic promoter representation as used in models 2-4, where only BSs are considered.

All models describe (1) interactions between TFs and DNA, coupled to (2) a simple scheme for gene expression. They were implemented with a common framework and set of parameters, in order to be as comparable as possible. Reactions were simulated with mass action kinetics and steady state results were evaluated (for details, see Sec. 3.3).

Phosphorylation signal.   I assume a constant overall TF dimer concentration $[T_2]$ on the time-scale of interest, since HKs change the phosphorylation state of TFs but not their number. I define the phosphorylation signal as the fraction of phosphorylated TF,

$$f^p = \frac{[T_2^p]_{t=0} + 0.5[T^p]_{t=0}}{[T_2]},\tag{3.1}$$

where T^p and T_2^p denote phosphorylated monomers and dimers, respectively. Unphosphory-lated TF is indicated by superscript *up*. The signal is not changed through reactions, but an initial value is set for each simulation. The implicit assumption behind this setup is that environmental changes are slower than changes in gene expression. Note that, since $[T_2]$ is constant, f^p is proportional to $[T^p]$ and $[T_2^p]$, but I chose to express the signal as a fraction rather than an absolute concentrations, because it is bounded between 0 and 1.

The four modes.   I implemented four different modes of phospho-activation using the following assumptions:

I) Phosphorylation-dependent dimerization: Phosphorylation of both involved monomers allows dimerization in solution or on the DNA.

II) Phosphorylation-dependent affinity: Phosphorylation increases the affinity between TF and DNA (independently of whether another BS is already bound).

III) Phosphorylation-dependent cooperativity: Phosphorylation increases TF-DNA affinity when two (neighboring) dimers bind successively.

IV) Phosphorylation-dependent gene expression: Only phosphorylated TFs bound to DNA can induce gene expression.

For every mode, I constructed a model in which this mode is phosphorylation-dependent while the other modes are not.

Since the models were evaluated in steady state, effectively only ratios between forward and reverse rates – here consistently given as K_A values – are relevant. Formally, however, all reactions are assigned a forward and a reverse rate which are indicated by superscript $+$ and $-$.

Model 1: Dimerization.   To study the effects of (phosphorylation-dependent) dimerization, I built a microscopic model in which the four REs of two BSs as well as monomers are resolved (Fig. 3.1A). In contrast, macroscopic models, which describe the system on the level of BSs and dimer binding, were applied for models 2-4 (Fig. 3.1B).
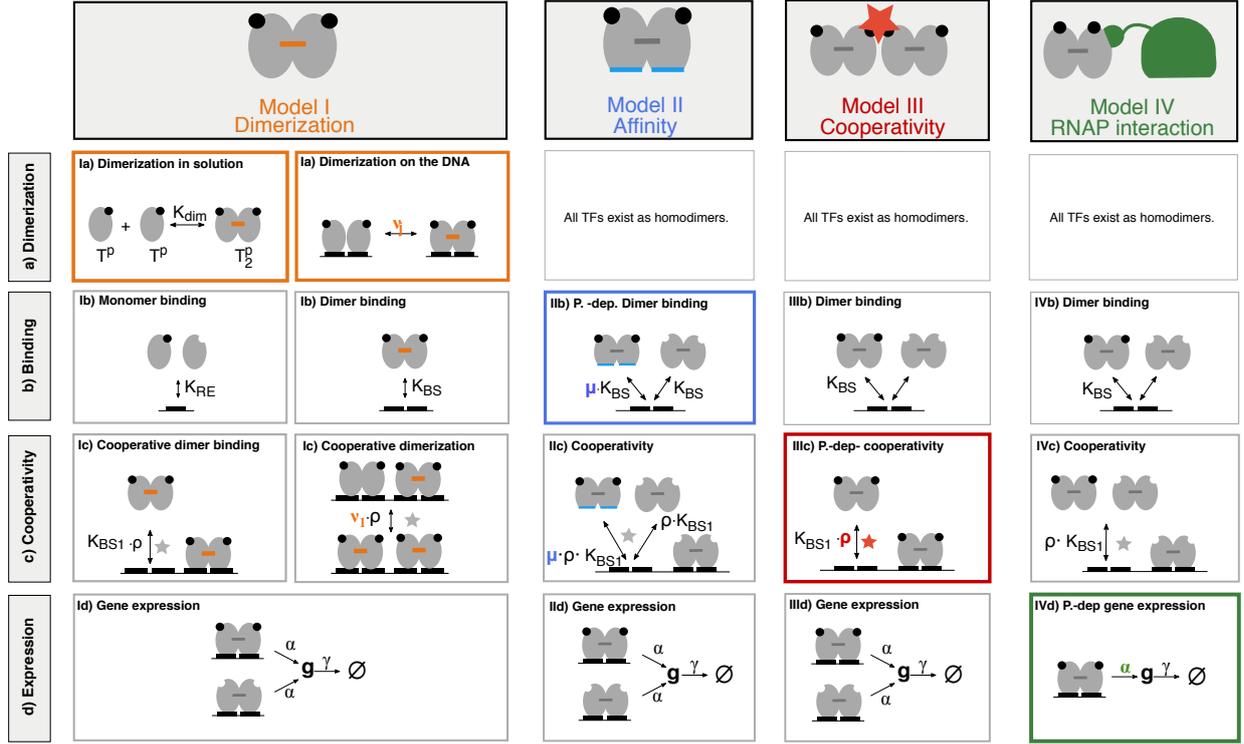
Figure 3.2: Models overview. Panels depict rules for the four models. Details in the main text. In all models, the phosphorylation state of TFs is an input signal. Within one simulation, it remains fixed.

In model 1, only phosphorylated TF can dimerize, either in solution or on the DNA. In solution, the dimerization constant $K_{dim} = k_{dim}^+/k_{dim}^-$ determines the equilibrium ratio between dimeric and monomeric T^p ($[T_2^p]_{eq}/[T^p]_{eq}^2 = K_{dim}$) (Fig. 3.2Ia). On the DNA, dimerization can occur between two phosphorylated monomers on neighboring REs (Fig. 3.2Ia). The dimerization strength on the DNA is modeled position-specific, in order to account for varying distances between REs or different RE types, which are found in natural sequences and are likely to affect the stability of dimers. For N_{RE} REs, dimer positions ($j \in 1, .., N_{RE} - 1$) are assigned dimerization constants $v_j = v_j^+/v_j^-$ which give the ratios between dimeric and monomeric T^p on RE_j and RE_{j+1}.

Even though both K_{dim} and v describe dimerization, they are independent model parameters. An important quantity is $\hat{v}_j = v_j/([T_2] \cdot K_{dim})$, which tells us how much more stable dimerization is on the DNA than in solution.

The DNA can be in different states and each state is a model species. In model 1, DNA states are named after the states of their REs, ordered from distal to proximal with respect to the RNAP binding site. REs can be empty (x) or occupied by a phosphorylated or unphosphorylated TF monomer (p or up). For example, a state named $\{up, x, p, p\}$ is occupied by an unphosphorylated monomer on RE1, and two phosphorylated monomers on RE3 and RE4. If two monomers form a dimer on the DNA, then this is indicated by a line between the two RE states, and the above state reads $\{up, x, p\text{-}p\}$.

Monomers can bind to empty REs with on and off-rates that depend on the RE's position, but not on the monomer's phosphorylation state (Fig. 3.2Ib). Also, the remaining REs' states do not influence the binding rates.

Any two neighboring REs make a potential BS for a TF dimer (Fig. 3.2Ib). BSs have the same index as their first (i.e. promoter-distal) RE, so that BS_j consists of RE_j and RE_{j+1}. Accordingly, a promoter with four REs has three potential BSs. Thermodynamic correctness (explanation below) requires the association constant K_{BSj} to be

$$K_{BSj} = \frac{K_{REj} \cdot K_{REj+1} \cdot \nu_j}{K_{dim}}. \tag{3.2}$$

For a derivation of this relation, see Sec. 2.1.

In model 1, binding of monomers is independent of phosphorylation, but dimer binding is indirectly dependent on phosphorylation, since only T^p can dimerize.

Similarly, cooperativity is formally phosphorylation-independent: Whenever any dimer is already bound to one of the BSs, binding of another dimer to the remaining BS occurs with binding constant $K_{BS} \cdot \rho$, where ρ denotes the cooperativity factor (Fig. 3.2Ic). However, since all dimers are phosphorylated, cooperativity is also indirectly dependent on phosphorylation.

For thermodynamic correctness, cooperative dimerization must be possible (Sec. 2.1). That means, dimerization needs to be modified by the cooperativity factor ρ in the special case when the resulting state is $\{p\text{-}p, p\text{-}p\}$, a promoter occupied with two dimers (Fig. 3.2Ic).

Every DNA state i is assigned a gene expression rate: when a DNA molecule is in state i, then gene product g is expressed with rate α_i^+. g is degraded globally with rate γ (Fig. 3.2Id). Thus, the expression constant $\alpha_i = \alpha_i^+/\gamma$ gives the equilibrium concentration of g that would be reached if the DNA was constantly in state i.

Since little is known about which BSs have to be occupied for gene expression to take place, i.e. how the expression constants should be distributed amongst the possible DNA states, I will look at three generic cases:

»any«	$\alpha_{\{t,x\}} = \alpha_{\{x,t\}} = \alpha_{\{t,t\}} = \alpha$	Any BS bound leads to gene expression.
»proximal«	$\alpha_{\{x,t\}} = \alpha_{\{t,t\}} = \alpha$	For expression, the proximal BS has to be bound.
»all«	$\alpha_{\{t,t\}} = \alpha$	For expression, both BSs have to be bound.

All other expression constants are $\alpha_i = 0$. Here, states are given in a macroscopic description (meaning BS and not RE states) and t is a placeholder for up or p, that is any state with a TF dimer bound. For model 1, t is meant to include all cases where both REs of the respective BS

are bound, irrespective of phosphorylation or dimerization state. For simplicity, I set $\alpha = 1$, so that gene product concentrations will be between 0 and 1. Referring to the classes of gene activation which are commonly used to describe TF-RNAP interaction (Sec. 1.1), scheme »all« can be interpreted as class III, scheme »proximal« as class II and scheme »any« as class I or II activation. The schemes can also be seen as a feature of promoter architecture, since the activation class is thought to be dependent on the relative positioning of TF and RNAP BSs.

Model 2: Phosphorylation-dependent affinity. We will see in Sec. 3.5 that many of the posed questions can also be answered with a macroscopic description of the model, which describes only dimers and BSs and which I use for models 2-4. In these, I do not model dimerization, but rather assume that all TFs exist as homodimers. Further, only the two non-overlapping BSs are considered. Because of the abstraction, DNA states are named after their BS states rather than after their RE states. Thus, BSs can be empty (x) or occupied by a phosphorylated or unphosphorylated TF homodimer (p or up).

In model 2, DNA binding affinity is explicitly enhanced by phosphorylation. In case of a phosphorylated dimer binding to the DNA, any association constant K_{BSj} is increased by a binding factor μ (Fig. 3.2IIb).

Cooperative dimer binding takes place between any two TF homodimers that successively bind to DNA (Fig. 3.2IIc). In case that phosphorylated TF takes part in cooperative interactions, both cooperativity and binding factor modify the association constant. Gene expression is independent of phosphorylation, with rules identical to those in model 1 (Fig. 3.2IId).

Model 3: Phosphorylation-dependent cooperativity. In model 3, DNA binding is independent of phosphorylation (Fig. 3.2IIIb), but for cooperative binding, both involved TF dimers have to be phosphorylated (Fig. 3.2IIIc). Gene expression rules are phosphorylation-independent and identical to those in model 1 and 2.

Model 4: Phosphorylation-dependent RNAP interaction. In my models, RNAP interaction does not interfere with TF binding, so that model 4 is completely phosphorylation-independent on the DNA binding level. Neither affinity nor cooperativity are affected by phosphorylation (Fig. 3.2IVb,IVc).

Phosphorylation-dependent RNAP interaction is implemented with the same three cases as for model 1-3, but only BSs occupied by phosphorylated TF lead to expression:

»any«	$\alpha_{\{p,x\}} = \alpha_{\{x,p\}} = \alpha_{\{p,p\}} = \alpha$	Any BS bound by T^p leads to gene expression.
»proximal«	$\alpha_{\{x,p\}} = \alpha_{\{p,p\}} = \alpha$	The proximal BS has to be bound by T^p.
»all«	$\alpha_{\{p,p\}} = \alpha$	Both BSs have to be bound by T^p.

All other expression constants are $\alpha_i = 0$. The standard parameter for expressing DNA states is set to $\alpha = 1$.

Initial concentrations. Initially, the concentration of empty DNA is 1, all other DNA concentrations are 0. TF concentrations are distributed according to the input signal:

microscopic	macroscopic
$[T^p] = f^p \cdot 2[T_2]$	$[T_2^p] = f^p \cdot [T_2]$
$[T^{up}] = (1 - f^p) \cdot 2[T_2]$	$[T_2^{up}] = (1 - f^p) \cdot [T_2]$

In model 1, which allows dimerization, $[T_2^p]_{t=0} = 0$, so that initially all TF is monomeric. Gene product does not exist initially ($[g]_{t=0} = 0$).

Mass conservation. The overall concentration of DNA does not change:

$$\sum_i [i] = 1. \tag{3.3}$$

Also, the overall concentration of TF (sum of molecules in solution and on the DNA) is constant, as well as the fraction of phosphorylated TF.

Gene product is produced and degraded during simulation and its mass is not conserved.

Detailed balance
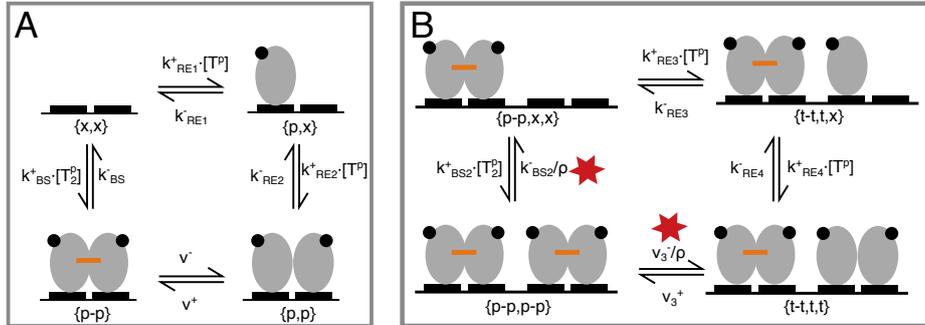
Figure 3.3: Cycles in dimer binding. A: Relation between monomer and dimer binding. **B:** Cooperative dimerization.

For models to be thermodynamically correct, they must fulfill detailed balance when at equilibrium. The definition of detailed balance is that every reversible reaction must be individually at equilibrium. In practice this means that whenever reaction cycles occur within a model, both possible ways from educt to product (which can be chosen arbitrarily) must show the same energy difference (or products of reaction rates along the path). Detailed balance imposes several constraints on the models:

- *Relation between monomer and dimer binding:* In model 1, a DNA state with a dimer bound can be reached either by TF dimerization in solution followed by dimer binding or by sequential monomer binding followed by dimerization on the DNA. The corresponding reaction cycle is shown in Fig. 3.3A. Application of the Wegscheider condition (Sec. 3.2) leads to the parameter constraint given by Eq. 3.2.

- *Cooperative dimerization:* When, in model 1, two dimers bind successively, they bind cooperatively, leading to a further reaction cycle (Fig. 3.3B). Again, the conditions for detailed balance can be found using the Wegscheider condition. The result is that dimerization on the DNA with the resulting state $\{p\text{-}p, p\text{-}p\}$ must be cooperative, i.e., it must occur with the modified constant $v_i \cdot \rho$.

- *Cooperative binding:* For cooperative binding, as found in all models, detailed balance requires that the cooperativity factor is independent of the order in which the dimers bind to the DNA. In a 3-BS model (used in Ch. 4), this means that cooperative binding to the middle BS is modulated by ρ^2 when both outer BSs are bound by T_2^p.

Model output and measures applied on the models

I performed deterministic simulations and as output I consider steady state concentrations of either DNA states or gene product g.

DNA occupancy. The model output is given as concentration of DNA in certain states. In equilibrium, the probability $p(i)$ of a state (or the fraction of time in which the promoter is in this state) can be derived from these by dividing the concentration of this state by the total promoter concentration:

$$p(i) = \frac{[i]}{\sum [i]}. \tag{3.4}$$

The DNA occupancy $p_{bound,BS}(f^p) = O_{BS}(f^p)$ gives the probability that one (or more) BS(s) is (are) bound as a function of the input signal. Occupancy of a certain BS is calculated as the sum of concentrations of all states in which the respective site is bound. Since the overall concentration of DNA is 1, the sum is equivalent to the probability. The change in occupancy upon phosphorylation (differential occupancy) of a single BS is given by

$$\Delta p_{bound} = O(f^p = 1) - O(f^p = 0). \tag{3.5}$$

Dynamic output range of gene expression. I use the dynamic output range of gene expression as a measure for the effectiveness of signal transduction. For a specific model with the

parameter settings a it is defined as

$$DOR_a = g_a(f^p = 1) - g_a(f^p = 0),$$

(3.6)

where g_a is the steady state gene output concentration.

Response shape. I apply a measure called δ, which gives the gradedness of a normalized response curve $f(x)$. At every data point i of the curve, the distance d_i to either 0 or 1 is calculated as follows:

$$d_i = \begin{cases} f(x_i) & \text{if } f(x_i) < 0.5 \\ 1 - f(x_i) & \text{if } f(x_i) \geq 0.5 \end{cases}$$

(3.7)

δ is then found by averaging over the distances:

$$\delta(f) = \frac{4\sum_i d_i}{N},$$

(3.8)

where N is the number of data points. If the response is a perfect switch, then $f(x)$ takes no intermediate values, but will always be either at baseline or take the maximum value, so that all $d_i = 0$ and, consequently, $\delta = 0$ (Fig. 3.7F). On the other hand, linear responses lead to an average distance of $d = 0.25$, so that $\lim_{N \to \infty} \delta = 1$ (Fig. 3.7E). Note that δ is underestimated by this numeric calculation and the error reduces with increasing number of data points. The measure δ applied here is inspired from the identically-named measure in Gunawardena (2005), which measures the switch quality of a function and is originally defined for continuous functions.

3.3 Model construction

The models were not manually constructed, but in a rule-based manner (Hlavacek et al., 2006), which means following a script. For all models, the same basic script was applied and the models differ in the rules for DNA binding and gene expression, which are defined at the beginning of the script. Especially for model 1, which features 113 species and 421 reactions, the construction script is more compact than the set of ODEs which constitutes the model. This is because even with simple binding rules, the number of species and reactions increases due to combinatorics. Thus, the rule-based approach saves both time and reduces errors from manual typing. A shared construction script further assures consistent nomenclature and equal setups for the models. This is useful for model evaluation, because it allows scripts that iterate over the models.

In this chapter there is one microscopic model (model 1) in which $N = 4$ binding units (4 REs) are considered, and three macroscopic models (models 2-4) which consider $N = 2$ binding units (2 BSs). In Ch. 4 I introduce two modifications of model 3 with one BS (model 3a) and three BSs (model 3b), which are also macroscopic (Tab. 3.2). For modeling, microscopic and macroscopic models can be largely treated in the same way (e.g., homodimeric TF in macroscopic models can be treated as monomers in the microscopic model).

Table 3.2: Overview of the models used in this thesis. For the number of reactions, matching binding and un-binding reactions are summarized into one reversible reaction.

Model	Phospho-mode	BSs	Abstraction	DNA states	Reactions
1	P.-dep. dimerization	2	microscopic	109	421
2	P.-dep. affinity	2	macroscopic	9	22
3	P.-dep. cooperativity	2	macroscopic	9	22
3a	P.-dep. cooperativity	1	macroscopic	3	6
3b	P.-dep. cooperativity	3	macroscopic	27	82
4	P.-dep. RNAP interaction	2	macroscopic	9	22

Definitions for model construction

Model species. For model construction, it is helpful to distinguish basic states and dimer states of DNA. Dimer states are DNA states in which dimerization is explicitly modeled. Therefore, in macroscopic models, all DNA states are basic. In microscopic models, basic states are those that contain only monomers (and no p-p).

Basic states. The set of basic DNA states contains $|S_b| = 3^N$ states, because each of the N binding units can be in either of three states (x, up or p). Model 1 has $|S_b| = 3^4 = 81$ basic states, models 2-4 have $|S_b| = 3^2 = 9$, model 3a has $|S_b| = 3^1 = 3$ and model 3b has $|S_b| = 3^3 = 27$.

Dimer states. Explicit dimer states exist only in model 1. The rules for dimers on the DNA are that any phosphorylated monomers (p) on two neighboring units j and $j + 1$ can form a dimer, but no p can be part of more than one dimer. This gives a maximum number of $\lfloor N/2 = 2 \rfloor$ dimers on one promoter. In model 1, there are $|S_d| = 1 + 3 \cdot 3^2 = 28$ dimer states: There is one state with two dimers, and three possible positions for one dimer where, for each, the remaining two positions lead to $3^2 = 9$ combinations.

Reaction details. In basic binding reactions, monomers/homodimers can bind to any empty position (x) in any DNA state. Standard association constants for binding/unbinding are given by the position and are K_{BSj} or K_{REj}. The K_A for a given basic binding reaction can be modified by affinity or cooperativity factors through rules so that $K_A = K_{BSj} \cdot \prod_k m_k$, where m_k are

modifiers for all rules that apply to this reaction (Tab. 3.3). Explicit dimer binding occurs only in model 1. T_2^p can bind to any two neighboring empty positions $(j, j + 1)$. The standard association constant is K_{BSj} and can be modified by the "microscopic phosphorylation-independent cooperativity" rule (Tab. 3.3).

Reaction rates. Because I am interested in the models in steady state, I consider only K_A values and not reaction rates for evaluation. In order to set up the model ODEs for simulation, however, k^+ and k^- rates are needed. Here, I simply set all $k^+ = 1$ and all $k^- = 1/K_A$.

Model settings. Settings for a particular model include the number of binding units N (which can mean REs or BSs), rules for basic and dimer binding and dimerization, and information whether it is microscopic or macroscopic.

Rules. Rules modify association constants for reactions depending on properties of the initial and resulting DNA state. I differentiate between rules for macroscopic (basic) binding, microscopic dimer binding and dimerization reactions. Relevant rules are evaluated while constructing the respective reactions (Tab. 3.3).

Table 3.3: Rules. t indicates a bound unit (*up* or p). The indicated condition is evaluated in order to find out whether the modifier m applies for the association constant. The last column gives the model(s) for which the rule is used.

Rule	Condition (for binding to position j)	m	Model
Dimer binding, microscopic			
P.-indep. cooperativity	The resulting state is $\{p\text{-}p, p\text{-}p\}$	ρ	1
Dimerization			
Cooperative dimerization	The resulting state is $\{p\text{-}p, p\text{-}p\}$	ρ	1
Dimer binding, macroscopic			
Phosphorylation-dependent affinity	The element that binds is p	μ	2
P.-indep. cooperativity	The resulting state is $\{t, t\}$	ρ	2,4
P.-dep. cooperativity	One *or* both elements at pos. $\{j - 1, j + 1\}$ equal p	ρ or ρ^2	3, 3a, 3b

Finding the model species

The set of basic states S_b can be found by spelling out all combinations of x, *up* and p distributed along N positions. For model 1, the set of dimer states S_d can be constructed from S_b by parsing the states for neighboring p and, if these are found, adding the corresponding dimer species to S_d. The procedure has to be repeated on the resulting dimer set for $\lfloor N/2 \rfloor - 1$ iterations (for

model 1 this is 1 iteration), in order to find DNA states with more than one dimer. Finally, since this method can detect dimer states more than once, S_d has to be reduced to a unique set.

Constructing the reactions

Basic binding reactions. Basic binding and unbinding reactions can be constructed from the set of all DNA states $S := S_b \cup S_d$. All binding units j in all states i in S are parsed and for every bound position $b_{i,j} \in \{up, p\}$, an unbinding and its reverse reaction are set up in which the resulting (unbound) DNA state is identical to i, except that element j is replaced by the empty unit x.

For possible modifications of the standard K_A, the conditions of all basic binding rules of the respective model are evaluated for each reaction that is set up.

Microscopic dimer reactions. Reactions for dimer binding and dimerization on the DNA can be constructed by parsing S_d. For every dimer p-p in every DNA state in S_d,

- set up a reversible dimer binding reaction where this dimer binds with K_{BS} and check microscopic dimer binding rules for possible modifications of the K_A and

- set up a reversible reaction for (un-)dimerization of this dimer on the DNA. The standard K_A is v. Check whether the cooperative dimerization rule applies.

Gene expression. For every DNA state in S, a gene expression reaction with the reaction equation $i \xrightarrow{\alpha_i^+} i + g$ is created, which can be done systematically by, again, parsing S. The reaction for degradation is $g \xrightarrow{\gamma} \emptyset$. The rates are given by the gene expression scheme (see Secs. 3.2, 4.3, 5.2).

3.4 Mode-specific DNA occupancy

To find out how promoter architecture and phospho-modes affect TF binding to the promoter, I scanned over a wide range of affinities for distal and proximal BS and calculated the steady state occupancy of both at $f^p = 0$ and $f^p = 1$. To predict the occupancy of a BS, the informative quantity is not the affinity constant alone, but rather the concentration of TF in relation to this constant. In Fig. 3.4, I therefore plotted relative concentrations, $c_j = [T_2] \cdot K_{BSj}$, where $[T_2]$ is held constant.

All four models show qualitatively similar DNA occupancy profiles for both unphosphorylated and phosphorylated state (shown as an example for $f^p = 0$ for model 1 in Fig. 3.4A). At low relative concentrations for both BSs ($c_1 \ll 1, c_2 \ll 1$), no BS is occupied (black quarter). If just one of the relative concentrations is strongly increased, then the respective BS gets bound to saturation, while the other is still unbound (green and blue quarters). A region in which both

BSs are occupied is found in the upper right corner, when concentrations for both BSs are high (cyan quarters).
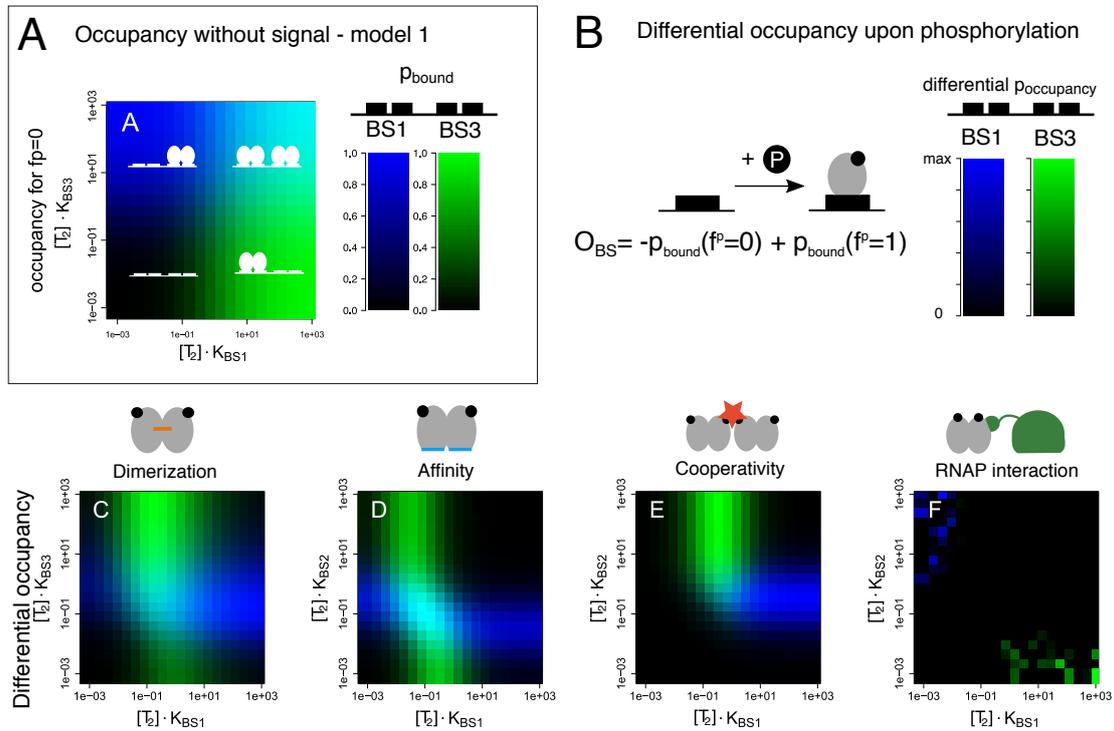

Figure 3.4: Differential DNA occupancy is mode-specific. This figure shows occupancy profiles for different combinations of relative TF concentrations ($c_j = [T_2] \cdot K_{BSj}$, explanation in text) for both BSs. Relative concentrations were log-uniformly scanned in the range $10^{-3} \leq c_j \leq 10^3$ (20 values each). **A:** Occupancy without signal ($f^p = 0$) is shown as an example for model 1. Occupancies are given as probabilities for BS1 (green channel in RGB), BS3 (blue channel in RGB) or both BSs (cyan = green + blue) to be bound. I assume equal RE affinities within one BS ($K_{RE1} = K_{RE2}, K_{RE3} = K_{RE4}$). **B:** Differential occupancy gives the change in occupancy upon phosphorylation. Panels **C-F:** Differential occupancy profiles for models 1-4. The profiles are normalized to their respective maximum value. The scattered spots in F arise from very slight inaccuracies inherent to numerical simulations and are only visible because of the normalization – none of the absolute values is larger than 10^{-14}.

In models 1-3, the regimes of one or more occupied BS increase upon phosphorylation, because phosphorylation directly or indirectly facilitates DNA binding. The region by which they increase is the phosphorylation-sensitive regime, as quantified by differential occupancy, which is the probability that a BS gets bound upon phosphorylation (Sec. 3.2). While all occupancy profiles look qualitatively similar (not shown), the changes induced by phosphorylation are model-specific (Fig. 3.4C-F).

When phosphorylation increases affinity (model 2), it equally supports binding transitions towards single-bound and double-bound state (Fig. 3.4D). They occur in a regime of intermediate affinity ($[T_2] \cdot K_{BS} \approx 1$), where T^{up} is mostly unbound, but T^p has an increased affinity sufficient for it to be mostly bound. The two responsive regions of BS1 and BS2 overlap in a small

regime where both relative concentrations are similar. In this region, phosphorylation changes the occupancy of both BSs at the same time (cyan).

In the cooperativity model (model 3), phosphorylation only supports transitions towards the double-bound state, since cooperativity stabilizes only this state (Fig. 3.4E). Also, simultaneous transitions of both BSs are less likely in this model, compared to model 2.

The occupancy range profile for model 1 (phosphorylation-induced dimerization), appears as a hybrid between the profiles for models 2 and 3 (Fig. 3.4C), since in this model dimerization is a prerequisite for both enhanced affinity and cooperativity.

In model 4, phosphorylation does not change promoter occupancy, irrespective of the BS affinities (Fig. 3.4F). This is because in this model, the effect of phosphorylation is solely at the gene expression level.

## 3.5 Rationale for a macroscopic model representation
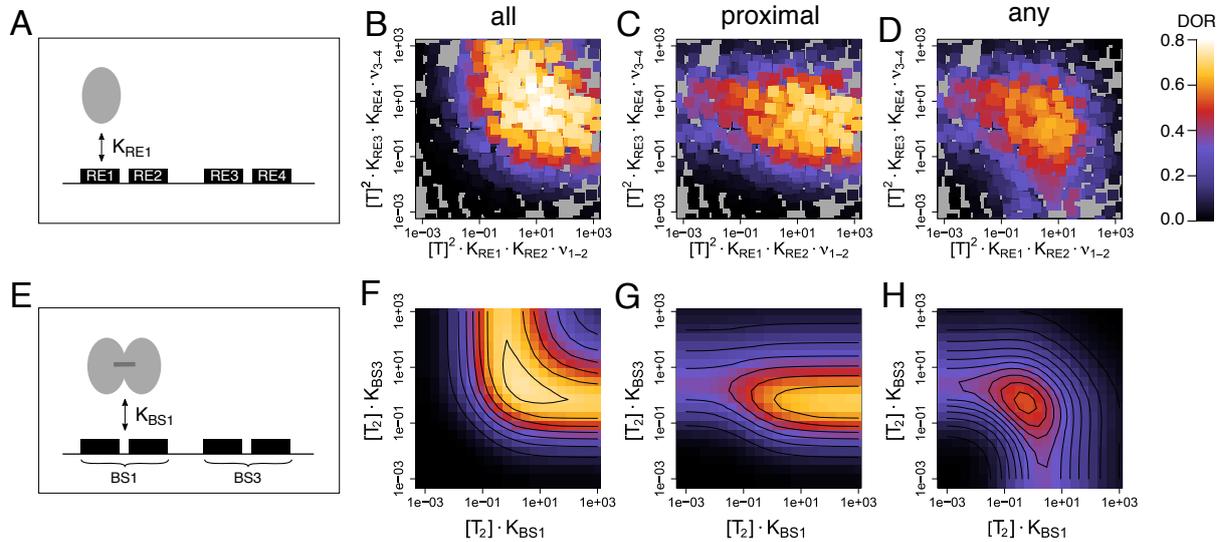
Figure 3.5: A macroscopic abstraction is sufficient to describe model 1. A: The microscopic version of model 1 resolves REs and TF monomers. **B-D**: *DOR* for microscopic model 1. For each data point, the values $K_{RE1\text{-}4}$ were sampled, then the resulting c_1 and c_3 were calculated using Eq. 3.2 and a square color-coded for *DOR* was plotted at the respective coordinates. **E**: The macroscopic version of model 1 abstracts to BSs and TF dimers. **F-H**: K_{BS1} and K_{BS3} were scanned (with $K_{RE1} = K_{RE2}, K_{RE3} = K_{RE4}$) and the resulting *DOR* is color-coded with the same scale as in B-D. In all simulations, $[T_2]$ was constant and $[T_2] \cdot K_{BS}$ was scanned by varying K_{BS}.

I measured the dynamic output range of gene expression (*DOR*) to quantify the effectiveness of signal transduction. This measure takes into account the three different output logics that

determine which BS(s) have to be bound for gene expression to take place (see Sec. 3.2). If a certain DNA state (e.g. "both BSs bound") is required for gene expression, then only those phosphorylation-induced changes in DNA occupancy that lead to this state also lead to a change in gene expression. This means, every gene expression logic can be seen as a mask to be applied on the differential occupancy profile, which filters out the relevant changes in occupancy.

The first question with respect to *DOR* is whether it is justified to regard model 1 in a macroscopic way – that is, on the level of dimers and with two BSs (Fig. 3.5E) – rather than on the level of monomers and REs (Fig. 3.5A).

Table 3.4: Influence of microscopic and macroscopic model parameters on *DOR*. A (log-uniform) sampling of the parameters K_{dim}, ρ, v_{1-3} and K_{RE1-4} was performed and the resulting *DOR* in model 1 was calculated. Correlation coefficients and corresponding p-values between the parameters' logarithm and the *DOR* are shown (calculated using R).

	»all«		»proximal«		»any«	
	CC	p	CC	p	CC	p
$[T] \cdot K_{RE1}$	0.01	$7e-01$	0.05	1e-01	-0.06	5e-02
$[T] \cdot K_{RE2}$	-0.09	$6e-03$	0.06	7e-02	-0.10	1e-03
$[T_2]^2 \cdot K_{BS1} \cdot K_{BS3}$	0.45	$8e-52$	0.40	4e-39	0.35	2e-29
$[T_2] \cdot K_{dim}$	-0.45	$1e-50$	-0.43	1e-46	-0.47	4e-57
v_1	0.33	$2e-27$	0.07	3e-02	0.30	3e-22
$\hat{v}_1 \cdot \hat{v}_3$	0.64	$2e-115$	0.60	7e-97	0.61	6e-102

A first sampling over most model parameters revealed no significant correlation between RE affinities and *DOR*, but a substantial one between the BS affinities and *DOR* (Tab. 3.4). Thus, to find out whether RE affinities only matter with respect to the resulting BS affinity, I performed another sampling over all four K_{RE} values, in which I fixed the remaining parameters to standard values (given in Sec. 2.3). From Eq. 3.2, I then calculated the resulting macroscopic affinities for BS1 and BS3 and plotted the *DOR* as a function of these (Fig. 3.5B-D). This gave the same result as scanning over BS affinities (Fig. 3.5F-H), thus corroborating the view that BSs are the relevant entities and a macroscopic view is justified.

The macroscopic view implies that the overlapping BS2 (RE2+RE3) is not helpful for effective signal transduction, since it blocks the state $\{p\text{-}p, p\text{-}p\}$ which, compared to $\{\_, p\text{-}p, \_\}$ (underscore signifies any state) increases the impact of phosphorylation by additional dimerization energy (two dimers vs. one) and cooperative forces.

A related question is whether it is justified to abstract the TF representation to dimers – or whether the distribution between monomers and dimers in solution (quantified by K_{dim}) is important. Counter-intuitively, dimerization strength in solution is both negatively correlated with the *DOR* (Tab. 3.4) and lowers the effective BS affinity (Eq. 3.2). The explanation for this is that

if dimer stability in solution exceeds dimer stability on the DNA, then the equilibrium distribution shifts away from DNA-bound TF, towards TF in solution. Therefore, the relevant quantities are $\hat{v}_j = v_j/([T_2] \cdot K_{dim})$, which compare dimer stabilities in solution and on the DNA and are contained in K_{BS}, again reducing the effective model parameters to the macroscopic ones.

In summary, a macroscopic abstraction is sufficient to describe the model and will therefore be used in the following.

3.6 Dynamic output range of gene expression
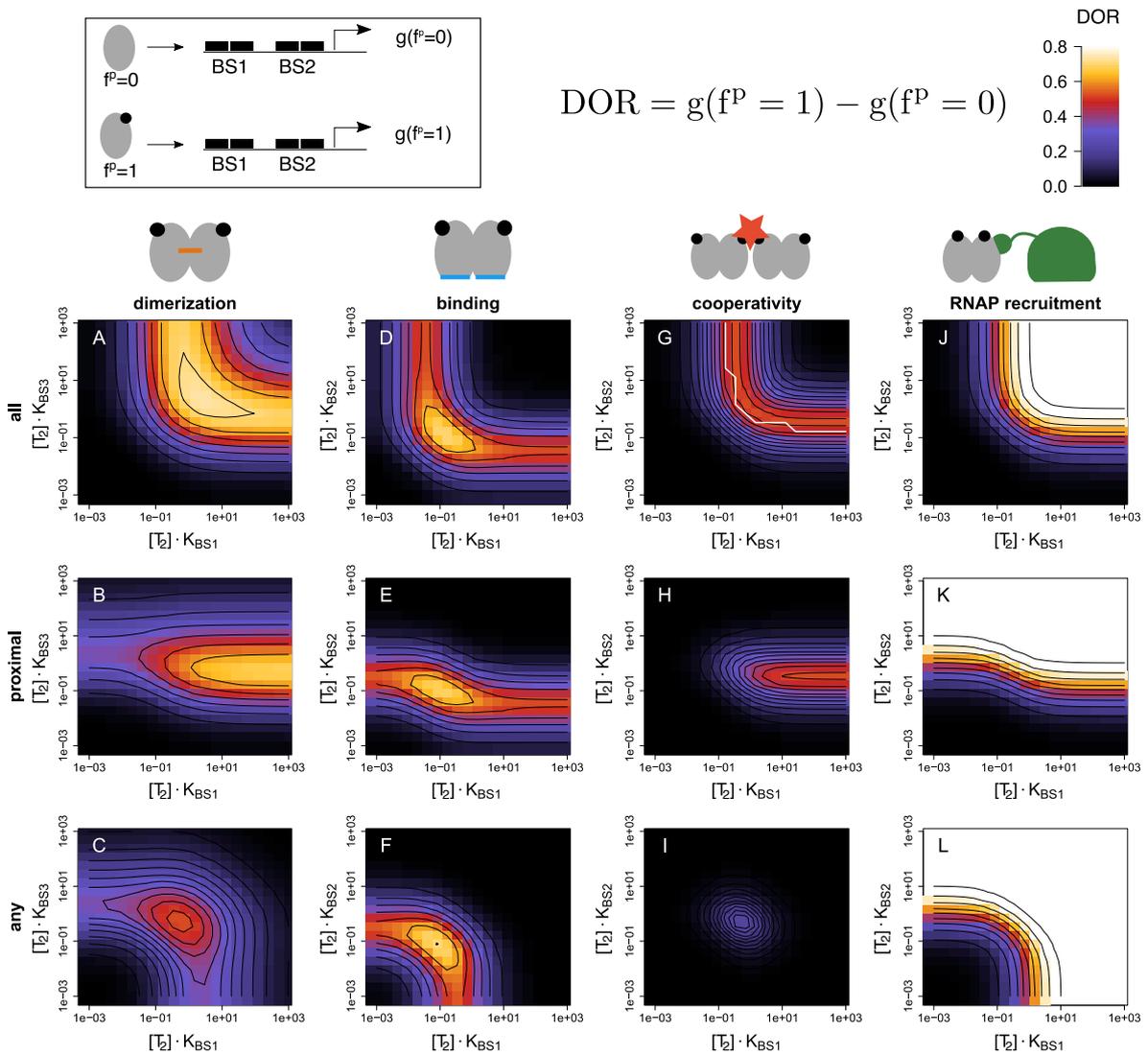
Figure 3.6: Optimizing the dynamic output range of gene expression. For each model (from left to right: model 1-4) and each expression scheme (from top to bottom: »all«, »proximal«, »any«) a scan over relative concentrations was performed and the computed *DOR* is shown color-coded. In all simulations, $[T_2]$ was constant and $[T_2] \cdot K_{BS}$ was scanned by varying K_{BS}. In panel G, the white isoline gives the parameter combination $1 + c_1 + c_2 \approx \sqrt{\rho} \cdot c_1 \cdot c_2$.

I compared the four modes with respect to their BS design requirements for effective signal transduction. For the effectiveness, I calculated the dynamic output range, which is the change in gene expression upon phosphorylation (Sec. 3.2).

In both model 1 and 2 phosphorylation directly or indirectly increases affinity. In these models both the »all« and »any« expression scheme show an optimum in DOR for equal affinities ($[T_2] \cdot K_{BS1} = [T_2] \cdot K_{BS2}$) (Fig. 3.6A, C, D, F). This is because BS identity includes the region where both BSs simultaneously get bound upon phosphorylation. In contrast, model 3 has no single-valued optimum DOR on the diagonal in scheme »all«, because simultaneous occupancy of both BSs upon phosphorylation hardly occurs in this model (Fig 3.6G). Rather, different combinations of affinities all lead to equal, optimal values for DOR. They form an isoline given by $1 + c_1 + c_2 \approx \sqrt{\rho} \cdot c_1 \cdot c_2$. This condition states that the probability for a double-bound state (RHS) is approximately balanced with other states (empty and single-bound, LHS), so that the promoter is in a responsive regime.

Exclusive expression schemes – that is schemes with strict requirements on which BSs have to be bound for gene expression – lead to larger affinity regimes of effective signal transduction in the dimerization and cooperativity models (1 and 3) (compare Fig. 3.6A, G with C, I). In both models, phosphorylation supports transitions towards the double-bound state (Fig. 3.4C, E). These transitions always lead to expression changes in the exclusive scheme »all« (Fig. 3.6A, G), but to lesser extents in scheme »proximal« (Fig 3.6H) and scheme »any« (Fig 3.6I). In the latter, the regime for this is very small and shows a reduced maximum DOR.

Phosphorylation does not increase the fraction of BSs bound in case of phosphorylation-dependent RNAP interaction (model 4), but the fraction of T^p on the promoter. Since the latter is a requirement for gene expression, the output at $f^p = 0$ will automatically be zero and the dynamic output range is optimized by maximizing the relevant BS affinities. Relevant are both BSs for »all« (Fig. 3.6J), BS2 for »proximal« (Fig. 3.6K) and any BS for »any« (Fig. 3.6L).

Phospho-factors include the dimerization factor ν, the affinity factor μ and the cooperativity factor ρ, i.e. those parameters that enhance the phosphorylation-dependent behavior and thus increase the impact of phosphorylation on gene expression. Increasing these factors strictly increases the measured DOR in all models for all tested expression schemes and combinations of relative concentrations (not shown).

In summary, for maximal effectiveness in signal transduction measured by DOR, the following promoter design requirements are predicted by the modeling results. (1) Exclusive expression schemes generally lead to larger affinity regimes of effective signal transduction. (2) For scheme »all«, phosphorylation-dependent dimerization and affinity lead to clear optima at equal affinities for both BSs, while phosphorylation-dependent cooperativity offers multiple combinations

of affinities that lead to optimal effectiveness. (3) For phosphorylation-independent TF binding in combination with phosphorylation-dependent gene expression (model 4), the rationale is that high affinities lead to a high dynamic output range.

## 3.7 The tunability of response shape
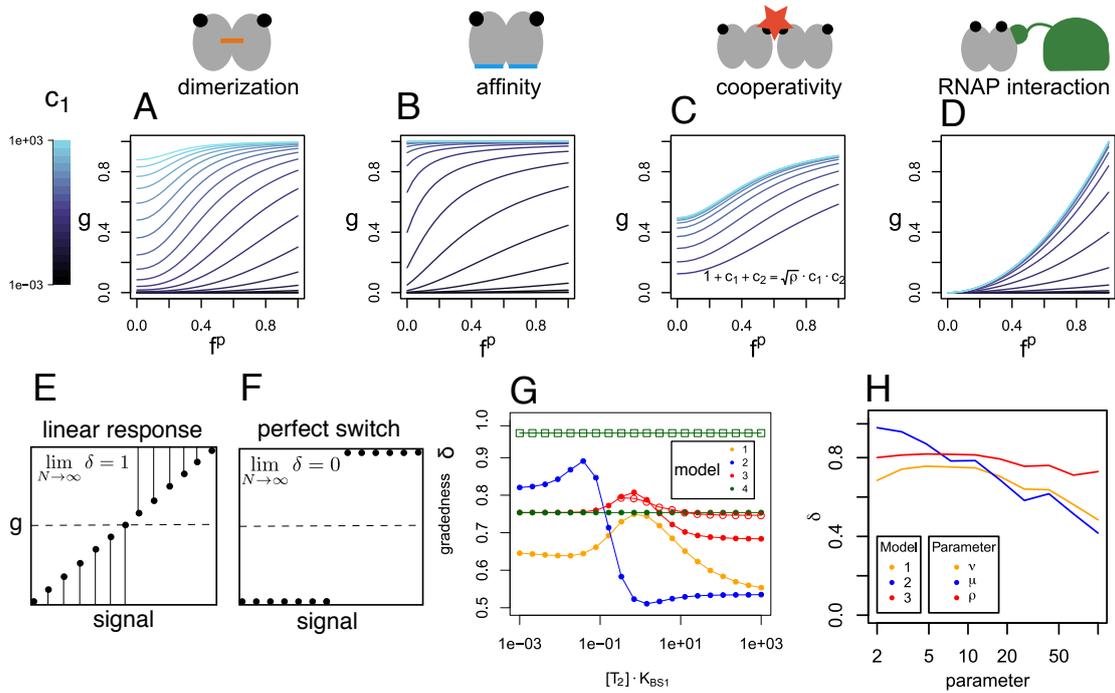
Figure 3.7: Modes allow tuning the response shape as a function of promoter architecture to different extents. A-D: Response curves for models 1-4, scheme »all«. Each line was calculated for a different value of K_{BS1} (color-coded). In all panels except C, $K_{BS2} = K_{BS1}$. In panel C, K_{BS2} is chosen to fulfill $1 + c_1 + c_2 = \sqrt{\rho} \cdot c_1 \cdot c_2$. **E-F**: For the calculation of δ, the distances (indicated by lines) to either baseline or maximum are summed up. In the limit, $\delta = 0$ for a perfect switch and $\delta = 1$ for a linear response (details: Sec. 3.2). **G**: Gradedness in the four models (color-coded) as a function of affinity, scheme »all«. Solid dots: $K_{BS2} = K_{BS1}$. Open circles: Model 3, K_{BS2} fulfills $1 + c_1 + c_2 = \sqrt{\rho} \cdot c_1 \cdot c_2$. Green squares: model 4, scheme »proximal«. **H**: Phospho-factors enhance switch-like behavior. For every parameter value, $c_1 = c_2$ was scanned between 10^{-3} and 10^3 for maximum *DOR* and δ was calculated for the respective response curve. Scheme »all«.

Not only the dynamic output range, but also the shape of a response can be an essential feature of a gene expression system.

A first glance at the response curves reveals a few mode-specific features: Directly or indirectly phosphorylation-dependent cooperativity allows a slight S-shape (Fig. 3.7A,C), which is not seen for phosphorylation-dependent affinity or RNAP interaction (Fig. 3.7B,D). The latter (model 4) constitutes a special case in which, independent of the BS affinities, all response shapes display quadratic (scheme »all«) or linear (scheme »any«) behavior. None of the curves in model 4 has an offset, because expression is strictly phosphorylation-dependent (Fig 3.7D).

Monotonic responses (which are given here), are often classified into graded or switch-like behavior. I applied the measure δ, which gives the gradedness of a response curve and takes the value 1 for a perfectly graded response (Fig. 3.7E) and 0 for a perfect switch (Fig. 3.7F) (see Sec. 3.2).

I find that the four modes allow tuning of gradedness through BS affinities to different extents (Fig. 3.7G). Small changes in BS affinities have a substantial impact on the response shape for phosphorylation-dependent affinity (blue line in Fig. 3.7G). δ measurements for model 3 are robust towards changes in overall affinity, compared to the other models. They are especially stable along the isoline of optimal effectiveness of signal transduction (filled and open red circles in Fig. 3.7G). Again, model 1 is a phenotypic combination of model 2 and 3. Model 4 has constant gradedness δ, which depends only on the expression scheme, in line with the constant response shapes seen above (green circles and squares in Fig. 3.7G).

As mentioned above, increasing phosphorylation dependence via the phospho-factors ν, μ and ρ in models 1-3 positively impacts the dynamic output range. These parameters also consistently augment the switch-like character of the response shape, as they decrease δ (Fig. 3.7H). The strongest effect on δ can be observed for μ in model 2.

In summary, the highest tunability of response shape due to BS affinities can be observed for phosphorylation-dependent affinity, as opposed to relatively stable and constant shapes for phosphorylation-dependent cooperativity and RNAP interaction (models 3 and 4), respectively. Further, increasing the phospho-factors (ν, μ and ρ) shifts the systems towards switch-like responses.

3.8 Summary and discussion

I compared four possible modes of phospho-activation in TCSs with respect to effectiveness of signal transduction and shape of response. I found that the ideal promoter architecture for effectiveness depends on both phospho-mode and expression scheme: Phosphorylation-dependent affinity and dimerization favor equal BS affinities, while cooperativity allows multiple combinations and phosphorylation-dependent expression operates best with high affinities. Restrictive expression schemes, that is schemes that require both BSs (»all«) or at least a certain one (»proximal«) to be bound, generally broaden the regime of effective signal transduction. Response shapes for the 2-BS promoter studied here appeared rather graded with the standard values for phospho-factors, for all modes, but became more switch-like when phospho-factors were increased. The results indicate that phosphorylation-dependent affinity allows for a higher degree of tunability of the response curve than phosphorylation-dependent cooperativity.

Two-component signal transduction from receptor activation to response regulator (RR) phos-

phorylation has been the subject of modeling efforts, especially with respect to robustness (Batchelor and Goulian, 2003; Shinar et al., 2007), but also concerning response shape and hysteretic behavior (Igoshin et al., 2008). But despite experimental evidence of different phospho-modes, these have not been addressed by modeling work yet. Models on gene expression are, of course, numerous and comprehensive theoretical studies have investigated the interplay between TF, DNA and RNAP, regarding putative logical operations (Buchler et al., 2003) or response shape (e.g. Estrada et al. (2016)) while considering cooperativity, promoter architecture and expression schemes. However, all of these studies regard gene expression as a function of active TF concentration, where active TF can bind to DNA and inactive TF has no function at all. This might not always be the case and may have led to overlooked options. For example, classical gene activation models identify equal BS affinities as a feature optimal for effective signal transduction (Bintu et al., 2005a), while many natural promoters do not display this architecture. This work therefore fills a gap by focusing on the impact of possible diversity arising from phospho-modes.

I chose to model a promoter with two BSs as a starting point, while all models can be easily extended to more BSs (and will be in Ch. 4). Clusters of BSs are a common feature of bacterial promoters (Collado-Vides et al., 1991) and several functions have been attributed to them in the literature: Giorgetti et al. (2010) investigated the graded response of human cells to TNFα, which is orchestrated by clusters of 5-6 BSs. They find that their experimental data is best fitted by a model with 6 BSs, non-cooperative binding (and an "additive" expression scheme, where all TF molecules attract RNAP independently). Thus, under certain conditions, additional BSs could serve for increasing a response's gradedness.

Conversely, Lengyel et al. (2014) find in a modeling study that high numbers of BSs in combination with a high threshold of BSs that need to be bound for expression (similar to scheme »all«) can increase the apparent Hill coefficient of the response curve. This study, too, assumes non-cooperative binding.

Gunawardena (2005) studied multi-site protein phosphorylation, but the model is still comparable to the ones used in this study: an input species (TF/kinase) acts on several sites (BSs/phosphorylation sites) of the same type and the sites' states are integrated to a single output (expression/fully phosphorylated protein). This study offers a more differentiated picture: the author states that additional sites can make the system a better switch, but only under certain parameter settings. Similar results are presented by Estrada et al. (2016), who find that (in equilibrium models), the apparent Hill coefficient of a gene expression system with cooperative TF binding is bounded by the number of BSs.

Again, these studies do not distinguish phospho-modes.

I did not perform any scans over the concentrations of TF or DNA. This is because, as long as TF is in excess over DNA, results depend only on relative concentrations ($[T_2] \cdot K_{BS}$). This

condition is fulfilled, because I chose a TF:DNA ratio of 100:1, so that even when both BSs are bound, the concentration of TF barely changes. $[T_2]$ only plays a role at concentrations so low that they are notably decreased by sequestration through DNA binding. Note that the "excess condition" can still be fulfilled when *relative* concentrations are low due to a low K_{BS}. It might not be fulfilled for low numbered TFs that bind to multiple target promoters. In such cases, the distribution of TF along the promoters may become of interest and models should be adapted accordingly (see also Sec. 6.3).

I assume here that DNA occupancy by TF affects binding of RNAP, but not vice versa. If recruitment of RNAP by TF was modeled as cooperative interaction with TFs, then, of course, RNAP binding would also affect TF binding (Bintu et al., 2005a; He et al., 2010). In steady state, this condition is difficult to interpret, because even if RNAP would have a stabilizing effect on TFs, it supposedly occupies its binding site only for a short time before transcription sets off. I would not expect qualitative changes when modeling RNAP as a species that can bind to DNA in a cooperative fashion with TF. Rather, I would merely expect a global shift of the optima towards lower relative concentrations, due to additional stabilization of TFs by RNAP.

Wherever possible, I modeled the modes of phospho-activation as strict dependencies, in order to construct generic cases. That is dimerization, cooperativity or expression are only possible with phosphorylated TF in the respective models. In model 2, however, phosphorylation-dependency is only partial, since binding is also possible for T^{up}, it is just more efficient for T^p. I did so, because certain experiments reported this behavior. A strict dependency would be representable with the model by choosing a high affinity factor μ and low BS affinities.

For phosphorylation-dependent affinity (model 2), I find that the response shape is sensitive to changes in relative TF concentration. A biological advantage of this feature could be that promoters of the same regulon can be varied in their response via small changes in BS affinities. On the population scale, high sensitivity in combination with small stochastic differences in intracellular TF concentration between bacteria could be a source of heterogeneity. Conversely, response shapes for the phosphorylation-dependent cooperativity model were relatively insensitive to changes in relative concentration. This could be a form of robustness towards changes in TF concentration which might be favored for systems that require a reliable response to certain conditions.

All the responses I found with my models are more graded than switch-like, even when phospho-factors (ν, μ, ρ) are increased up to 100. This means, for switch-like responses, either very high phosphorylation parameters or additional BSs (see above) might be necessary.

Understanding the interplay of TFs, DNA and RNAP suffers from a number of uncertainties, regarding applied experimental methods and interpretation of the results which is in turn related

to the identification of a system's relevant functional features.

Oligomerization state of RRs. There is considerable diversity between TFs concerning their oligomeric state which leads to some confusion about their DNA binding properties. For example, most RR TFs are found to form dimers in solution, although the proportion between monomers and dimers appears to vary between different TFs and sometimes also between studies on the same TF. Several TFs such as *B. subtilis* PhoP seem to exist mostly in dimeric form, independent of phosphorylation (Liu and Hulett, 1997), even though this has been questioned by other findings (Sinha et al., 2008). Some, like PhoP in *Salmonella enterica*, have been found to be in a mixture of monomer and dimer (Perron-Savard et al., 2005). For others, like members of the OmpR/PhoB subfamily in *E. coli*, dimerization is enhanced by or even entirely dependent on phosphorylation (Gao et al., 2008). Surprisingly, model 1 suggests that dimerization in solution alone is not informative about the TFs' DNA binding ability. While isolated studies on TFs can provide information on whether they are able to dimerize at all, this study indicates that the question relevant for DNA binding remains whether dimerization is possible on the DNA and whether the dimer is more stable on the DNA than in solution.

What is the relevant binding unit? Studies on the oligomeric state of a TF are closely related to the question what the relevant binding unit is. Does it make a difference whether TFs bind as monomers to REs (which is sometimes hypothesized for those that do not form dimers in solution) or as dimers to BSs? And, is the pairwise classification of REs into BSs justified, or do overlapping BSs also have a function? To this end, my dimerization model indicates that, at least with the mechanism modeled here, the entities which determine signaling effectiveness in 2-BS promoters are BSs. Further, an arrangement with an overlapping BS would, again with respect to signaling effectiveness, be inferior to pairwise grouping of REs. The latter arrangement comes at the advantage of optimal exploitation of dimerization energies. Thus, my results suggest BSs as the relevant unit.

What is the underlying phospho-mode? Observations in many studies do not unambiguously identify or differentiate a mechanism. For instance, several studies state "self-association" (Lejona et al., 2004; Gao et al., 2008), where it is unclear whether this means cooperativity or dimerization in solution or on the DNA. I argue that this is an important distinction. My results suggest that "self-association" meaning (1) dimerization strength in solution alone is not an informative quantity (see above), while (2) dimerization on the DNA could result in features that resemble both dimer cooperativity and affinity. Importantly, (3) dimer cooperativity shows different promoter architecture requirements and a different response shape from dimerization. The lesson learned from modeling is therefore that we should be careful to distinguish these effects, both in experiments and by terminology.

In *in vitro* studies on promoters, binding is typically tested for the promoter region as a whole,

and in some cases, single REs or BSs of more complex promoters are removed to prove their contribution to gene expression. Unfortunately, many of these experiments fail to determine the exact mode of phospho-activation, because "enhanced expression" or "enhanced binding to the promoter as a whole" due to phosphorylation are observations that can indicate different functions of phosphorylation. For example, the observation of "enhanced affinity to the promoter region upon phosphorylation" could be obtained with either phosphorylation-dependent dimerization, affinity or cooperativity. Here, an additional source of uncertainty could be the methods used for measuring BS affinities (Head et al., 1998).

Also, mutational studies which find that a removal of a certain BS decreases gene expression allow at least two fundamentally different interpretations: TF bound to this site could be responsible for recruiting RNAP, or it could be necessary for cooperatively recruiting TF bound to another site, which then contacts RNAP. These difficulties are reflected by sometimes contradictory results or theories concerning the mechanisms of phospho-activation in gene expression: For instance, OmpR binding to P_{ompF} has been found to be cooperative by some studies (Huang et al., 1997; Rampersaud et al., 1994), which was contradicted by others (Head et al., 1998).

In fact, my models for phosphorylation-dependent dimerization and affinity show partially similar results, for both dynamic output range and response shape. Dimerization is therefore a putative explanation for experimentally observed enhanced affinity, also to a single BS. However, the differences between phosphorylation-dependent affinity, cooperativity and RNAP interaction are substantial in this study: These modes differ in their requirements for promoter architecture and in the response shapes the systems can generate. Therefore, learning about the phospho-mode should be of fundamental interest when researching TCS-dependent gene expression.

I believe that my results can help finding phospho-mode and expression scheme of TCSs when applied on sequence data: For example, BSs with asymmetric affinities could be a hint for phosphorylation-induced cooperativity, while more than one near-ideal BS sequence would suggest phosphorylation-dependent RNAP interaction as implemented in model 4.

Expression schemes can be rather well distinguished based on *DOR* profiles. In all my models, scheme »all« tolerates high affinities of one of the BSs (no matter which one), »any« tolerates low affinities of one BS, and »proximal« poses tight restrictions on K_{BS2}, while tolerating both high and low affinities for the distal BS. These findings might be especially useful when comparing sequences of several promoters from the same regulon. Here, the observed variability in BS affinity (as approximated by deviation from consensus) could give hints at which BS high or low affinities are tolerated, and thus, which expression scheme might apply.

Hence, sequence data analysis combined with modeling predictions could be used in addition to experiments, especially when the latter give contradictory results.

In synthetic biology, the engineered systems are to fulfill a specific task, which in turn can

require their optimization for effectiveness, response shape or offset. Since mutating BSs is thought to enable the tuning of their affinities (Stormo and Fields, 1998), the results of this chapter could be a helpful basis for the engineering of TCSs with known phospho-modes. On the other hand, for TCSs with unknown mode, the fact that model 1 behaves as a hybrid between models 2 and 3 could be an advantage, because it might allow designs that fit several phospho-modes. For example, my models predict that when the exact mode is not known, a relatively safe choice to construct a system with high dynamic output range is to use BSs of equal affinities.

I investigated both response shape and the effectiveness of signal transduction (dynamic output range) in gene expression. While I suppose that a high dynamic output range is generally favorable in signal transduction systems, the optimal response shape may depend on context. Therefore, other objective functions may be the subject of future research. For instance, in another modeling study, Bintu et al. (2005a) used the maximum sensitivity of gene expression to TF concentration (log-log slope) as objective function. Locally, high sensitivity could either indicate an overall high dynamic output range, or also a switch-like response shape. Interestingly, in a model for synergistic activation, which is comparable to model 2, »all«, Bintu et al. (2005a) come to the conclusion that sensitivity is maximized for equal BS affinities – the same conditions that I find to maximize the dynamic output range.

Another frequently discussed feature of gene expression systems is their reliability in terms of either robustness to variable initial conditions or stochasticity due to low copy numbers of the involved proteins. For example, Bundschuh et al. (2003) modeled a simple gene expression system where a TF suppresses its own transcription and found that in this context TF dimerization reduces intrinsic fluctuations of the output TF concentration.

It will be interesting to compare my models also with respect to stochastic effects, which could be parametric or intrinsic fluctuations. This will require resolving binding constants into forward and reverse rates and a more detailed description of the expression itself, including separate species for RNAP, mRNA and protein. However, this is beyond the scope of this work.

On DNA binding and gene expression level, investigating robustness towards changing (initial) conditions would primarily concern fluctuations in TF concentration (which is a parameter and not the signal in my study), because many bacterial TFs are known to have low copy numbers.

Chapter 4

Signal integration and promoter architecture in *B. subtilis* quorum sensing

4.1 Introduction

In *B. subtilis* quorum sensing (QS) two signals converge in the activation of the QS master regulator ComA, which is a response regulator (RR) functioning as a transcription factor (TF). A strain-specific signal (CSF) causes the release of ComA from a sequestering reaction, and a species-specific signal (ComX) leads to phosphorylation of ComA (see Sec. 1.2). From the point of view of evolutionary theory, cooperative behavior as triggered by QS is only useful when placed under kin selection control. In this chapter I therefore investigate whether – despite the signal convergence observed at the level of signaling pathway architecture – the promoter architecture offers ways in which the control of the strain-specific signal over QS can be increased. I will also discuss the implication of my findings for the evolution of QS-dependent promoters.

Promoter architecture in *B. subtilis* QS

So far there are 11 operons known to be regulated by ComA (Comella and Grossman, 2005; Wolf et al., 2016). They typically show a weak BS for σ^A, the housekeeping σ factor for RNAP in *B. subtilis* (Nakano et al., 1991b; Mueller et al., 1992; Lazazzera et al., 1999; Jarmer et al., 2001). Most ComA-dependent promoters feature a core module composed of two binding sites, an inverted repeat (IR) and a direct repeat (DR), each of which consist of two recognition elements (REs) (Fig. 4.1). The IR is more distal with respect to the RNAP binding site and has a consensus sequence of TTGCGG-4nt-CCGCAA from which no known promoter deviates in more than three nucleotides (Nakano and Zuber, 1993; Comella and Grossman, 2005). The DR in some cases overlaps the -35 element of the RNAP binding site and has the idealized motif TTGCGG-5nt-TTGCGG. The spacer between IR and DR varies between 4 and 13 nucleotides (nt). An exception to the core module is found in P_{rapI}. This promoter lacks the IR and consists

of a single DR only (Wolf et al., 2016). In two other promoters, additional BSs exist upstream of the core module: P_{srfA} contains an additional IR and P_{pel} comprises three additional DRs that overlap each other as well as with the core module's IR (Nakano et al., 1991b; Wolf et al., 2016).
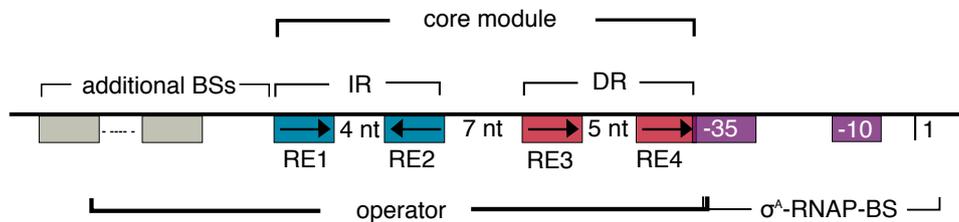

Figure 4.1: Promoter architecture of QS-dependent genes in *B. subtilis*. The core module of ComA-dependent promoters in *B. subtilis* consists of an inverted repeat (IR) and a direct repeat (DR), which have two REs each. The DR can overlap the -35 element of the RNAP BS. Spacer lengths within and between BSs are indicated. The transcription start site is marked as 1.

Complex functionality of ComA-dependent promoters

ComA binds to the DNA. Electrophoretic mobility shift assays (EMSAs) showed that ComA binds to the BSs found by sequence analysis. All REs of the core module contribute to ComA binding and the affinity increases towards the consensus/idealized sequence (Griffith and Grossman, 2008; Wolf et al., 2016). BS identity (i.e. whether it is IR or DR) does not influence the affinity of ComA to single BSs (Wolf et al., 2016).

BSs as functional units. Evidence that the two half sites of each of the two core module BSs form a functional unit comes mainly from investigating the spacers between the RE: With respect to both ComA binding and gene expression, mutations in spacer lengths within IR and DR are hardly tolerated (Wolf et al., 2016), while, the spacer between IR and DR of P_{rapF} shows greater flexibility than the spacers within binding sites (Griffith and Grossman, 2008). Interaction between the BSs (or the TFs bound to them) is therefore presumably weaker than between the REs of one BS, but nevertheless existent.

Which BSs contribute to gene expression? *In vivo*, it was confirmed that RE1-RE4 all contribute to gene expression and that mutations away from or towards consensus decrease or increase gene expression, respectively (Griffith and Grossman, 2008; Wolf et al., 2016; Nakano and Zuber, 1993). In addition, assays have been conducted in which deletion or mutation of one BS or RE was compensated for by mutating another one towards consensus.

- In *srfA* deleting the additional IR upstream of the core module largely abolishes transcription. This can be compensated for by a perfectly symmetric core module IR (Nakano and

Zuber, 1993) and even overcompensated by a fully consensus RE3 (Fig. 2a in Griffith and Grossman (2008)).

- In *rapF* mutations away from consensus in RE3 can be compensated for by changes towards consensus in RE2 (Griffith and Grossman, 2008).

An uncontroversial conclusion from this kind of experiments is that IR and RE3 contribute to gene expression. In which way, however, is disputable.

Cooperative binding. Nakano and Zuber (1993) concluded from the compensation assays that ComA molecules bound to different BSs interact cooperatively, so that a strong upstream IR could facilitate binding to the core module IR. Even earlier, the mere existence of two binding sites had already led Nakano et al. (1991b) to proposing cooperative binding. A more direct way to show cooperative binding are quantitative EMSA, which indeed indicated a Hill coefficient >1 (Wolf et al., 2016).

The role of phosphorylation. Phosphorylation is known to increase the affinity of ComA to P_{srfA} as a whole (Roggiani and Dubnau, 1993). However, no published results are available on the question by which mechanism phosphorylation supports DNA binding.

Open questions

Even though they are all activated by the same TF, ComA-dependent promoters differ with respect to the number and types of BSs, and their similarity to consensus. The exact function-alities of the promoter – and with this the reason for promoter diversity – are still unclear. New and to date unpublished findings from my experimental collaborators (Complex Adaptive Traits research group of Dr. Ilka Bischofs) concerning the role of phosphorylation will be presented in Sec. 4.2. They motivated the following theoretical study on signal processing at the promoter level. The main question is whether QS-dependent genes can be differentially activated by the two QS signals, depending on promoter architecture.

4.2 New insights from experimental data

The following experimental results were provided by the Complex Adaptive Traits research group of Dr. Ilka Bischofs at the Max Planck Institute for Terrestrial Microbiology in Marburg.

Gene expression studies

To test the influence of the two *B. subtilis* QS signals on gene expression, several knock-out strains of components of both pathways were constructed, which additionally contained fluo-

rescence reporters fused to different ComA-dependent promoters. Compared to wt, gene expression from both a single-BS and a double-BS promoter decreased when the un-sequestering pathway was impaired (Fig. 4.2, yellow boxes). Mutations in the ComX/ComP phosphorylation pathway, however, only had negative effects on the expression from a double BS promoter, while the effect on the single BS promoter was positive (Fig. 4.2, blue boxes). Thus, the effect of ComA phosphorylation through ComP appears to depend on the promoter architecture.
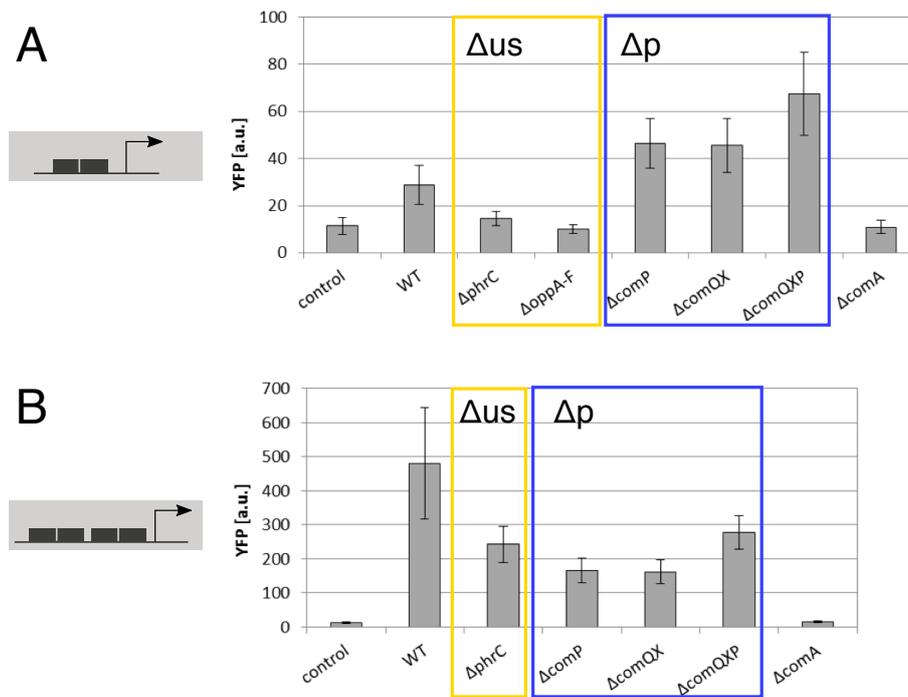
Figure 4.2: Experimental gene expression studies. Fluorescence reporter expression from **A:** a perfect direct repeat (PDR) and **B:** a PIR-PDR (perfect inverted repeat - perfect direct repeat) segment subject to mutations in the ComA un-sequestering (yellow boxes) and ComX/ComP phosphorylation pathways (blue boxes). The respective promoter was fused to a YFP reporter. On the x-axes, the names of the deleted genes are given. Graphs adapted from Diana Wolf and Valentina Rippa. Experimental procedures are described in Wolf et al. (2016).

DNA binding studies

In EMSA experiments, the ability of unphosphorylated ComA and ComA phosphorylated by ComP to bind to a single BS was measured and no difference was found: Unphosphorylated and phosphorylated ComA bind a perfect direct repeat (PDR) (Fig. 4.3A) and PIR (perfect inverted repeat, not shown) equally well. This indicates that phosphorylation through ComP does not impact ComA binding to a single BS.

Next, the PIR-PDR core motif was tested. Here, a lower dissociation constant (K_D) and a steeper binding curve were observed in case of phosphorylation. This indicates cooperative binding of phosphorylated ComA to the 2-BS core module (Fig. 4.3B). Phosphorylation-dependent cooperative DNA binding can explain why mutations in the phosphorylation pathway only impair

gene expression from a double BS promoter (Fig. 4.2), since in a single-BS promoter, no co-operativity is possible. It does not explain the observed increase in gene expression from a single-BS promoter due to knock-outs in the ComP phosphorylation pathway. One possible explanation for this phenomenon might be a negative feedback through ComA-dependent Rap genes.

Overall, the presented results point towards phosphorylation-dependent cooperativity of ComA molecules that bind to neighboring BSs.
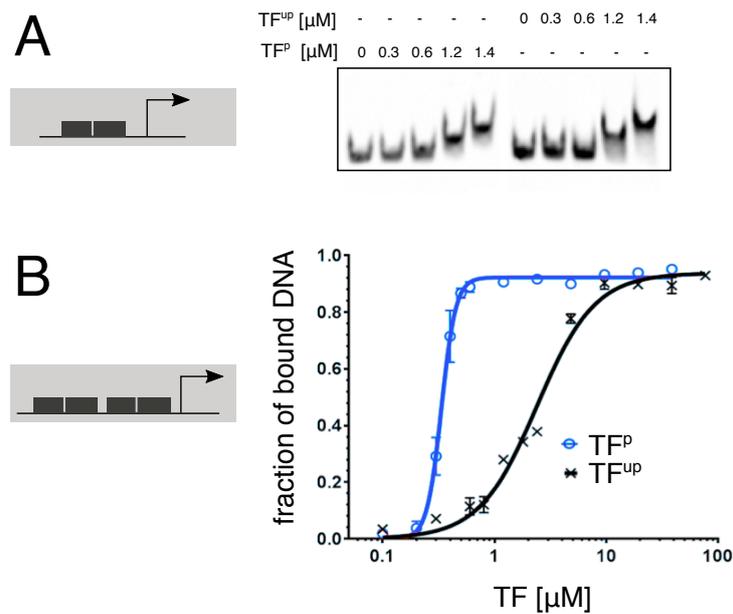
Figure 4.3: Experimental DNA binding studies.. **A:** Binding to a PDR in an EMSA (experiment by Valentina Rippa). Phosphorylated ComA (TF^p) was obtained through incubation with ComP before loading the samples on the gel. **B:** Binding to PIR-PDR. Each data point comes from the comparison of band intensities in an EMSA. Phosphorylated ComA (TF^p) was obtained through incubation with ComP before loading the samples on the gel. Graph from Valentina Rippa. Experimental procedures are described in Wolf et al. (2016).

4.3   Models for QS-dependent gene expression in *B. subtilis*

Rationale of the models

The experiments presented above indicate phosphorylation-dependent cooperativity for binding of ComA to the DNA. Therefore, model 3 from Ch. 3, which assumes this dependency, serves as a basis for the modeling study in this chapter.

The following rationale is used for the application of the model:

1. As in Ch. 3 I assume a constant TF concentration over the time course of simulation. In this chapter I consider two signals that change the state of the TF: the fractions of phosphorylated and un-sequestered TF. In line with experimental findings (Core and Perego,

2003), I assume that the phosphorylation and un-sequestering states of the TF are independent of each other.

2. ComA is a dimer in solution (Griffith and Grossman, 2008; Hobbs et al., 2010) and thought to bind to the DNA as a dimer (see Sec. 4.1). Further, known ComA-dependent promoters display only full BSs (and no single REs), justifying my approach to abstract from monomers and REs and to model only dimers and BSs.

3. The core module of ComA-dependent promoters consists of two BSs, which is consistent with model 3. However, for this chapter I also adapted the model to one and three BS(s), respectively.

4. The experiments presented in Sec. 4.2 gave no hints for phosphorylation-dependent affinity (increased K_A value) or phosphorylation-dependent dimerization – which, as results in Ch. 3 suggested, would also lead to increased affinity. Therefore, I suppose that the only effect of phosphorylation is to allow cooperativity.

5. ComA promoters have a weak RNAP BS (Nakano et al., 1991a; Mueller et al., 1992; Lazazzera et al., 1999; Jarmer et al., 2001), suggesting that no basal activity is expected in the absence of TF. Therefore, in my models gene expression is strictly dependent on TF bound to the promoter.

6. I consider the three generic expression schemes introduced in Sec. 3.2, because the exact DNA occupancy conditions for gene expression are unclear.

Model overview

Input signals. The model input describes the phosphorylation and sequestering states of the TF, leading to the four TF dimer species T_{us}^{up}, T_{us}^{p}, T_{s}^{up} and T_{s}^{p}. While the overall dimer concentration $[T_2]$ is fixed, the signals

$$f_{us} = \frac{[T_{us}]}{[T_2]} = \frac{[T_{us}^p] + [T_{us}^{up}]}{[T_2]} \qquad \text{and} \qquad f^p = \frac{[T^p]}{[T_2]} = \frac{[T_{us}^p] + [T_s^p]}{[T_2]} \tag{4.1}$$

give the fractions of un-sequestered and phosphorylated TF, respectively (Fig. 4.4A). Un-sequestering is a general requirement for binding to the DNA, and phosphorylation is a requirement for cooperative binding.

DNA binding. DNA binding follows the same rules as presented in Sec. 3.2, with the extension that only un-sequestered TF can bind to any BS: In general, binding to BS_j occurs with the association constant K_{BSj}, independent of phosphorylation (Fig. 4.4B, left). An exception

occurs when phosphorylated TF (T_{us}^p) binds next to a BS which is already bound by another T_{us}^p. In this case, phosphorylation-dependent cooperativity leads to an increase in affinity by the cooperativity factor ρ (Fig. 4.4B, right). This rule is applicable for 2-BS promoters, as well as for 1- or 3-BS promoters – even though, of course, on 1-BS promoters it has no effect. For 3-BS promoters, detailed balance requires that the affinity is increased by the factor ρ for each neighboring bound T_{us}^p, that is when the middle BS is the last to be bound by T_{us}^p, then the K_A is increased by ρ^2. TFs on a fully occupied 3-BS promoter can therefore be stabilized by two cooperative interactions.

A Input signals

B DNA binding   **C** Gene expression

D Promoters
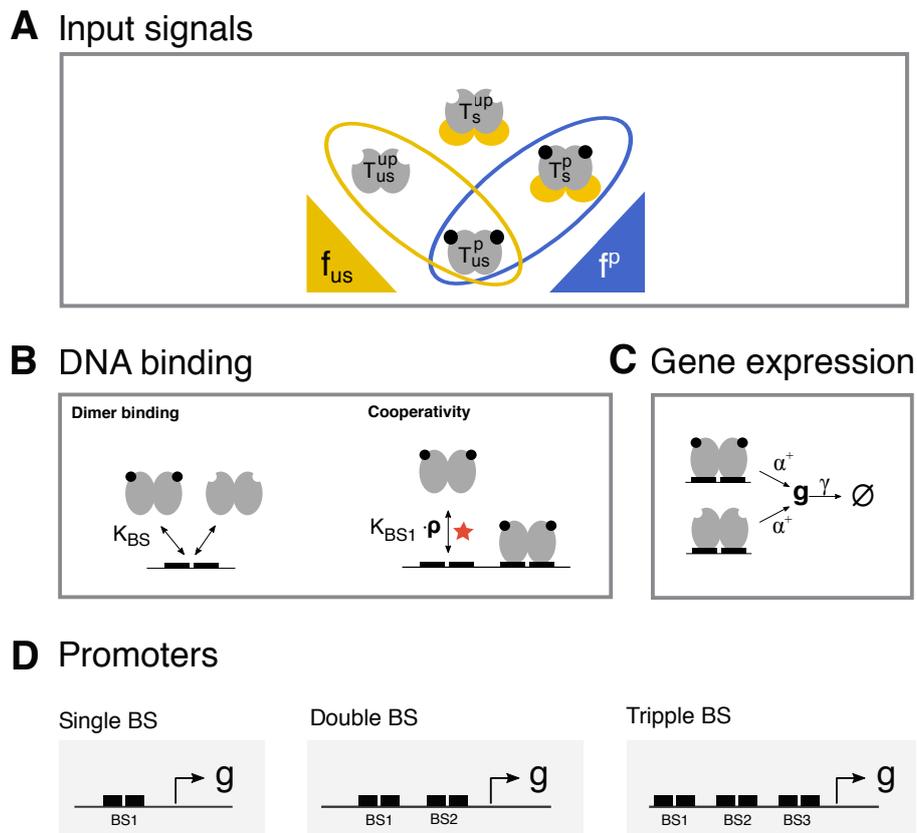
Figure 4.4: Models with phosphorylation-dependent cooperativity and sequestering-dependent binding. A: Inputs are the fractions of un-sequestered and phosphorylated TF. **B:** TF binding to any BS requires un-sequestering. Dimer binding occurs with the association constant K_{BS}. If any neighboring BS is already bound by T_{us}^p, the K_A is increased by the cooperativity factor ρ. **C:** Expression is phosphorylation-independent and occurs with rate α^+. Expression schemes are explained in Sec. 3.2. **D:** Gene expression from promoters with one, two and three BSs is investigated in this chapter.

Gene expression. I use the same rules for gene expression as in model 3: A promoter in state i expresses gene product g with rate α_i^+, while degradation occurs with rate γ (Fig. 4.4C), leading to the expression constant $\alpha_i = \alpha_i^+/\gamma$. The expression constants are independent of phosphorylation. The three expression schemes »all«, »proximal« and »any«, which determine the BSs

that have to be bound for expression, are also the same as for model 3. For a 1-BS promoter, all three schemes lead to the same logic: Expression can take place, if the BS is bound. In the 3-BS model, the schemes have the following expression constants:

»any« $\alpha_{\{t,x,x\}} = \alpha_{\{x,t,x\}} = \alpha_{\{x,x,t\}} = \alpha_{\{t,t,x\}} = \alpha_{\{t,x,t\}} = \alpha_{\{x,t,t\}} = \alpha_{\{t,t,t\}} = \alpha$

»proximal« $\alpha_{\{x,x,t\}} = \alpha_{\{t,x,t\}} = \alpha_{\{x,t,t\}} = \alpha_{\{t,t,t\}} = \alpha$

»all« $\alpha_{\{t,t,t\}} = \alpha$

Here, t signifies a bound BS. All other expression constants are $\alpha_i = 0$.

Promoters.   In this chapter I compare promoters with one (model 3a), two (model 3) and three (model 3b) BSs. All BSs are numbered from distal to proximal with respect to their position relative to the RNAP BS (Fig. 4.4D).

Relative impact of the phosphorylation signal

To compare the effects of both signals on gene expression, I measure the relative impact of phosphorylation, defined as

$$
ri(f^p) = \frac{\overbrace{\langle g_{f_{us}}(f^p = 1) - g_{f_{us}}(f^p = 0)\rangle_{f_{us}}}^{\text{average impact of } f^p}}{\underbrace{\langle g_{f^p}(f_{us} = 1) - g_{f^p}(f_{us} = 0)\rangle_{f^p}}_{\text{average impact of } f_{us}} + \underbrace{\langle g_{f_{us}}(f^p = 1) - g_{f_{us}}(f^p = 0)\rangle_{f_{us}}}_{\text{average impact of } f^p}}. \tag{4.2}
$$

The absolute impact of phosphorylation for a particular value of f_{us} is calculated as the difference between $\max_{f_{us}}(g)$ and $\min_{f_{us}}(g)$ which in this model is equal to $g_{f_{us}}(f^p = 1) - g_{f_{us}}(f^p = 0)$, i.e. the change in gene expression that phosphorylation can make for this value of f_{us}. As f_{us} can take different values, a uniformly spaced scan over f_{us} (5 values between 0 and 1) is performed to obtain the average impact. This value is then normalized by the sum of both the average impact of f_{us} (calculated analogously) and f^p. In all cases for which this measure was applied, the gene expression behavior was checked and found to be monotonic with respect to both f_{us} and f^p.

4.4   The relative impact of phosphorylation on expression from 1- and 2-BS promoters

First, I compared gene expression from a single BS (1-BS) and a double BS (2-BS) promoter by creating gene expression profiles. These are scans over the input conditions f^p and f_{us} (Fig. 4.5). Expression from a 1-BS promoter depends only on the extent of un-sequestered TF (Fig. 4.5A).

This is not surprising, since in this model, in accordance with the experiments (Sec. 4.2), phosphorylation only affects cooperative binding to two BSs, a setting that is not given here. Expression from a 2-BS promoter depends on both phosphorylation and un-sequestering: The TF must be un-sequestered in order to bind to the DNA and, in addition, phosphorylation increases binding via cooperativity (Fig. 4.5B). Thus, a fundamental qualitative difference between the two promoter classes becomes obvious: Only expression from 2-BS promoters depends on phosphorylation.
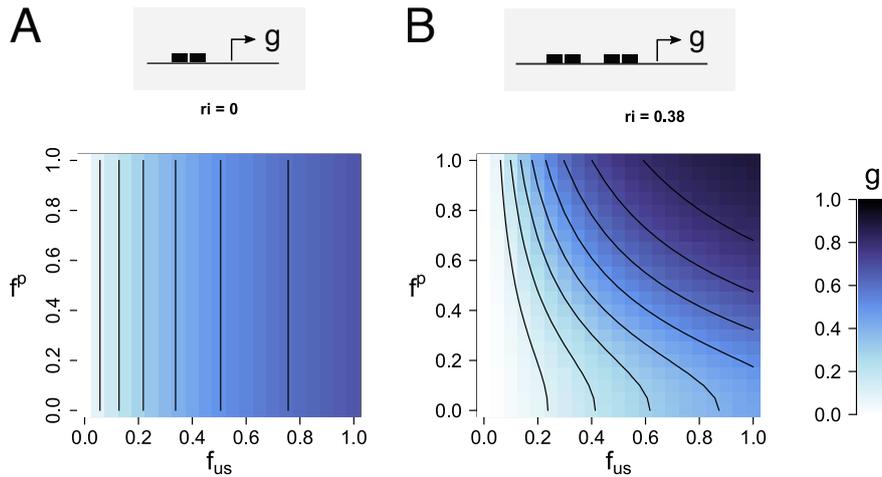
Figure 4.5: Model output. Gene expression outputs from **A:** a 1-BS and **B:** a 2-BS promoter are shown as 2-D scans over the input signals f_{us} and f^p. The relative impact of phosphorylation, ri (Sec. 3.2), is also given. Non-standard parameters: $[T_2] \cdot K_{BS} = 2$. Standard parameters: see Sec. 2.3.

A closer look on Fig. 4.5B shows another interesting outcome: At least for the applied parameter setting (see legend Fig. 4.5), expression from a 2-BS model is also dominated by f_{us}. This is logical from the mechanistic point of view, because un-sequestering is a strict requirement for binding, while phosphorylation only increases affinity under certain circumstances. However, as the phosphorylation signal is the one with higher specificity, this is not what I expect from the biological point of view. Instead, I expect that phosphorylation has a substantial impact on gene expression which would increase kin selection and thereby the payoff of QS-dependent cooperative behavior.

The question I will address in the following is therefore: What promoter features, if any, can increase the impact of phosphorylation over the impact of un-sequestering?
To be able to answer this question, I consider the relative impact of phosphorylation ri, which gives the average change of gene expression due to phosphorylation, normalized by the total impact of phosphorylation and un-sequestering (see Sec. 4.3). ri_a for a parameter set a is 1 when the output g depends entirely on f^p and ri_a is 0 if the output is completely independent of f^p. In our example, the expression from a 1-BS promoter is independent of phosphorylation, so that

ri = 0 (Fig. 4.5A). The expression from a 2-BS promoter with standard parameters is influenced by phosphorylation, but dominated by un-sequestering, leading to *ri* = 0.38 (Fig. 4.5B).

4.5 An analytic approximation of the gene regulatory function for better understanding

I applied a method by Ahsendorf et al. (2014), which allows the formulation of a gene regulatory function (GRF) that gives the steady state concentration of gene output as a function of the input signals and model parameters. The method involves representing the model as a weighted graph in which the possible DNA states are nodes and reactions that lead from one state to another are edges, the weights of which are defined by the reaction rates. The analytic solution comes at the assumption that DNA binding does not change the concentration of free TF, so that the reaction rates depend on a fixed parameter $[T_2]$, which linearizes them and simplifies the graph description. Apart from this assumption, the underlying representation is equivalent to the model described in Sec. 4.3. The graph then allows calculating the probabilities of all the DNA states and from these, a gene GRF can be set up.

I applied this method to the 2-BS model, and performed additional substitutions and transformations in order to give them a simple, understandable format. Details on the method and its application to the model are given in Sec. 2.4.

The concentration of gene output g for a 2-BS promoter in steady state equals

$$g = \frac{\overbrace{\alpha_{\{t,x\}} f_{us} c_1}^{\text{state } \{t,x\}} + \overbrace{\alpha_{\{x,t\}} f_{us} c_2}^{\text{state } \{x,t\}} + \overbrace{\alpha_{\{t,t\}} f_{us}^2 c_1 c_2 [1 + (f^p)^2 (\rho - 1)]}^{\text{state } \{t,t\}}}{\underbrace{1 + f_{us} c_1 + f_{us} c_2 + f_{us}^2 c_1 c_2 [1 + (f^p)^2 (\rho - 1)]}_{\text{normalization term}}}, \tag{4.3}$$

where $c_j = [T_2] \cdot K_{BSj}$ gives the relative concentration of TF for BSj. In the numerator, we find the contributions of the states with proximal, distal and both BSs bound, weighed by the respective expression constant α. The denominator is the sum of all possible states and therefore effectively a saturation term. The fraction saturates at 1 for high relative concentrations. The expression constants α_i for states i have the same unit as g and give the steady state concentrations of g that would be reached if the promoter was constantly in state i. The GRF emphasizes that only the double-bound state $\{t,t\}$ depends on f^p. A high contribution from this state to gene expression will thus positively affect the impact of phosphorylation. Understanding this will help interpret the simulation results which I will present in the following.

Cooperativity increases the impact of phosphorylation

An obvious parameter that increases the impact of phosphorylation is the cooperativity factor ρ. In a large scan over relative concentrations and expression schemes, for every tested parameter combination an increase in ρ lead to an increase in ri (not shown). In the GRF, ρ is a prefactor of the only f^p-dependent term in the numerator. Thus, increasing ρ automatically changes the weights of the contributions to gene expression in favor of the f^p-dependent term. Since the effect of changing ρ is always the same, I will not further consider this parameter in the following. It will be constant at $\rho = 10$ (standard parameter).

4.6 Restrictive expression increases the impact of phosphorylation
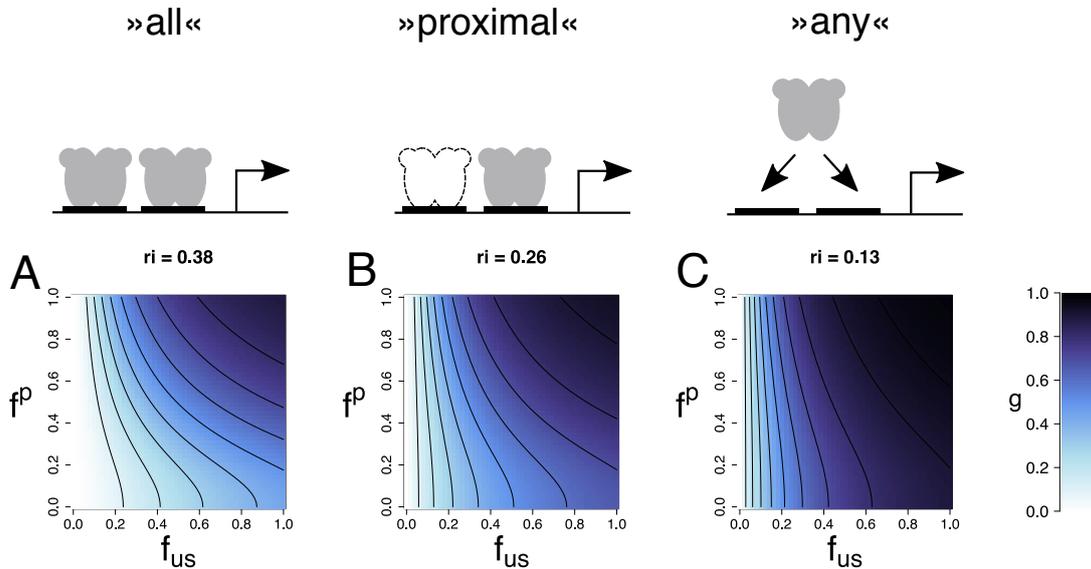
Figure 4.6: Gene expression schemes and the impact of phosphorylation. Upper line: Schematic drawings for the three tested generic expression schemes. From left to right: »all« BSs, the »proximal« or »any« BS have to be occupied for gene expression to take place with constant α. **A-C**: Gene expression profiles for the above expression schemes. Non-standard parameters: $c_1 = c_2 = 2$.

We do not know for sure which BSs have to be occupied in order for gene expression to take place. Therefore I compare three generic schemes: both BSs have to be occupied (»all«), only the proximal BS has to be occupied (»proximal«) or expression occurs as soon as any BS (»any«) is occupied by a TF (Fig. 4.6, upper line). In the GRF (Eq. 4.3), the three schemes are represented in the settings of α. The GRF suggests that stronger weightings of the contribution term for the double bound state $\{t, t\}$ through expression constants should increase the impact of phosphorylation. In other words, the more restrictive an expression scheme is to the double-bound state, the higher the expected impact of phosphorylation. I first tested this hypothesis for

the case of equal BSs affinities.

Maximum restriction to state $\{t,t\}$ is given in scheme »all«: In this scheme, $\alpha_{\{t,x\}} = \alpha_{\{x,t\}} = 0$ and only $\alpha_{\{t,t\}} = 1$. Indeed, this scheme yields the highest ri of the three tested schemes (Fig. 4.6A). Here, only the double-bound promoter state can contribute to expression and this state is supported by phosphorylation. However, the un-sequestering signal is a strict requirement for binding, so that f_{us} still dominates expression.

In scheme »any« the weights between the promoter states are distributed equally ($\alpha_{\{t,x\}} = \alpha_{\{x,t\}} = \alpha_{\{t,t\}} = 1$) and any occupied BS is enough to trigger gene expression. In this scheme the double bound state is not necessary for gene expression. Therefore, expression is strongly dominated by un-sequestering, yielding a much lower ri than in scheme »all« (Fig. 4.6C). Only at high f_{us}, phosphorylation starts to play a role.

The ri for scheme »proximal« lies between that of the other two schemes (Fig. 4.6B), which is in line with the scheme being intermediately restrictive.

In summary, the more restrictive an expression scheme is to the double-bound promoter state, the higher the impact of phosphorylation.

4.7 Asymmetric affinities increase the impact of phosphorylation

So far, equal affinities for both BSs were assumed. The next question is therefore, whether the affinity distribution can increase phosphorylation-dependence of gene expression. To answer this, I scanned over a large range of relative concentrations for both BSs and calculated the resulting ri (Fig. 4.7A-C).

First, the scans confirm that the schemes »all« and »proximal« (Fig. 4.7A,B) are more suitable for phosphorylation-impacted gene expression than scheme »any« (Fig. 4.7C), in so far that (1) the maximum ri is higher and (2) the affinity regimes of high ri are larger for these schemes.

Within the parameter ranges tested here, tuning of BS affinities can increase the ri up to ≈ 0.5 in schemes »all« and »proximal«, which corresponds to roughly equal contributions of phosphorylation and un-sequestering signal. Optimal combinations for both schemes are asymmetric affinities, meaning one high and one low affinity BS (Fig. 4.7D,E). Why is an asymmetric setup useful? As discussed in Sec. 3.4, an effective way to occupy a BS with the help of cooperativity is when another BS is bound per default in the absence of signal, which is achieved by high affinity. The BS to be occupied upon phosphorylation should have a lower affinity, so that the signal is necessary for occupancy. For scheme »all«, it does not matter which BS is of high

affinity, leading to two optimal designs (Fig. 4.7D). In scheme »proximal«, the BS to be occupied via cooperativity is BS2, because a high affinity BS2 would lead to immediate expression without the need of phosphorylation-dependent cooperativity. Thus, the optimal solution is one where c_1 is high and c_2 is low (Fig. 4.7E).

By contrast, the best possible design with respect to *ri* in scheme »any« are equal, intermediate affinities (Fig. 4.7F). Because of the ability of each occupied BS to recruit RNAP on its own, the only way to establish phosphorylation-dependence is simultaneous occupancy of both BSs (compare also Sec. 3.4).

In summary, in schemes »all« and »proximal« asymmetric affinities help increase the impact of phosphorylation when an expression-critical BS has a low affinity.
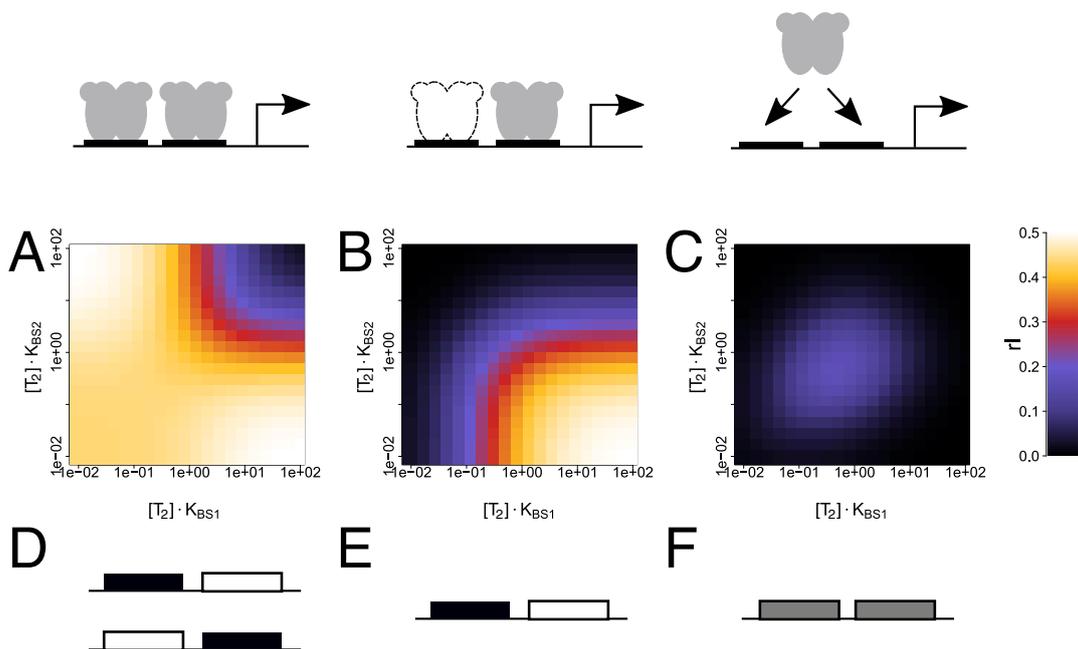
Figure 4.7: Asymmetric BS affinities increase the impact of phosphorylation. **A-C:** *ri* profiles. For each expression scheme, $[T_2] \cdot K_{BS}$ for both BSs were scanned over 20 values between 0.1 and 10. To calculate *ri* for a given combination, f^p and f_{us} were 2D-scanned from 0 to 1 (5 values, uniform sampling). **D-F:** The combination of relative concentrations that maximizes *ri* is displayed schematically. Light and dark BSs represent low and high relative concentrations.

4.8 The number of binding sites increases the impact of phosphorylation

Yet another variation between QS-dependent promoters is the number of BSs. Above, I already compared a 1-BS to a 2-BS promoter. The difference between them was of qualitative nature, because a single BS does not allow for cooperativity and thus a second BS is necessary for phosphorylation to affect gene expression. Now I ask whether additional BSs can help increase *ri*. To study their effect, I compared a 2-BS and a 3-BS model as an example (details: Sec. 4.3).

In the 3-BS model, as in the 2-BS model, each BS_j has an individual relative concentration c_j and cooperative binding of T_{us}^p requires at least one neighboring BS to be bound by T_{us}^p. The expression schemes are also analogous to the 2-BS model.
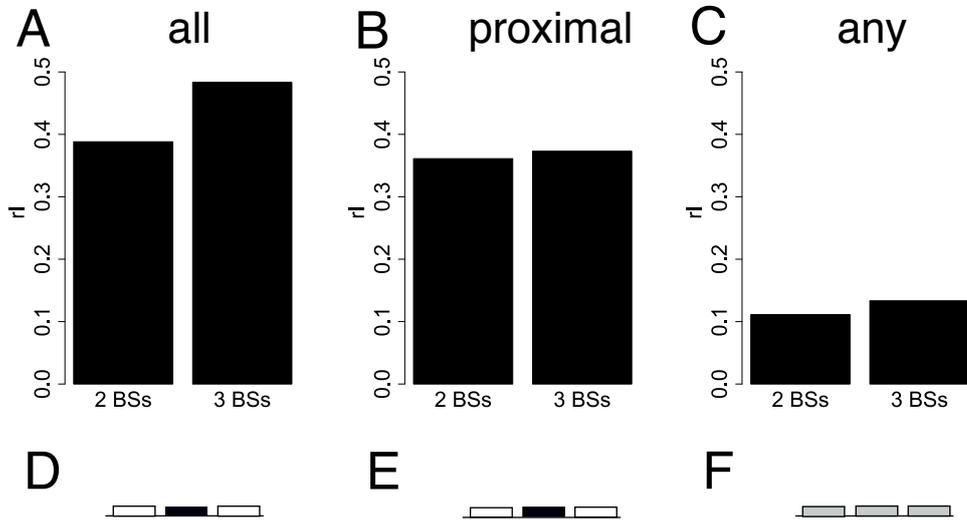

Figure 4.8: The number of binding sites. A-C: Maximum *ri* for 2- and 3-BS promoters. For each bar in the histogram, a double (2 BSs) or triple (3 BSs) scan over $[T_2] \cdot K_{BS}$ (10 log-uniformly distributed values between 0.2 and 5) for all BSs was performed and the maximum value is displayed. The cooperativity factor was set to $\rho = 5$. **D-F**: The sets of relative concentrations which maximize *ri* are shown schematically for every expression scheme. Black, gray and white BSs represent high, intermediate and low relative concentrations, respectively.

In all three expression schemes, the third BS increases the maximum possible impact of phosphorylation in a scan over relative concentrations (Fig. 4.8A-C). The effect of the third BS is largest for scheme »all«. However, these findings are only true, when the scan boundaries for the relative concentrations are narrowed down compared to those in Fig. 4.7, because otherwise, high values for *ri* are reached already with the 2-BS model (not shown).

Like for the 2-BS promoter, asymmetric affinities increase the impact of phosphorylation in schemes »all« and »proximal«: The combination of affinities that maximizes *ri* includes one high-affinity middle BS and two low affinity outer BSs (Fig. 4.8D,E).
Also for the »any« scheme, a third BS does not change the optimal design which consists of BSs of uniform, intermediate affinities (Fig. 4.8F). Here, the third BS simply helps to increase the overall cooperativity and thus facilitates the simultaneous occupancy of BSs.

In conclusion, a third BS is useful, especially when there are restrictions on the relative concentration that do not allow the 2-BS model to reach a maximum *ri*.

4.9   Summary and discussion

This comprehensive study sheds light on the question which features in the promoter architecture of QS-dependent genes in *B. subtilis* can maximize the impact of the ComX phosphorylation QS signal, compared to the CSF un-sequestering signal. I find that the more restrictive an expression scheme is to the double bound state, the higher the relative impact of phosphorylation. In all expression schemes, except »any«, asymmetric BS affinities further help increase the relative impact. Finally, in all tested expression schemes, additional BSs increase the relative impact.

In this chapter I investigated the same three generic expression schemes as in Ch. 3, even though some knowledge on ComA-dependent RNAP recruitment is available. From experiments, we know that the proximal BS (BS2) is essential for expression from ComA-dependent promoters (Fig. 2 in Wolf et al. (2016)), which argues against the scheme »any«. However, it cannot be ruled out that a stronger BS1 would compensate for the loss of BS2, which is why I still investigated the scheme. Against the scheme »all« speaks that expression from a 1-BS promoter takes place (Fig. 4.2A). Still, expression from a 2-BS promoter is much stronger, indicating that recruitment by two TFs might be necessary for full induction of a promoter, which is phenomenologically similar to scheme »all«.

The measure *ri* contains an implicit assumption on the distribution of input signals, as uniform scans from 0 to 1 are performed. It is possible that in the natural system, not all phosphorylation or un-sequestering states between 0 and 1 occur, or that they do not occur equally often (no uniform distribution), but that certain phosphorylation or un-sequestering states occur particularly often or rarely. Finally, the mutual distribution might play a role: For example, high un-sequestering states might be correlated with high phosphorylation states, because bacteria tend to produce either both signals or none of them. However, we can only speculate about the distributions, as natural conditions are hard to assess experimentally. Therefore, I consider the uniform scan over all theoretically possible values as the most general and therefore preferable approach.

Wolf et al. (2016) collected all known IR-DR sequences in QS-dependent promoters in *B. subtilis* and computed an idealized sequence of the IR-DR motif from these, using MEME analysis (Bailey et al., 2009). Given these results, one can count the number of deviations from the consensus sequence in all IR-DR promoter regions (Tab. 4.1). On average, the IR (which corresponds to BS1 in the 2-BS model) deviates from consensus by 2.0 bases, while the DR (BS2 in the 2-BS model) has 3.4 mismatches. Thus, the proximal BS (DR) is the more degenerate one. However, there is also large variation between the tested genes. The standard deviations of the number of mismatches are 0.9 and 2.0 for BS1 and BS2, respectively. For one gene, *rapF*, the relation is

reversed and BS1 shows more mismatches than BS2. In P_{rapA}, both BSs have one mismatch. For the other seven genes, the relation holds.

Table 4.1: Degenerate proximal BS in *B. subtilis* quorum sensing. For the known QS-dependent genes in *B. subtilis*, the numbers of deviations from the idealized sequence in the core module are given (data from Fig. 1A in Wolf et al. (2016)).

Gene	BS1	BS2
rapA	1	1
rapF	3	1
pel	1	2
srfAA	2	3
degQ	2	3
rapC	2	4
rapE	1	5
fapR	3	5
lutP	3	7
Average	**2.0**	**3.4**
sd	**0.9**	**2.0**

The fact that we see asymmetric BS affinities – the design predicted to maximize *ri* in two out of three expression schemes in this study – in the natural QS-dependent promoters supports the hypothesis that the impact of phosphorylation can be one part of an objective function in the regulation of these genes. If we assume that this is the case, then the observed degenerate DR in ComA-dependent promoters would suggest an expression scheme related to the »all« or »proximal« scheme – because only for these schemes, asymmetric affinities increase *ri*. Indeed, the fact that the naturally degenerate BS is almost always the proximal one, further hints at a »proximal«-like expression scheme which only leads to high *ri* when the degenerate BS is the proximal one. Finally, the observed promoter diversity could mean that the respective genes depend on phosphorylation to different extents.

We saw that restrictive expression, asymmetric BS affinities and additional BSs are features that increase the impact of phosphorylation over un-sequestering. We also know that at least the latter two features vary between natural QS-dependent promoters (see Tab. 4.1 and Sec. 4.1). This corroborates the hypothesis presented in Sec. 4.1, namely that these genes might weight differently between the two QS signals. If weighting is possible, then an evolutionarily plausible strategy could be to place genes that regulate costly processes under tighter control by phosphorylation. This way, their benefit would be more exclusive for close relatives. Such a type of regulation would be in line with evolutionary theory as presented in Sec. 4.1.

Evidence pointing towards this theory can be obtained from mapping the function of genes to their promoters' number of BSs. This relation is useful to look at, because the presented results indicate that the number of BSs is a feature that, independent of other promoter properties, always increases the impact of phosphorylation. In Tab. 4.2 I grouped the QS-dependent genes by their number of BS.

Table 4.2: Number of BSs and functions of ComA-dependent genes

	Genes	**Functions**
1 BS	ICE*Bs1*	ICE (jumping gene)
2 BSs	*rapA*, *C*, *E* and *F*	signaling
	degQ	(nutrient) degradative enzymes, antimicrobiotics
	lutP	lactate uptake
	fapR operon	control of membrane lipid synthesis
3 and more BSs	*pel*	processing of plant cell walls
	srfA	antibiotics production, swarming, biofilm formation, competence

The only known ComA-dependent gene with 1 BS – and therefore supposedly not under the influence of phosphorylation through the ComX pathway – is an integrative and conjugative element (ICE), called ICE*Bs1*. ICEs are transferred via horizontal gene transfer. Most of the time they reside in the host genome, but under certain conditions they excise, pass a copy to a new host through a mating pore and then reintegrate into the DNA (review: Johnson and Grossman (2015)). ICE*Bs1* of *B. subtilis* (Burrus et al., 2002) contains, apart from excision and transfer genes, a *rapI/phrI* operon, which is regulated by ComA (Wolf et al., 2016). RapI activates excision and transfer, while PhrI is secreted and re-imported by Opp and counteracts its cognate Rap (Auchtung et al., 2005, 2016). Thus, ComA activates excision and transfer in case of high bacterial density, while PhrI prevents excision and transfer in case of high density of cells that already contain ICE*Bs1*.

Since ICE*Bs1* brings no apparent benefit to its host, it might be more meaningful to think about the inherent purpose of the ICE, which is its own replication. From this point of view, generally high bacterial density (as coded for by all QS signals) is a useful criterion for replication attempts, while restriction of transfer to close relatives (as coded for by ComX) brings no advantage for the ICE.

Most ComA-dependent genes have a promoter consisting of the 2-BS core module. Their functions have in common that they are often connected to signaling or acquisition of nutrients: The *rap/phr* genes *rapC* and *F* take part in QS signaling (Solomon et al., 1996; Bongiorni et al., 2005) and *rapA* and *E* in sporulation signaling (Perego et al., 1994; Perego, 1997; Jiang et al., 2000). The *fapR* operon contains a master regulator for membrane lipid synthesis which can

induce changes in the membrane composition as an adaptation to environmental conditions (Schujman et al., 2003). *lutP* codes for a lactate permease and thus enables growth on this nutrient source (Kunst et al., 1997; Chai et al., 2009). *degQ* leads to the production of secreted degradative enzymes and antibiotics (Kunst et al., 1975; Yang et al., 1986; Amory et al., 1987; Tsuge et al., 1999, 2005).

Two operons, *pel* and *srfA*, are known to be regulated by promoters containing clusters of BSs (Nakano et al., 1991b; Wolf et al., 2016). *pel* codes for pectate lyase C, an enzyme that is secreted and extracellularly degrades pectin, a major plant cell wall component (Nasser et al., 1993).

The *srfA* operon contains

- *srfAA*, *srfAB*, *srfAC* and *srfAD*: four genes necessary for the non-ribosomal surfactin production (Nakano et al., 1988),

- *comS*, a competence regulator (D'Souza et al., 1994; Hamoen et al., 1995) and

- *yxcA*, a gene with unknown function, which might be related to stress response (Zuber, 2001).

Surfactins, a group of antibiotics, are produced by many *B. subtilis* strains as well as closely related *Bacilli* (Stein, 2005). The *B. subtilis* surfactin is a very powerful surfactant (Carrillo et al., 2003) and is effective against various bacteria, viruses and fungi, as well as against erythrocytes (Singh and Cameotra, 2004). It is also a strict, but not sufficient requirement for swarming (Kearns et al., 2004) and important for structuring biofilms (Branda et al., 2001). *B. subtilis* is a soil bacterium and often found in symbiosis with plants (Berg, 2009), where it is thought that the bacterium protects the plants against pathogens in exchange for nutrients. In this context, Bais et al. (2004) could show that surfactin is necessary both for biofilm formation of *B. subtilis* on *Arabidopsis* and for protection of the plant against infection by *Pseudomonas syringae*. Within *srfAB1*, but out of frame, lies *comS*, which is necessary for the expression of late competence genes (D'Souza et al., 1994; Hamoen et al., 1995).

The only QS-dependent gene with presumably no intra-species specific control is a jumping gene, of which it can be speculated that it integrated at a site suitable for general density-dependent replication and for which discrimination between strains comes with no apparent benefit. Virtually all other known QS-dependent genes display at least two BSs and are therefore to some extent under the control of increased kin recognition. The distinction between two and more BSs was further enlightening. While processes regulated by genes with two BSs are more basic and include mostly signaling and acquisition of nutrients, the more sophisticated, higher level functions like antibiotic production, swarming, competence or plant interaction are regulated by genes that are under control of three or more BSs. The number of BSs therefore

appears to be correlated with the complexity of the respective genes' functions, suggesting that elaborate behaviors are indeed placed under tighter intra-species QS control than the more basic ones.

A reason why this relation evolved could be that complex behaviors are costly processes, the benefits of which should therefore only be available for kins. Indeed, there are arguments supporting a correlation between costly and complex behavior. First of all, simple processes like nutrient acquisition bring direct benefits, i.e. food, while symbiosis with plants or swarming might not. In addition, a study by Oslizlo et al. (2014) strongly suggests that production of surfactin comes with large metabolic cost. They show that overexpression of *srfA* leads to a decreased fitness level compared to the wt. Further, they provide experimental evidence pointing to surfactin production-induced metabolic burden as the cause of decreased fitness.

Chapter 5

Multiplexing of *B. subtilis* quorum sensing signals

5.1 Introduction

In Ch. 4 I looked at the processing of the two quorum sensing (QS) signals ComX and Phr in *B. subtilis* at the level of single promoters. I also discussed that it would be useful to know the distribution of these signals in natural environments, for the task of designing a promoter towards processing the signals in a specific way. Unfortunately, there are no quantitative measurements on this in the literature. This gave the motivation for the work presented in this chapter, which proposes a synthetic multiplexing system composed of two QS-dependent promoters. The aim is to infer the two QS signals from the multiplexing system's output and thus help to study their occurrence *in vivo*.

Multiplexing in the technical sense refers to a process in which several signals are pooled and transmitted over a common medium (Ohm and Lüke, 2010). Subsequently, they are restored, which is called *demultiplexing*. While in engineering systems are man-made and well-defined, biological pathways are often highly interconnected and not entirely resolved. Therefore, I will now define criteria for multiplexing, which I consider useful in a biological context. I will consider pathways that exhibit an hourglass structure, meaning that signals converge to and from a single entity, which can, for instance, be a molecule or a cell. For successful multiplexing, (a) inference about the input should be possible from the output. In both technical and biological systems, inference is never perfect and inevitably comes with uncertainty. (b) The outputs must exhibit orthogonality, that is, different outputs should contain information about different inputs. As for inference, also orthogonality could be fulfilled only partly, when outputs contain overlapping information about the inputs.

From an operational point of view, "perfect" multiplexing and decision making can be seen as two ends of a spectrum. In decision making, which might also occur in hourglass-structured

63

pathways, the inputs are processed to a common result. This can then influence several outputs. However, all of the outputs are results of the same decision, for example, parts of developmental programs.

In the *B. subtilis* QS pathway an hourglass structure is given, because the ComX and Phr signals get modulated onto the same TF – in the form of its un-sequestering and phosphorylation state – and the TF can then activate several promoters. In the previous chapter we saw that, depending on their architecture, these promoters are expected to be impacted by the two signals to different extents. This indicates that the orthogonality criterion is likely to be fulfilled. Finally, *in vitro* studies suggest that binding to a 1-BS promoter depends entirely on one of the signals (Phr) (Sec. 4.2), implying that inference about the Phr signal should be possible from the output. The remaining question to be considered theoretically in this chapter is whether also the ComX signal could be decoded. As experiments (Sec. 4.2) suggest that a 2-BS promoter's output depends on both inputs, I propose a multiplexing setup with two promoters that drive two distinct reporter genes. Here, YFP and CFP are chosen as an example (Fig. 5.1). These two reporters could be transformed into the same bacterial cell. The scope for this chapter will be to predict, using a model that represents these two promoters, if and under which circumstances such a setup could be useful as a multiplexing device for experimental purposes.
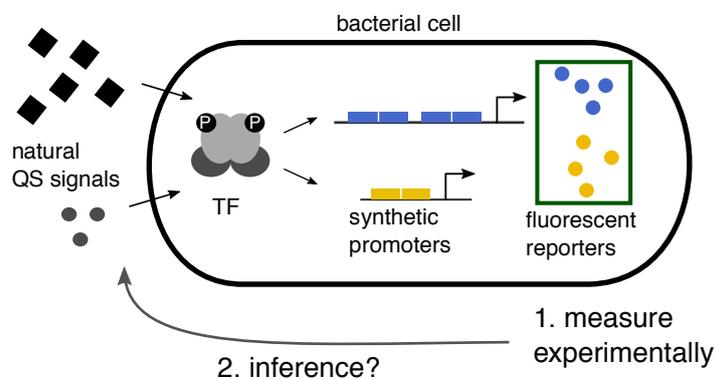
Figure 5.1: Proposed multiplexing system. The idea is to learn about the natural occurrence of *B. subtilis* QS signals by setting up a synthetic measuring device. This includes the introduction of two synthetic ComA-dependent promoters with one and two BSs, fused to fluorescence reporters. The question of this chapter is, under which circumstances inference of the natural signals from the experimentally measured fluorescence output is possible.

5.2   Model for the proposed multiplexing system

To model the proposed multiplexing system, I used the same modeling setup as in Ch. 4, which describes expression from two generic promoters, one with a single BS and another with a double BS promoter. The gene outputs are named Y and C in this chapter, because they rep-

resent two distinct reporters, for example YFP and CFP. Both models feature phosphorylation-dependent cooperativity, which appears to be the case in QS-dependent promoters in *B. subtilis* (Sec. 4.2).
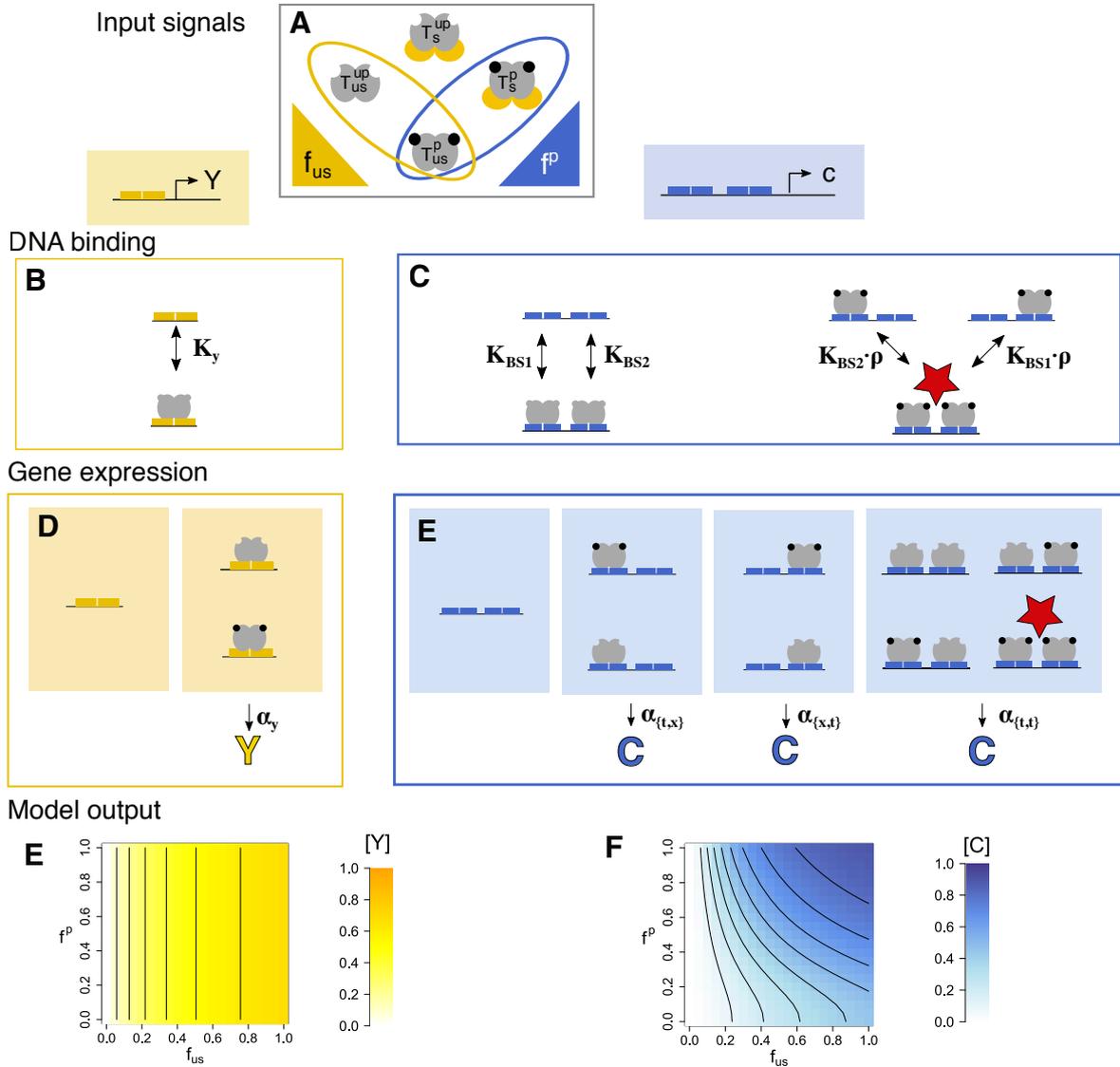

Figure 5.2: Models describing the hypothetic multiplexing system. A: Input signals are the fractions of phosphorylated and un-sequestered TF. **B**: DNA binding module for a 1-BS promoter (promoter Y). **C** DNA binding for a 2-BS promoter (promoter C). **D,E**: Gene expression for promoter Y and C. **E,F**: Model outputs as functions of the two input signals. Same computations as in Fig. 4.5.

The model input describes the phosphorylation and sequestering state of the TF dimer. The signals

$$f_{us} = \frac{[T_{us}]}{[T_2]} \qquad \text{and} \qquad f^p = \frac{[T^p]}{[T_2]} \tag{5.1}$$

65

give the fractions of un-sequestered and phosphorylated TF, respectively (Fig. 5.2A).

DNA binding to a single BS is independent of phosphorylation, but requires un-sequestering. For promoter Y, the association constant is K_y (Fig. 5.2B), for promoter C, the K_A depends on the BS identity: The BS distal to the transcription start site has the affinity K_{BS1}, the proximal BS has the affinity K_{BS2} (Fig. 5.2C, left). Cooperativity requires at least two BSs, which is why it can only occur at promoter C. If two TFs bind sequentially, then cooperativity is modeled as an increase in K_A by the cooperativity factor ρ for the second TF that binds (Fig. 5.2C, right).

Gene expression depends on which BS is bound, which I refer to as the promoter's functional state. At promoter Y, gene expression occurs with rate α_y^+, as soon as the BS is occupied (Fig. 5.2D). At promoter C, occupancy of BS1, BS2 or both leads to expression with rates $\alpha_{\{t,x\}}^+$, $\alpha_{\{x,t\}}^+$ and $\alpha_{\{t,t\}}^+$. Gene products Y and C are globally degraded with rate γ, so that the expression constants $\alpha = \alpha_i^+/\gamma$ give the average concentrations of gene product for promoters that are constantly in state i.

Approximation error

In order to quantify how well an input signal (in this case f^p) can be inferred from a simulation output, I use the approximation error E, which describes the discrepancy between estimated and simulated values for f^p. For a particular parameter set a, E_a is calculated by scanning over the inputs f^p and f_{us} (for both, 11 uniformly spaced values between 0 and 1 are tested). For each input combination, a deterministic simulation is performed, yielding steady state concentrations for $[Y]_a(f_{us})$ and $[C]_a(f_{us}, f^p)$. The according estimate $\hat{f}^p\left([Y]_a(f_{us}), [C]_a(f_{us}, f^p)\right)$ is retrieved by applying Eq. 5.11. The estimation error for a particular input combination is then given by $\Delta E_a(f_{us}, f^p) = |f^p - \hat{f}^p|$ and the overall estimation error is calculated by averaging over both inputs:

$$E_a = \frac{\langle \Delta E_a \rangle_{f_{us}, f^p}}{\langle f^p \rangle_{f_{us}}} \tag{5.2}$$

By normalizing to $\langle f^p \rangle_{f_{us}}$, E_a is given as fractional deviation.

When additive noise is assumed, the above procedure for calculating $\Delta E_a(f_{us}, f^p)$ is repeated $n = 10$ times, where each time a different noise term is added on the simulation outcome of C and Y. This means, $[Y]_a(f_{us}) + \eta_{y,i}$ and $[C]_a(f_{us}, f^p) + \eta_{c,i}$ are used in the calculation of the estimate $\hat{f}^p$. The noise terms η are drawn from a normal distribution with mean $\mu = 0$ and standard deviation $\sigma^2 = 0.001$.

## 5.3 Decoding ability of the model
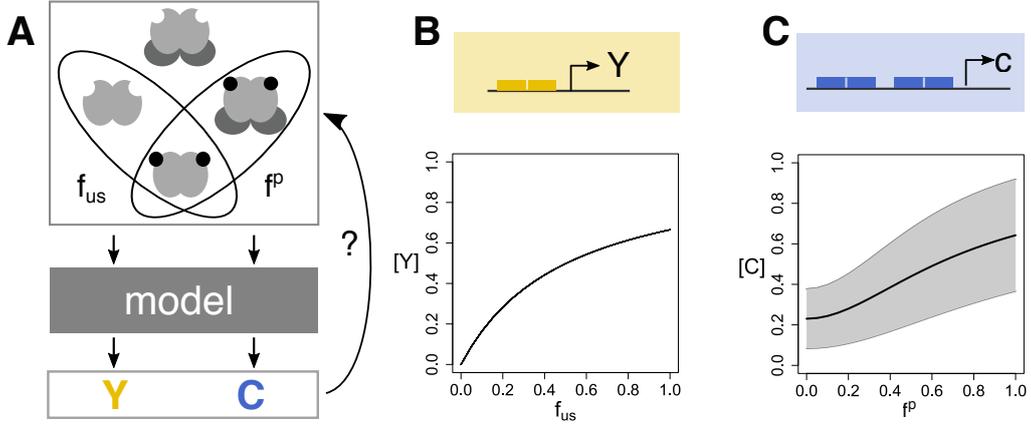
Figure 5.3: Un-sequestering can be decoded from Y, but phosphorylation cannot be decoded from C. **A:** The decoding idea in this chapter is to infer the input signals f_{us} and f^p from the model outputs Y and C. **B:** f_{us} can be decoded from $[Y]$. For 21 uniformly spaced values of f^p and f_{us} between 0 and 1, an $f_{us} - [Y]$ curve is plotted. All curves are identical. **C:** For every tested value of f^p, 21 values of f_{us} (uniformly spaced between 0 and 1) were scanned and the mean value (solid curve) and standard deviation (gray area) for $[C]$ (simulation with scheme »all«) are shown. Parameters: $c_1 = c_2 = c_y = 2$.

The overall question is whether the input signals can be decoded from the outputs Y and C (Fig. 5.3A). Therefore, a reasonable starting point is to check whether there is a chance to directly do so. When the input-output relation between f_{us} and Y (Fig. 5.2E) is known, measuring the output Y allows inference of f_{us}. There is no uncertainty caused by f^p, because f^p has no influence on Y (Fig. 5.3B). In other words, the $f_{us} - [Y]$ curve is the same for all values of f^p. On the other hand, C depends on both f_{us} and f^p (Fig. 5.2F). Therefore, f^p cannot be unambiguously determined from C, because the additional influence of f_{us} leads to uncertainty (Fig. 5.3C).

5.4 Exact decoding of the input signals

Since the signal f^p cannot be directly decoded from the input-output curve, I used the analytic solution presented in Sec. 4.5, which gives the steady state concentration of gene output as a function of the input signals and model parameters. I also applied the same framework to analogously derive the gene regulatory function (GRF) for the 1-BS model (Sec. 2.4). The idea is that the GRFs allow to express f^p as a function of C and Y, by simply building their reverse. The resulting GRFs for Y and C are

$$[Y] = \frac{f_{us} \cdot c_y}{1 + f_{us} \cdot c_y} \alpha_y \qquad (5.3)$$

and

$$[C] = \frac{\overbrace{\alpha_{\{t,x\}} \cdot f_{us} \cdot c_1}^{\text{BS1 bound}} + \overbrace{\alpha_{\{x,t\}} \cdot f_{us} \cdot c_2}^{\text{BS2 bound}} + \overbrace{\alpha_{\{t,t\}} \cdot f_{us}^2 \cdot c_1 \cdot c_2 \cdot [1+(\rho-1)(f^p)^2]}^{\text{both BSs bound}}}{\underbrace{1 + f_{us}(c_1 + c_2) + f_{us}^2 \cdot c_1 \cdot c_2 \cdot [1+(\rho-1)(f^p)^2]}_{\text{saturation term}}},$$

(5.4)

respectively. Eq. 5.3 confirms that Y depends on the un-sequestering signal f_{us} only and not on the phosphorylation signal. The function consists of a term for average promoter occupancy, multiplied by α_y, the average concentration of Y for full occupancy. The occupancy term contains a saturation element in the denominator, so that at high relative TF concentrations, occupancy approaches 1.

The GRF for C is more complex, since it contains contributions from different functional promoter states, i.e. $\{t,x\}$ (BS1 occupied), $\{x,t\}$ (BS2 occupied), and $\{t,t\}$ (both BSs occupied). The contributions are weighted by the respective expression constants α, which determine the expression scheme. We see in Eq. 5.4 that all contributions depend on the un-sequestering signal and only the contribution from fully occupied promoter depends on the phosphorylation signal. Since Eqs. 5.3 and 5.4 constitute two equations with two variables (f_{us} and f^p), they can now be solved for the two signals. For f_{us}, the solution is

$$f_{us} = \frac{[Y]}{K_y \cdot [T_2] \cdot (\alpha_y - [Y])}.$$

(5.5)

The general solution for f^p, which is valid for all expression schemes, is

$$f^p = \frac{\sqrt{Y^2(K_y(K_1' + K_2') - K_{12}' + C(K_1 - K_y)(K_2 - K_y)) + \alpha_y K_y Y(-K_1' - K_2' + C(K_1 + K_2 - 2K_y)) + \alpha_y^2 C K_y^2}}{\sqrt{K_1 K_2 (\rho - 1) Y^2 (\alpha_{12} - C)}}.$$

(5.6)

In this equation, the abbreviations

$$
\begin{array}{llll}
K_1 = K_{BS1} & K_1' = \alpha_1 \cdot K_1 & \alpha_1 = \alpha_{\{t,x\}} & C = [C] \\
K_2 = K_{BS2} & K_2' = \alpha_2 \cdot K_2 & \alpha_2 = \alpha_{\{x,t\}} & Y = [Y] \\
& K_{12}' = \alpha_{12} \cdot K_1 \cdot K_2 & \alpha_{12} = \alpha_{\{t,t\}} &
\end{array}
$$

are used in order to save space. In the following, these solutions will be called *exact decoding*. Note that the exactness is not given with respect to the simulation results, but with respect to the GRFs (Eqs. 5.3 and 5.4), which approximate the simulation results. The exact decoding for f_{us}

gives the un-sequestering signal as a function of Y (the measured output) and requires knowledge on two system parameters, α_y and K_y. The phosphorylation signal is given as a function of two measured outputs, Y and C, and its solution requires knowledge on eight parameters from both expression systems. Comparison of the exact decoding and the simulation outcome verifies its accuracy (Fig. 5.4).
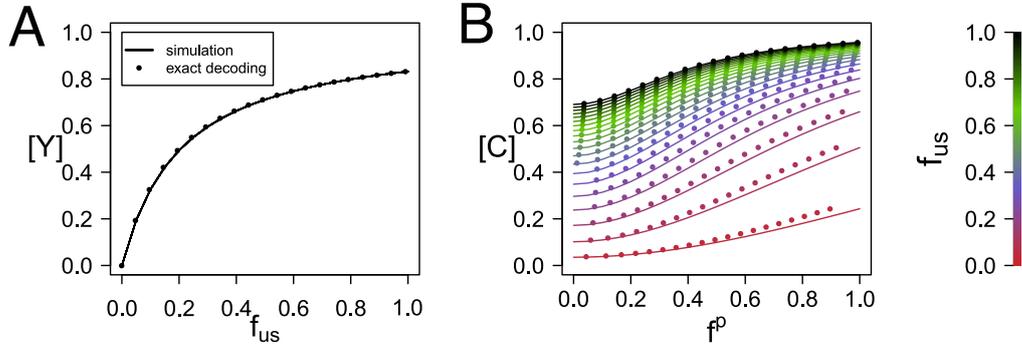

Figure 5.4: Verification of the exact decoding. A: A 2D-scan over f_{us} and f^p was performed and the simulation results (scheme »all«) for $[Y]$ (**A**) and $[C]$ (**B**) are plotted as lines. In B, f_{us} is color-coded. For comparison, the simulation results for $[Y]$ and $[C]$ were used to obtain the exact decodings for f_{us} and f^p (shown as dots), using Eqs. 5.5 and 5.6. The slight differences seen between simulation and decoding are due to the fact that the decoding is derived from the GRFs for C and Y. Parameters: $c_1 = c_2 = c_y = 2$.

5.5 Parameter reduction by approximated decoding

The exact decoding for f^p comes at the major disadvantage that its practical application requires estimates on eight parameters. Four of them (α_y, $\alpha_{\{t,x\}}$, $\alpha_{\{x,t\}}$ and $\alpha_{\{t,t\}}$) describe the expression scheme which, as explained above, is not known. The others are BS affinities (K_{BS1}, K_{BS2} and K_y), that can be measured *in vitro*, and the cooperativity factor, which could be estimated by comparing binding to single and double BSs. However, the retrieved values always come with measuring errors, which propagate through the calculation of f^p. It therefore seems favorable to find an estimate for f^p that depends on fewer parameters. To this end, my approach was to first find condensed approximations for the GRFs (Eqs. 5.3 and 5.4), so that solving for the input signals would lead to a simpler equation than Eq. 5.6, which would require fewer parameters for experimental application.

Simplifying the GRFs involved the following assumptions:

1. Expression from promoter C can be described by one of the three generic schemes, »all«, »proximal« and »any«.

2. Both promoters operate in the non-saturated regime.

3. In promoter C, expression from the fully-occupied promoter dominates, that is, the promoter states with just one BS bound do not significantly contribute to gene expression (even if they formally contribute, as is the case in schemes »proximal« and »any«).

Considering the three expression schemes »all«, »proximal« and »any« splits the problem into three cases, while in each case only one expression constant remains ($\alpha_c = 1$).

The second assumption is that in the non-saturated regime, the saturation term of the GRFs would lose influence so that it could be approximated by 1:

$$\lim_{t_1 \to 0, t_2 \to 0} [C] \approx \frac{\alpha_1 t_1 + \alpha_2 t_2 + \alpha_{1,2} t_1 t_2 [1 + (\rho - 1)(f^p)^2]}{1} \tag{5.7}$$

$$\lim_{t_y \to 0} [Y] \approx \alpha_y \cdot t_y \tag{5.8}$$

Here, I summarize $t_j = f_{us} \cdot c_j$. The validity of the assumption was tested by 2D Taylor expansion for all three expression schemes, with c_1, c_2 and c_y around 0 (Sec. 2.4). The results are:

$$
\begin{array}{ll}
GRF_Y & [Y] \approx \alpha_y \cdot t_y \\
GRF_C, \text{»all«} & [C] \approx \alpha_c \cdot t_1 \cdot t_2 [1 + (\rho - 1)(f^p)^2)] \\
GRF_C, \text{»proximal«} & [C] \approx \alpha_c \cdot \left(t_2 + (f^p)^2 t_1 \cdot t_2 (\rho - 1) - t_2^2 \right) \\
GRF_C, \text{»any«} & [C] \approx \alpha_c \cdot \left(t_1 + t_2 + t_1 \cdot t_2 (\rho - 1)(f^p)^2 - t_1 t_2 - t_2^2 - t_1^2 \right)
\end{array}
\tag{5.9}
$$

Grey terms indicate Taylor correction terms which I expect to approach zero in the limit. The Taylor expansion shows that, indeed, in the non-saturated regime, the denominators of Eqs. 5.4 and 5.3 vanish, while the numerators largely remain.

To further simplify Eqs. 5.9, I assume that expression from the fully-occupied promoter dominates, so that the contribution terms for single BSs bound in Eq. 5.9 can be neglected. This leads to the approximations and conditions given in Tab. 5.1.

Table 5.1: Approximations for gene regulatory functions.

	Approximation	Condition for assumption 3
»all«	$[C] \approx \alpha_c \cdot t_1 \cdot t_2 [1 + (\rho - 1)(f^p)^2)]$	no further condition needed
»proximal«	$[C] \approx \alpha_c \cdot t_1 \cdot t_2 (f^p)^2 (\rho - 1)$	$t_1 (\rho - 1) >> 1$
»any«	$[C] \approx \alpha_c \cdot t_1 \cdot t_2 (f^p)^2 (\rho - 1)$	$(\rho - 1) t_1 t_2 >> t_1 + t_2$

The approximations (Tab. 5.1) can be solved for f^p and yield

$$\widehat{f^p} = \sqrt{\underbrace{\frac{[C]}{[Y]^2} \frac{K_y^2}{K_1 \cdot K_2} \frac{\alpha_y^2}{\alpha_c} \frac{1}{(\rho-1)}}_{\phi} - \underbrace{\frac{1}{\rho-1}}_{\rho'}} \qquad \text{in scheme »all« and} \qquad (5.10)$$

$$\widehat{f^p} = \sqrt{\frac{[C]}{[Y]^2}} \sqrt{\underbrace{\frac{K_y^2}{K_1 \cdot K_2} \frac{\alpha_y^2}{\alpha_c} \frac{1}{(\rho-1)}}_{\phi}} \qquad \text{in scheme »proximal« and »any«.} \qquad (5.11)$$

These approximations still contain six parameters. However, compared to the exact decoding (Eq. 5.6), in schemes »proximal« and »any« they are all summarized to a single scalar, which I call ϕ. The advantage here is that ϕ could be treated as a phenomenological constant, which is individually measured for the expression system of interest. In scheme »all« (Eq. 5.10) there is an additional constant $\rho' = 1/(\rho-1)$. Interestingly, for sufficiently large values of ρ, this value will approach 0, in which case Eq. 5.11 can also be used for scheme »all«.

Eq. 5.11 implies that f^p can be approximately inferred from $[C]/[Y]^2$. Indeed, a mapping of $[C]/[Y]^2$ to f^p (Fig. 5.5B) comes with a much lower standard deviation than the $[C]$-f^p mapping (Fig. 5.5A). The intuition here is that by dividing by $[Y]^2$, the impact of f_{us} on $[C]$ can be roughly eliminated, so that $[C]/[Y]^2$ mostly depends on f^p.
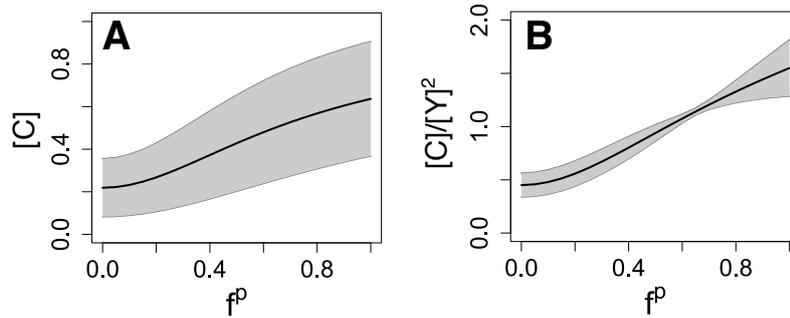
Figure 5.5: The phosphorylation signal can be inferred from a combination of C and Y. A: For comparison, Fig. 5.3B is shown again. **B**: The figure was created in the same way as A, but $[C]/[Y]^2$ instead of $[C]$ was plotted.

5.6 Design requirements for high quality decoding
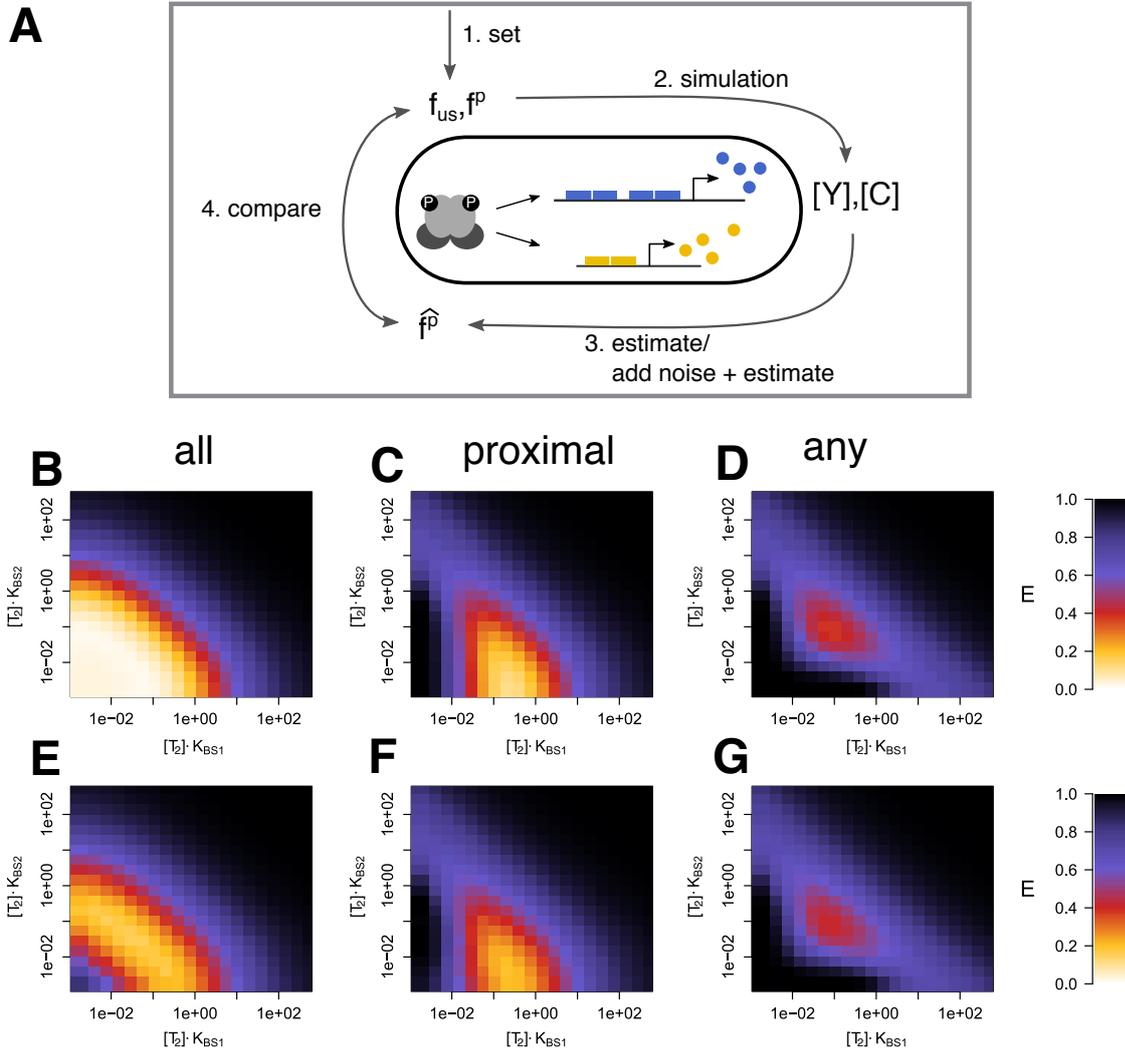
Figure 5.6: Design requirements for approximated decoding. **A**: Measurement of decoding error: Values for f_{us} and f^p are set and a simulation subsequently yields $[Y]$ and $[C]$. From these, $\hat{f}^p$ is estimated using Eq. 5.11 and compared to the initial setting of f^p. Details in Sec. 5.2. **B-D**: Estimation error E for schemes »all«, »proximal« and »any«. $c_y = 0.045$, $\rho = 500$. **E-G:** Estimation error under the assumption of additive noise. Parameters: $N = 10$, $sd = 0.001$, $c_y = 0.045$, $\rho = 500$.

To more precisely determine the parameter regimes in which the approximation is valid, I measured the decoding error E, which is the relative deviation of the estimated $\hat{f}^p$ from the "real" f^p. E was obtained by simulating with predefined inputs (f_{us}, f^p) and then comparing the inputs with the resulting estimates (Fig. 5.6A). For a particular parameter set, I averaged the deviations over different input values (Details are given in Sec. 5.2).

As expected from the assumptions that lead to the approximation for $\hat{f}^p$, low values for K_y and high values for ρ lead to the lowest decoding errors in all expression schemes (not shown). I therefore set these two parameters to $c_y = 0.045$ and $\rho = 500$ an then looked at the BSs design

of promoter C.

In schemes »all« and »proximal«, estimation errors close to 0 can be reached within the tested parameter regions. For scheme »all«, the main requirement for a good approximation of f^p is the unsaturated regime of both BSs (Fig. 5.6B), confirming Tab 5.1. In scheme »proximal«, an additional requirement is $c_1(\rho - 1) > 1$ (Tab. 5.1), which increases the estimation error in the low-affinity region of BS1 and thus limits the regime of good approximation compared to scheme »all« (Fig. 5.6C). In scheme »any«, also the low-affinity region for BS2 is not suitable for the approximation, leaving only a confined regime of equal BS affinities in which the approximation could be applied with low quality (Fig. 5.6D). The assumption in Tab. 5.1 $((\rho - 1)t_1t_2 >> t_1 + t_2)$ requires that both t_1 and t_2 are >1, so that multiplying t_1 and t_2 leads to larger values than adding them up. This is in conflict with the non-saturation assumption.

The deterministic simulations that were used so far for calculating the estimation error neglect measurement noise that has to be expected in *in vivo* experiments. To get an idea of its impact on the quality of approximation, I applied additive noise on the simulation outcomes and recalculated the estimation error E (Fig. 5.6E-G) (method details: Sec. 5.2). The results are qualitatively similar, but show increased non-functional areas, especially in the low-affinity regimes. These results point towards optima in BS affinities in all three expression schemes, due to a trade-off between susceptibility to noise and good fulfillment of the approximation's assumptions (non-saturated regime).

5.7 Summary and discussion

By deriving an analytic solution for the model steady state, I first demonstrated that the information of both input signals is contained in the outputs. Second, I found an approximated solution for the inputs which requires only one phenomenological parameter and which I therefore consider more useful for experimental purposes. Last, I tested the validity of the approximation and found design criteria for the promoters under which the best performance is expected.

The proposed approximated solution for the phosphorylation signal is in principle valid for all three tested expression schemes, suggesting that it would also hold for intermediates of these. However, the schemes »all« and »proximal« appear more suitable for the task. Differences exist in the design requirements that have to be met. For scheme »any«, the model predicts that relative concentrations, i.e. BS affinities, have to be very precisely tuned in order to yield a low approximation error. As mentioned earlier in this work, engineering BS affinities is generally feasible through point mutations (Stormo and Fields, 1998). What I consider the two most critical aspects are the requirement of a high cooperativity factor, which cannot easily be engineered, and the susceptibility to noise that could impair reliable statements on the two signals:

In the simulations shown in Fig. 5.6E-G, quite low additive noise terms were used, which still lead to often detrimental effects on the quality of decoding. However, it should be noted that for these calculations, a very basic assumption of additive measurement noise was made, because the model as such is not built for the study of stochastic effects of any kind (to be further discussed in Sec. 6.3). The aim of the calculations was to call attention to the expected trade-off between susceptibility to noise, which requires high relative TF concentrations, and the fulfillment of the non-saturation assumption, which requires low relative concentrations. The results should be considered as being of qualitative nature. The tolerance to noise and thus the exact optima presumably depend on the absolute concentrations of TF and DNA in bacterial cells and the expression strength (here modeled as α), as well as kinetic factors, all of which were not explicitly treated in this study. Therefore, I conclude that this study qualitatively predicts an optimal BS affinity, due to the trade-off mentioned above, but cannot predict quantitatively which amounts of noise will be tolerated by the experimental system. If noise causes problems in the experimental set-up, a method to reduce its effect on the quality of decoding could be population averaging.

The main requirement for a multiplexing device as suggested here is a TF with two signal-induced modifications that fulfill two different functions (or modes of activation). Even though I am not aware of other examples of this type, this criterion might help identify such TFs.

So far, multiplexing was discussed as a potential synthetic device to learn about the distribution of natural QS systems. Having demonstrated in theory that multiplexing is thinkable based on ComA and two ComA-dependent promoters, another question is, whether it also exists in nature.

In a multiplexing system like the one suggested here, where two signals are modulated onto one molecule, an obvious advantage could be saving material. This holds as long as the modifications of the TF are less costly than the production of a second TF. As another reason, I speculate that downstream computations might work better when both signals are on one molecule, because then there is only one messenger concentration which is subject to noise, rather than two. For comparison, it has also been proposed that the amount of phosphorylated RR is more robust to fluctuations in the amount of HK and ATP, when both phosphorylation and de-phosphorylation are executed by the HK (as opposed to an additional phosphatase) (Batchelor and Goulian, 2003; Shinar et al., 2007).

Even though this chapter suggests a synthetic system to be used in experiments, the idea is based on a natural signaling pathway which contains all the necessary "parts", except the fluorescence reporter output. The question is therefore, whether multiplexing is a suitable alternative hypothesis to the one suggested in Sec. 4.9, namely that different QS-dependent genes can weight the

signals differently. But even though inference might be possible from the various genes' outputs, demultiplexing, i.e. the subsequent decoding of the phosphorylation signal, is not achieved in this scenario, because no gene depends on phosphorylation alone. This means that for multiplexing to occur, there would have to be a demultiplexing unit – a further layer of interconnection between QS-dependent genes, downstream of directly QS-dependent gene expression. The approximated decoding (Eq. 5.11) also gives a hint to how this could be accomplished: Since eliminating the influence of f_{us} can be approximated with $[C]/[Y]^2$, it can be speculated that the demultiplexing unit could be an effector molecule which is positively impacted by the phosphorylation-dependent gene output and negatively regulated by the sequestering-dependent expression output. Further, a 2:1 stoichiometry of the two outputs (to accomplish $[Y]^2$) would be favored, for example, the sequestering-dependent output could be active as a dimer only. When researching the functions of QS-dependent genes in *B. subtilis* (Sec. 4.9), I found no evidence for such a downstream interconnection. Rather, it appeared that each gene has a specific function or array of functions. However, since (indirectly) ComA-dependent genes are numerous (Comella and Grossman, 2005), it cannot be excluded that such an interconnection might be discovered in the future.

Also, the suspected function of multiplexing as a tool for saving material can be doubted in the context of ComA, because the sequestering molecule Rap exceeds ComA in molecular mass (RapC: 45.47 kDa, ComA: 23.98 kDa, Mäder et al. (2011)).

In conclusion, I do not consider multiplexing a likely scenario in the natural *B. subtilis* QS pathway.

Chapter 6

General discussion

6.1   General findings

The pattern that crystallizes in this thesis is that the optimal promoter architecture depends on both the processing task (i.e. objective function) and – in case of RR TFs – the phospho-mode. In Ch. 3 I focused on effective signal transduction (*DOR*) in 2-BSs promoters and found that different phospho-modes require distinct patterns of BS affinities. Ch. 4 shows that for a specific phospho-mode, as found in the *B. subtilis* quorum sensing (QS) pathway, the weighting of two input signals – the fractions of un-sequestered and phosphorylated TF – can be tuned via the number of BSs and their affinities. In Ch. 5 the modeling results suggested that for a multiplexing task, a combination of two different promoters is suitable. Comparing the outcomes of Ch. 3 and Ch. 4 we see the difference that the investigated processing task can make: For a 2-BS promoter with phosphorylation-dependent cooperativity, high relative impact (*ri*) of phosphorylation favors asymmetric affinities, while in order to reach high absolute impact (*DOR*), the optimum isoline includes both asymmetric and equal affinities.

For the investigated tasks, effective signal transduction and high relative impact of the phosphorylation signal, »all« and »proximal« appear to be useful expression schemes. That is, strict requirements regarding which BSs have to be bound for expression lead to a stronger dependence of the output on the phosphorylation signal, for larger affinity regimes. Of course, for other purposes, for instance the design of logical gates (Buchler et al., 2003), an expression scheme that resembles »any«, might turn out useful.

In the case study on *B. subtilis* QS (Chs. 4 and 5) I explored the signal processing options for a particular pathway structure: In theory, for the same pathway architecture, varying the promoter architecture still allows for different ways of signal processing, including multiplexing and weighting of the signals.

6.2 Importance and possible application of the findings

The presented results may support the interpretation of experimentally determined pathway or promoter architectures in bacteria: Since Ch.3 gives the optimal promoter designs for different expression schemes and phospho-modes, this knowledge could be used to, vice versa, infer these features from given promoter sequence data. In Ch.4 I applied this strategy by comparing sequence data of QS-dependent promoters to make predictions on their functionality. I then used these predictions to underpin a hypothesis on the evolution of QS-dependent signal processing in *B. subtilis*. Further, Chs. 4 and 5 demonstrated that signal convergence at a TF, as observed in many pathway architectures, does neither necessarily mean that (1) all downstream genes respond to these signals in the same way nor that (2) large parts of the information has to be lost. Rather this thesis introduces promoter-based strategies to determine the weighting between the signals for a particular gene or to restore information on the input signals from two genes' outputs. These findings underscore the importance of promoter architecture and should be kept in mind when interpreting signaling pathway architectures.

The relations between architecture and functionality determined in this work could also inform design problems in synthetic biology: As mentioned in Sec. 1.1 TCSs are – due to both their plethora of functionalities and their modularity – frequently used as parts of engineered pathways (Ninfa, 2010). These, of course, should carry out a well defined processing task while potentially fulfilling further criteria to be determined by the experimentalist. To this end, Ch. 3 provides a comprehensive foundation for choosing a TCS with a suitable phospho-mode, and tuning the BS affinities, in order to optimize signal transduction effectiveness or response shape. Ch. 5 gives a specific example for how modeling can lead to design suggestions for a synthetic system with a well-defined processing task.

6.3 Outlook

As design questions depend strongly on the definition of the processing task at hand and the objective functions, this work can only cover a few of all possible options. In Ch. 3 I intentionally kept the objective functions basic, to allow for a comprehensive understanding of the differences between phospho-modes, while in Chs. 4 and 5 objective functions were designed according to the investigated processing task. The here envisaged objective functions were effectiveness of signal transduction (DOR), response shape (δ), relative impact (ri) and decoding ability (low decoding error E). Other frequently studied features of signal transduction systems are their robustness to changing inputs or parameters or their susceptibility to intrinsic noise (Elowitz et al., 2002), which will also be interesting to investigate regarding the models from this thesis. However, studying these features is not trivial, since their calculation requires (in the case of extrinsic influences) knowledge or assumptions on parameter distributions, which differ

highly between systems, and (in case of intrinsic noise) knowledge on kinetic parameters – as opposed to their ratios (e.g. K_A values), which were sufficient for the steady state descriptions used in this work. In the latter case, this means that the stochastic model setup has either twice the number of parameters, compared to the setup used here, or other methods, such as variations of the random telegraph model, that allow for simplified descriptions, have to be applied (Paulsson, 2005).

It should be noted that, even in the simple setup of the models here, implicit assumptions on the input signals were inevitable for calculating the applied objective functions. For the effectiveness of signal transduction and relative impact of the phosphorylation signal, I assumed uniform distributions of the input(s). Depending on the environmental conditions and given the presumably non-linear translation from environmental signal to phosphorylated RR through the HK, this is likely to be an oversimplification. Therefore, future perspectives include experiments that investigate such distributions in natural environments (which I motivate, for example, in Ch. 5) or computational studies which consider different input distributions and thus allow statements on the distributions that a system might be optimized for. An example for such a study on QS is the aforementioned publication by Mehta et al. (2009).

In Ch. 3 the four models so far only represent a few, intentionally simplified, generic scenarios, even though the modeling framework allows for more differentiated models. Other known features or theories could easily be implemented: For example, different types of BSs (IR, DR, ER) could, because of the distinct dimer orientation they enforce, cause distinct ability for dimer cooperativity, or overall affinity, which could be modeled by BS type-specific cooperativity or affinity factors. Further, with additional BSs, options for cooperative interaction show increasing complexity. Also hybrids of more than one phospho-mode are thinkable: For instance, both phosphorylation-dependent dimerization and dimer cooperativity have been stated for OmpR (Huang et al., 1997; Gao et al., 2008).

Results in this work are based on the assumption that TF binding to the promoter does not significantly change its concentration. The assumption was made explicitly in the derivation of analytic solutions, but also implicitly in the numeric simulations, because in these, only binding to one promoter was considered per simulation, thus neglecting possible other promoters or unspecific sites competing for the TF. Especially for TFs with large regulons and/or low copy numbers, the distribution of TFs along the genome will become of interest. This might be the case for *B. subtilis* QS, since the regulon constitutes roughly 20 BSs. Unfortunately, we have little knowledge on the intracellular concentration of ComA. Still, computational studies could at least give hints to the TF copy numbers or numbers of BSs at which these effects become relevant. Again, such approaches require different models from those used in this work. One obvious solution could be to combine the binding to whole sets of promoters into one ODE

model. This comes at the cost of bulky models, which I therefore suggest to implement in a rule-based fashion, i.e. from a preferably compact script, similar to the approach presented here (Sec. 3.3), in order to avoid chaos. Another option is the framework suggested by Bintu et al. (2005b), which is largely equivalent to the one used for analytical approximations in this work (Ahsendorf et al., 2014), but includes terms describing the distribution of TF molecules on the genome. The application of this framework will lead to larger, harder-to-handle equations than the ones presented here, and will require research on the expected number of (un-/specific) BSs as well as TF copy numbers.

Bibliography

Ahsendorf, T., Wong, F., Eils, R., and Gunawardena, J. (2014). A framework for modelling gene regulation which accommodates non-equilibrium mechanisms. *BMC Biology*, 12(1):102.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008). Molecular Biology of the Cell. *Garland Science, 5th edition*, pages 433–436.

Albright, L. M., Huala, E., and Ausubel, F. M. (1989). Prokaryotic signal transduction mediated by sensor and regulator protein pairs. *Annual Review of Genetics*, 23(1):311–336.

Amory, A., Kunst, F., Aubert, E., Klier, A., and Rapoport, G. (1987). Characterization of the *sacQ* genes from *Bacillus licheniformis* and *Bacillus subtilis*. *Journal of Bacteriology*, 169(1):324–333.

Ansaldi, M., Marolt, D., Stebe, T., Mandic-Mulec, I., and Dubnau, D. (2002). Specific activation of the *Bacillus* quorum-sensing systems by isoprenylated pheromone variants. *Molecular Microbiology*, 44(6):1561–1573.

Auchtung, J. M., Aleksanyan, N., Bulku, A., and Berkmen, M. B. (2016). Biology of ICE*Bs1*, an integrative and conjugative element in *Bacillus subtilis*. *Plasmid*, 86:14–25.

Auchtung, J. M., Lee, C. A., and Grossman, A. D. (2006). Modulation of the ComA-dependent quorum response in *Bacillus subtilis* by multiple Rap proteins and Phr peptides. *Journal of Bacteriology*, 188(14):5273–5285.

Auchtung, J. M., Lee, C. A., Monson, R. E., Lehman, A. P., and Grossman, A. D. (2005). Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proceedings of the National Academy of Sciences*, 102(35):12554–12559.

Ay, A. and Arnosti, D. N. (2011). Mathematical modeling of gene expression: a guide for the perplexed biologist. *Critical Reviews in Biochemistry and Molecular Biology*, 46(2):137–151.

Baikalov, I., Schröder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R. P., and Dickerson, R. E. (1996). Structure of the *Escherichia coli* response regulator NarL. *Biochemistry*, 35(34):11053–11061.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37(suppl_2):W202–W208.

Bais, H. P., Fall, R., and Vivanco, J. M. (2004). Biocontrol of *Bacillus subtilis* against infection of Arabidopsis roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiology*, 134(1):307–319.

Barbieri, C. M., Wu, T., and Stock, A. M. (2013). Comprehensive analysis of OmpR phosphorylation, dimerization, and DNA binding supports a canonical model for activation. *Journal of Molecular Biology*, 425(10):1612–1626.

Batchelor, E. and Goulian, M. (2003). Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *Proceedings of the National Academy of Sciences*, 100(2):691–696.

Beales, N. (2004). Adaptation of microorganisms to cold temperatures, weak acid preservatives, low pH, and osmotic stress: a review. *Comprehensive Reviews in Food Science and Food Safety*, 3(1):1–20.

Berg, G. (2009). Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. *Applied Microbiology and Biotechnology*, 84(1):11–18.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005a). Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Developmen*, 15(2):125–135.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005b). Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*, 15(2):116–124.

Bongiorni, C., Ishikawa, S., Stephenson, S., Ogasawara, N., and Perego, M. (2005). Synergistic regulation of competence development in *Bacillus subtilis* by two Rap-Phr systems. *Journal of Bacteriology*, 187(13):4353–4361.

Boucher, P. E., Murakami, K., Ishihama, A., and Stibitz, S. (1997). Nature of DNA binding and RNA polymerase interaction of the *Bordetella pertussis* BvgA transcriptional activator at the *fha* promoter. *Journal of Bacteriology*, 179(5):1755–1763.

Branda, S. S., González-Pastor, J. E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001). Fruiting body formation by *Bacillus subtilis*. *Proceedings of the National Academy of Sciences*, 98(20):11621–11626.

Browning, D. F. and Busby, S. J. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2(1):57.

Buchler, N. E., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences*, 100(9):5136–5141.

Bundschuh, R., Hayot, F., and Jayaprakash, C. (2003). The role of dimerization in noise reduction of simple genetic networks. *Journal of Theoretical Biology*, 220(2):261–269.

Burrus, V., Pavlovic, G., Decaris, B., and Guédon, G. (2002). The ICE*St1* element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid*, 48(2):77–97.

Carrillo, C., Teruel, J. A., Aranda, F. J., and Ortiz, A. (2003). Molecular mechanism of membrane permeabilization by the peptide antibiotic surfactin. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1611(1):91–97.

Chai, Y., Kolter, R., and Losick, R. (2009). A widely conserved gene cluster required for lactate utilization in *Bacillus subtilis* and its involvement in biofilm formation. *Journal of Bacteriology*, 191(8):2423–2430.

Collado-Vides, J., Magasanik, B., and Gralla, J. D. (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiological Reviews*, 55(3):371–394.

Comella, N. and Grossman, A. D. (2005). Conservation of genes and processes controlled by the quorum response in bacteria: characterization of genes controlled by the quorum-sensing transcription factor ComA in *Bacillus subtilis*. *Molecular Microbiology*, 57(4):1159–1174.

Core, L. and Perego, M. (2003). TPR-mediated interaction of RapC with ComA inhibits response regulator-DNA binding for competence development in *Bacillus subtilis*. *Molecular Microbiology*, 49(6):1509–1522.

Csonka, L. N. (1989). Physiological and genetic responses of bacteria to osmotic stress. *Microbiological Reviews*, 53(1):121–147.

Darwin, A. J., Tyson, K. L., Busby, S. J., and Stewart, V. (1997). Differential regulation by the homologous response regulators NarL and NarP of *Escherichia coli* K-12 depends on DNA binding site arrangement. *Molecular Microbiology*, 25(3):583–595.

De Jong, H. (2002). Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology*, 9(1):67–103.

de Ronde, W. and ten Wolde, P. R. (2014). Multiplexing oscillatory biochemical signals. *Physical Biology*, 11(2):026004.

De Wulf, P., McGuire, A. M., Liu, X., and Lin, E. C. (2002). Genome-wide profiling of promoter recognition by the two-component response regulator CpxR-P in *Escherichia coli*. *Journal of Biological Chemistry*, 277(29):26652–26661.

Dong, X.-R., Li, S.-F., and DeMoss, J. (1992). Upstream sequence elements required for NarL-mediated activation of transcription from the *narGHJI* promoter of *Escherichia coli*. *Journal of Biological Chemistry*, 267(20):14122–14128.

Driks, A. (2002). Overview: development in bacteria: spore formation in *Bacillus subtilis*. *CMLS Cellular and Molecular Life Sciences*, 59(3):389–391.

D'Souza, C., Nakano, M. M., and Zuber, P. (1994). Identification of *comS*, a gene of the *srfA* operon that regulates the establishment of genetic competence in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences*, 91(20):9397–9401.

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186.

Estrada, J., Wong, F., DePace, A., and Gunawardena, J. (2016). Information integration and energy expenditure in gene regulation. *Cell*, 166(1):234–244.

Even-Tov, E., Bendori, S. O., Valastyan, J., Ke, X., Pollak, S., Bareia, T., Ben-Zion, I., Bassler, B. L., and Eldar, A. (2016). Social evolution selects for redundancy in bacterial quorum sensing. *PLoS Biology*, 14(2):e1002386.

Foerster, J. and Pahle, J. (2018). The Copasi R connector, a high-level R API for Copasi. *http://jpahle.github.io/CoRC*.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3-4):601–620.

Fuqua, W. C., Winans, S. C., and Greenberg, E. P. (1994). Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *Journal of Bacteriology*, 176(2):269.

Gallego del Sol, F. and Marina, A. (2013). Structural basis of Rap phosphatase inhibition by Phr peptides. *PLoS Biology*, 11(3):e1001511.

Gao, R., Mack, T. R., and Stock, A. M. (2007). Bacterial response regulators: versatile regulatory strategies from common domains. *Trends in Biochemical Sciences*, 32(5):225–234.

Gao, R., Tao, Y., and Stock, A. M. (2008). System-level mapping of *Escherichia coli* response regulator dimerization with FRET hybrids. *Molecular Microbiology*, 69(6):1358–1372.

Giorgetti, L., Siggers, T., Tiana, G., Caprara, G., Notarbartolo, S., Corona, T., Pasparakis, M., Milani, P., Bulyk, M. L., and Natoli, G. (2010). Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Molecular Cell*, 37(3):418–428.

Gotoh, Y., Eguchi, Y., Watanabe, T., Okamoto, S., Doi, A., and Utsumi, R. (2010). Two-component signal transduction as potential drug targets in pathogenic bacteria. *Current Opinion in Microbiology*, 13(2):232–239.

Griffith, K. L. and Grossman, A. D. (2008). A degenerate tripartite DNA-binding site required for activation of ComA-dependent quorum response gene expression in *Bacillus subtilis*. *Journal of Molecular Biology*, 381(2):261–275.

Guet, C. C., Elowitz, M. B., Hsing, W., and Leibler, S. (2002). Combinatorial synthesis of genetic networks. *Science*, 296(5572):1466–1470.

Gunawardena, J. (2005). Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proceedings of the National Academy of Sciences*, 102(41):14617–14622.

Hamoen, L. W., Eshuis, H., Jongbloed, J., Venema, G., and Sinderen, D. (1995). A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. *Molecular Microbiology*, 15(1):55–63.

Hansen, A. S. and O'Shea, E. K. (2015). Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *Elife*, 4:e06559.

Hasty, J., McMillen, D., Isaacs, F., and Collins, J. J. (2001). Computational studies of gene regulatory networks: *in numero* molecular biology. *Nature Reviews Genetics*, 2(4):268.

Hayashi, K., Kensuke, T., Kobayashi, K., Ogasawara, N., and Ogura, M. (2006). *Bacillus subtilis* RghR (YvaN) represses *rapG* and *rapH*, which encode inhibitors of expression of the *srfA* operon. *Molecular Microbiology*, 59(6):1714–1729.

He, H. and Zahrt, T. C. (2005). Identification and characterization of a regulatory sequence recognized by *Mycobacterium tuberculosis* persistence regulator MprA. *Journal of Bacteriology*, 187(1):202–212.

He, X., Samee, M. A. H., Blatti, C., and Sinha, S. (2010). Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Computational Biology*, 6(9):e1000935.

Head, C. G., Tardy, A., and Kenney, L. J. (1998). Relative binding affinities of OmpR and OmpR-phosphate at the *ompF* and *ompC* regulatory sites. *Journal of Molecular Biology*, 281(5):857–870.

Hlavacek, W. S., Faeder, J. R., Blinov, M. L., Posner, R. G., Hucka, M., and Fontana, W. (2006). Rules for modeling signal-transduction systems. *Science's STKE*, 2006(344):re6.

Hobbs, C. A., Bobay, B. G., Thompson, R. J., Perego, M., and Cavanagh, J. (2010). NMR solution structure and DNA-binding model of the DNA-binding domain of competence protein A. *Journal of Molecular Biology*, 398(2):248–263.

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

Huang, K.-J., Lan, C.-Y., and Igo, M. M. (1997). Phosphorylation stimulates the cooperative DNA-binding properties of the transcription factor OmpR. *Proceedings of the National Academy of Sciences*, 94(7):2828–2832.

Igoshin, O. A., Alves, R., and Savageau, M. A. (2008). Hysteretic and graded responses in bacterial two-component signal transduction. *Molecular Microbiology*, 68(5):1196–1215.

Jarmer, H., Larsen, T. S., Krogh, A., Saxild, H. H., Brunak, S., and Knudsen, S. (2001). Sigma A recognition sites in the *Bacillus subtilis* genome. *Microbiology*, 147(9):2417–2424.

Jiang, M., Grau, R., and Perego, M. (2000). Differential processing of propeptide inhibitors of Rap phosphatases in *Bacillus subtilis*. *Journal of Bacteriology*, 182(2):303–310.

Johnson, C. M. and Grossman, A. D. (2015). Integrative and conjugative elements (ICEs): what they do and how they work. *Annual Review of Genetics*, 49:577–601.

Katsir, G., Jarvis, M., Phillips, M., Ma, Z., and Gunsalus, R. P. (2015). The *Escherichia coli* NarL receiver domain regulates transcription through promoter specific functions. *BMC Microbiology*, 15(1):174.

Kauffman, S., Peterson, C., Samuelsson, B., and Troein, C. (2003). Random Boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences*, 100(25):14796–14799.

Kearns, D. B., Chu, F., Rudner, R., and Losick, R. (2004). Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Molecular Microbiology*, 52(2):357–369.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A., et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, 390(6657):249.

Kunst, F., Pascal, M., Lepesant-Kejzlarova, J., Lepesant, J.-A., Billault, A., and Dedonder, R. (1975). Pleiotropic mutations affecting sporulation conditions and the syntheses of extracellular enzymes in *Bacillus subtilis* 168. *Biochimie*, 56(11):1481–1489.

Lazazzera, B. A., Kurtser, I. G., McQuade, R. S., and Grossman, A. D. (1999). An autoregulatory circuit affecting peptide signaling in *Bacillus subtilis*. *Journal of Bacteriology*, 181(17):5193–5200.

Leisner, M., Stingl, K., Frey, E., and Maier, B. (2008). Stochastic switching to competence. *Current Opinion in Microbiology*, 11(6):553–559.

Lejona, S., Castelli, M. E., Cabeza, M. L., Kenney, L. J., Véscovi, E. G., and Soncini, F. C. (2004). PhoP can activate its target genes in a PhoQ-independent manner. *Journal of Bacteriology*, 186(8):2476–2480.

Lengyel, I. M., Soroldoni, D., Oates, A. C., and Morelli, L. G. (2014). Nonlinearity arising from noncooperative transcription factor binding enhances negative feedback and promotes genetic oscillations. *Papers in Physics*, 6:060012.

Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., et al. (2005). Engineering *Escherichia coli* to see light. *Nature*, 438(7067):441.

Li, J. and Stewart, V. (1992). Localization of upstream sequence elements required for nitrate and anaerobic induction of *fdn* (formate dehydrogenase-N) operon expression in *Escherichia coli* K-12. *Journal of Bacteriology*, 174(15):4935–4942.

Liu, W. and Hulett, F. M. (1997). *Bacillus subtilis* PhoP binds to the *phoB* tandem promoter exclusively within the phosphate starvation-inducible promoter. *Journal of Bacteriology*, 179(20):6302–6310.

Liu, W., Qi, Y., and Hulett, F. M. (1998). Sites internal to the coding regions of *phoA* and *pstS* bind PhoP and are required for full promoter activity. *Molecular Microbiology*, 28(1):119–130.

Long, T., Tu, K. C., Wang, Y., Mehta, P., Ong, N., Bassler, B. L., and Wingreen, N. S. (2009). Quantifying the integration of quorum-sensing signals with single-cell resolution. *PLoS Biology*, 7(3):e1000068.

Mackey, M. C., Santillán, M., and Yildirim, N. (2004). Modeling operon dynamics: the tryptophan and lactose operons as paradigms. *Comptes Rendus Biologies*, 327(3):211–224.

Mäder, U., Schmeisky, A. G., Flórez, L. A., and Stülke, J. (2011). Subti Wiki—a comprehensive community resource for the model organism *Bacillus subtilis*. *Nucleic Acids Research*, 40(D1):D1278–D1287.

Magnuson, R., Solomon, J., and Grossman, A. D. (1994). Biochemical and genetic characterization of a competence pheromone from *B. subtilis*. *Cell*, 77(2):207–216.

Marianayagam, N. J., Sunde, M., and Matthews, J. M. (2004). The power of two: protein dimerization in biology. *Trends in Biochemical Sciences*, 29(11):618–625.

McAdams, H. H., Srinivasan, B., and Arkin, A. P. (2004). The evolution of genetic regulatory systems in bacteria. *Nature Reviews Genetics*, 5:169 – 178.

Mehta, P., Goyal, S., Long, T., Bassler, B. L., and Wingreen, N. S. (2009). Information processing and signal integration in bacterial quorum sensing. *Molecular Systems Biology*, 5(1).

Miller, M. B. and Bassler, B. L. (2001). Quorum sensing in bacteria. *Annual Reviews in Microbiology*, 55(1):165–199.

Mirouze, N., Parashar, V., Baker, M. D., Dubnau, D. A., and Neiditch, M. B. (2011). An atypical Phr peptide regulates the developmental switch protein RapH. *Journal of Bacteriology*, 193(22):6197–6206.

Mizuno, T. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Research*, 4(2):161–168.

Mueller, J., Bukusoglu, G., and Sonenshein, A. (1992). Transcriptional regulation of *Bacillus subtilis* glucose starvation-inducible genes: control of *gsiA* by the ComP-ComA signal transduction system. *Journal of Bacteriology*, 174(13):4361–4373.

Nakano, M., Magnuson, R., Myers, A., Curry, J., Grossman, A., and Zuber, P. (1991a). *srfA* is an operon required for surfactin production, competence development, and efficient sporulation in *Bacillus subtilis*. *Journal of bacteriology*, 173(5):1770–1778.

Nakano, M. and Zuber, P. (1993). Mutational analysis of the regulatory region of the *srfA* operon in *Bacillus subtilis*. *Journal of Bacteriology*, 175(10):3188–3191.

Nakano, M. M., Marahiel, M., and Zuber, P. (1988). Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. *Journal of Bacteriology*, 170(12):5662–5668.

Nakano, M. M., Xia, L., and Zuber, P. (1991b). Transcription initiation region of the *srfA* operon, which is controlled by the comP-comA signal transduction system in *Bacillus subtilis*. *Journal of Bacteriology*, 173(17):5487–5493.

Nasser, W., Awade, A., Reverchon, S., and Robert-Baudouy, J. (1993). Pectate lyase from *Bacillus subtilis*: molecular characterization of the gene, and properties of the cloned enzyme. *FEBS Letters*, 335(3):319–326.

Ng, W.-L. and Bassler, B. L. (2009). Bacterial quorum-sensing network architectures. *Annual Review of Genetics*, 43:197–222.

Ninfa, A. J. (2010). Use of two-component signal transduction systems in the construction of synthetic genetic networks. *Current Opinion in Microbiology*, 13(2):240–245.

Ogura, M. and Fujita, Y. (2007). *Bacillus subtilis rapD*, a direct target of transcription repression by RghR, negatively regulates *srfA* expression. *FEMS Microbiology Letters*, 268(1):73–80.

Ohm, J.-R. and Lüke, H. D. (2010). *Signalübertragung: Grundlagen der digitalen und analogen Nachrichtenübertragungssysteme*. Springer-Verlag.

Oslizlo, A., Stefanic, P., Dogsa, I., and Mandic-Mulec, I. (2014). Private link between signal and response in *Bacillus subtilis* quorum sensing. *Proceedings of the National Academy of Sciences*, 111(4):1586–1591.

Paulsson, J. (2005). Models of stochastic gene expression. *Physics of Life Reviews*, 2(2):157–175.

Perego, M. (1997). A peptide export–import control circuit modulating bacterial development regulates protein phosphatases of the phosphorelay. *Proceedings of the National Academy of Sciences*, 94(16):8612–8617.

Perego, M. and Brannigan, J. A. (2001). Pentapeptide regulation of aspartyl-phosphate phosphatases. *Peptides*, 22(10):1541–1547.

Perego, M., Hanstein, C., Welsh, K. M., Djavakhishvili, T., Glaser, P., and Hoch, J. A. (1994). Multiple protein-aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in *B. subtilis*. *Cell*, 79(6):1047–1055.

Perego, M., Higgins, C., Pearce, S., Gallagher, M., and Hoch, J. (1991). The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. *Molecular Microbiology*, 5(1):173–185.

Perron-Savard, P., De Crescenzo, G., and Le Moual, H. (2005). Dimerization and DNA binding of the *Salmonella enterica* PhoP response regulator are phosphorylation independent. *Microbiology*, 151(12):3979–3987.

Piazza, F., Tortosa, P., and Dubnau, D. (1999). Mutational analysis and membrane topology of ComP, a quorum-sensing histidine kinase of *Bacillus subtilis* controlling competence development. *Journal of Bacteriology*, 181(15):4540–4548.

Pottathil, M., Jung, A., and Lazazzera, B. A. (2008). CSF, a species-specific extracellular signaling peptide for communication among strains of *Bacillus subtilis* and *Bacillus mojavensis*. *Journal of Bacteriology*, 190(11):4095–4099.

Pottathil, M. and Lazazzera, B. A. (2003). The extracellular Phr peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. *Frontiers in Bioscience*, 8:d32–d45.

Qi, Y. and Hulett, F. M. (1998). PhoP~P and RNA polymerase σ^A holoenzyme are sufficient for transcription of Pho regulon promoters in *Bacillus subtilis*: PhoP~P activator sites within the coding region stimulate transcription *in vitro*. *Molecular Microbiology*, 28(6):1187–1197.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rampersaud, A., Harlocker, S. L., and Inouye, M. (1994). The OmpR protein of *Escherichia coli* binds to sites in the *ompF* promoter region in a hierarchical manner determined by its degree of phosphorylation. *Journal of Biological Chemistry*, 269(17):12559–12566.

Roggiani, M. and Dubnau, D. (1993). ComA, a phosphorylated response regulator protein of *Bacillus subtilis*, binds to the promoter region of *srfA*. *Journal of Bacteriology*, 175(10):3182–3187.

Rudner, D. Z., LeDeaux, J. R., Ireton, K., and Grossman, A. D. (1991). The spo0K locus of *Bacillus subtilis* is homologous to the oligopeptide permease locus and is required for sporulation and competence. *Journal of Bacteriology*, 173(4):1388–1398.

Schujman, G. E., Paoletti, L., Grossman, A. D., and de Mendoza, D. (2003). FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Developmental Cell*, 4(5):663–672.

Setty, Y., Mayo, A. E., Surette, M. G., and Alon, U. (2003). Detailed map of a cis-regulatory input function. *Proceedings of the National Academy of Sciences*, 100(13):7702–7707.

Shea, M. A. and Ackers, G. K. (1985). The OR control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of Molecular Biology*, 181(2):211–230.

Shinar, G., Milo, R., Martínez, M. R., and Alon, U. (2007). Input–output robustness in simple bacterial signaling systems. *Proceedings of the National Academy of Sciences*, 104(50):19931–19935.

Shmulevich, I., Dougherty, E. R., and Zhang, W. (2002). From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. *Proceedings of the IEEE*, 90(11):1778–1792.

Singh, P. and Cameotra, S. S. (2004). Potential applications of microbial surfactants in biomedical sciences. *Trends in Biotechnology*, 22(3):142–146.

Sinha, A., Gupta, S., Bhutani, S., Pathak, A., and Sarkar, D. (2008). PhoP-PhoP interaction at adjacent PhoP binding sites is influenced by protein phosphorylation. *Journal of Bacteriology*, 190(4):1317–1328.

Smits, W. K., Bongiorni, C., Veening, J.-W., Hamoen, L. W., Kuipers, O. P., and Perego, M. (2007). Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in *Bacillus subtilis*. *Molecular Microbiology*, 65(1):103–120.

Solomon, J. M., Lazazzera, B. A., and Grossman, A. D. (1996). Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes & Development*, 10(16):2014–2024.

Stein, T. (2005). *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. *Molecular Microbiology*, 56(4):845–857.

Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. *Annual Review of Biochemistry*, 69(1):183–215.

Stormo, G. D. and Fields, D. S. (1998). Specificity, free energy and information content in protein–DNA interactions. *Trends in Biochemical Sciences*, 23(3):109–113.

Tortosa, P., Logsdon, L., Kraigher, B., Itoh, Y., Mandic-Mulec, I., and Dubnau, D. (2001). Specificity and genetic polymorphism of the *Bacillus* competence quorum-sensing system. *Journal of Bacteriology*, 183(2):451–460.

Tran, L.-S. P., Nagai, T., and Itoh, Y. (2000). Divergent structure of the ComQXPA quorum-sensing components: molecular basis of strain-specific communication mechanism in *Bacillus subtilis*. *Molecular Microbiology*, 37(5):1159–1171.

Tsuge, K., Ano, T., Hirai, M., Nakamura, Y., and Shoda, M. (1999). The genes *degQ*, *pps*, and *lpa-8* (*sfp*) are responsible for conversion of *Bacillus subtilis* 168 to plipastatin production. *Antimicrobial Agents and Chemotherapy*, 43(9):2183–2192.

Tsuge, K., Inoue, S., Ano, T., Itaya, M., and Shoda, M. (2005). Horizontal transfer of Iturin A operon, *itu*, to *Bacillus subtilis* 168 and conversion into an Iturin A producer. *Antimicrobial Agents and Chemotherapy*, 49(11):4641–4648.

Tyson, K., Bell, A., Cole, J., and Busby, S. (1993). Definition of nitrite and nitrate response elements at the anaerobically inducible *Escherichia coli nirB* promoter: interactions between FNR and NarL. *Molecular Microbiology*, 7(1):151–157.

Tyson, K., Cole, J., and Busby, S. (1994). Nitrite and nitrate regulation at the promoters of two *Escherichia coli* operons encoding nitrite reductase: identification of common target heptamers for both NarP- and NarL-dependent regulation. *Molecular Microbiology*, 13(6):1045–1055.

Veening, J.-W., Hamoen, L. W., and Kuipers, O. P. (2005). Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. *Molecular Microbiology*, 56(6):1481–1494.

Von Hippel, P. H., Revzin, A., Gross, C. A., and Wang, A. C. (1974). Non-specific DNA binding of genome regulating proteins as a biological control mechanism: 1. The *lac* operon: equilibrium aspects. *Proceedings of the National Academy of Sciences*, 71(12):4808–4812.

Walker, M. and DeMoss, J. (1994). NarL-phosphate must bind to multiple upstream sites to activate transcription the *narG* promoter of *Escherichia coli*. *Molecular Microbiology*, 14(4):633–641.

Waters, C. M. and Bassler, B. L. (2005). Quorum sensing: cell-to-cell communication in bacteria. *Annual Review of Cell and Developmental Biology*, 21:319–346.

Waters, C. M. and Bassler, B. L. (2006). The *Vibrio harveyi* quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers. *Genes & Development*, 20(19):2754–2767.

Wegscheider, R. (1901). Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reactionskinetik homogener Systeme. *Monatshefte für Chemie und verwandte Teile anderer Wissenschaften*, 22:849–906.

Weinrauch, Y., Guillen, N., and Dubnau, D. (1989). Sequence and transcription mapping of *Bacillus subtilis* competence genes *comB* and *comA*, one of which is related to a family of bacterial regulatory determinants. *Journal of Bacteriology*, 171(10):5362–5375.

Weinrauch, Y., Penchev, R., Dubnau, E., Smith, I., and Dubnau, D. (1990). A *Bacillus subtilis* regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-transduction systems. *Genes & Development*, 4(5):860–872.

Whitaker, W. R., Davis, S. A., Arkin, A. P., and Dueber, J. E. (2012). Engineering robust control of two-component system phosphotransfer using modular scaffolds. *Proceedings of the National Academy of Sciences*, 109(44):18090–18095.

Wolf, D., Rippa, V., Mobarec, J. C., Sauer, P., Adlung, L., Kolb, P., and Bischofs, I. B. (2016). The quorum-sensing regulator ComA from *Bacillus subtilis* activates transcription using topologically distinct DNA motifs. *Nucleic Acids Research*, 44(5):2160—2172.

Wolfram Research, I. (2018). Mathematica. *Wolfram Research, Inc.*, Version 10.0.

Wray, G. A. (2007). The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics*, 8(3):206.

Yang, M., Ferrari, E., Chen, E., and Henner, D. (1986). Identification of the pleiotropic *sacQ* gene of *Bacillus subtilis*. *Journal of Bacteriology*, 166(1):113–119.

Zuber, P. (2001). A peptide profile of the *Bacillus subtilis* genome. *Peptides*, 22(10):1555–1577.

Danksagung

Acknowledgement

List of Figures

List of Tables