UPTEC X 05 010
FEB 2005

ISSN 1401-2138

HANNA HÄRDIN

Modeling and simulation of stimulation of motility in *Synechocystis* PCC 6803

Master's degree project

UPPSALA UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 05 010	**Date of issue 2005-02**

Author

Hanna Härdin

Title (English)

Modeling and simulation of stimulation of motility in *Synechocystis* PCC 6803

Abstract

When the cyanobacterium *Synechocystis* PCC 6803 is exposed to blue light, it responds by moving away from the light source. There are also indications that, when exposed to glucose, the bacterium responds by decreasing its motility. These processes are controlled by a large and complex network of genes, a so-called genetic regulatory network. Because of the lack of information about the components of the network underlying these processes in *Synechocystis* PCC 6803, several alternative networks can be imagined. These possible networks have been modeled and simulated. Since almost no quantitative data on the system is available, a qualitative simulation method that can handle these constraints has been used. By analyzing the simulation results, it has been possible to select experiments that would discriminate effectively between possible networks.

Keywords

Qualitative modeling and simulation, genetic regulatory networks, chemotaxis, phototaxis, motility, *Synechocystis* PCC 6803, piecewise-linear differential equations

Supervisors

Hidde de Jong
Helix bionformatics, INRIA Rhône-Alpes, Grenoble

Scientific reviewer

Måns Ehrenberg
Department of Cell & Molecular Biology, Uppsala University

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **49**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Modellering och simulering av stimulans av motilitet i *Synechocystis* PCC 6803

Hanna Härdin

Sammanfattning

Vissa bakterier kan röra sig mot, alternativt ifrån, en ljuskälla. Det finns också bakterier vars rörelseförmåga påverkas av vissa typer av kemikalier. I bakterier med sådana egenskaper finns molekyler som har förmåga att registrera ljus eller kemikalier. Dessa molekyler för vidare informationen inuti bakterien genom att reagera med andra molekyler. Signalerna resulterar i att bakterien bygger upp, eller bryter ner, det molekylära maskineri som får bakterien att röra sig. De signalerande molekylerna som för vidare informationen genom bakterien utgör ett stort och komplext så kallat genetiskt regulatoriskt nätverk.

Bakterien *Synechocystis* PCC 6803 rör sig bort från blått ljus. Man har också visat att sockermolekylen glukos tycks minska bakteriens rörlighet. Eftersom man inte har fullständig information om det genetiska regulatoriska nätverk som ligger bakom dessa beteenden finns flera tänkbara alternativa nätverk. I detta arbete har dessa tänkbara nätverk modellerats och simulerats. Genom att analysera simuleringsresultat har det varit möjligt att föreslå experiment vars resultat kan hjälpa till att urskilja bland de möjliga nätverken. Urskiljningen bland de möjliga nätverken kan göras genom att de nätverk vars simuleringsresultat ej överensstämmer med resultat från de föreslagna experimenten förkastas. Genom att med hjälp av modellering och simulering föreslå informativa experiment effektiviseras det experimentella arbetet.

Examensarbete 20 p i Molekylär bioteknikprogrammet

Uppsala universitet, februari 2005

Contents

1 Introduction

When exposed to light, some species of bacteria respond by moving towards or away from it. Similarly, when exposed to a concentration gradient of some certain molecule, the bacteria might move along or against the direction of the gradient. These processes, known as phototaxis and chemotaxis respectively, have evolved in order to lead the bacteria to places with optimal light intensity, towards nutrients, or away from toxic molecules.

In phototactic and chemotactic organisms a light or chemical signal causes induced expression of proteins responsible for building up the motility apparatus, *e.g.* pilin, which is one of the constituents of the tail-like motor pili, that some bacteria use for their motility. Between the light or chemical input and the final motility output, molecules arranged in a regulatory network are passing on the signal through the cell. These molecular networks generally involve a large amount of components and a complex system of interactions between these, constituting a variety of feedback loops that regulate the system. Due to the complex regulatory structure often including numerous feedback loops, it is difficult to understand what influence a change of the concentration of one molecule would have on the other components, or how a change in environmental conditions is propagated in the system.

To get an understanding of how the global behaviour of this kind of networks emerges from local interactions, mathematical modeling and simulation are essential tools. By collecting the knowledge about the system in a mathematical model and simulate it, it is possible to learn more about how changes of certain parameters, such as concentrations or environmental signals, are propagated through the system, and what consequences these changes have on the overall behaviour of the network. If the information about the system is incomplete, different hypothetical systems can be simulated. Comparing the simulation results with the outcome from experiments might make it possible to exclude some of the hypothetical models. In this way, modeling and simulation might help to suggest those experiments that would discriminate most efficiently between the hypotheses about the system. By concentrating on these proposed experiments, the experimental work may become more efficient.

Because of the complex structures of chemotaxis and phototaxis networks, it is of interest to use modeling and simulation to study these processes. In order to do this, the model cyanobacterium *Synechocystis* PCC 6803 would be a suitable organism to study. This bacterium, whose entire genome sequence is determined [17], is known to be phototactic and there are vague indications that it also might exhibit chemotaxis as a response to glucose.[1] Its motility is accomplished by pili [1]. Unlike the flagella generating process, which has been well-studied in many organisms, the pili generating process has not been studied in much detail. Therefore, in order to learn more about the pili-generating process, it would be interesting to apply modeling and simulation methods to this process in *Synechocystis* PCC 6803.

As a consequence of the fact that not much research has been done on the motility pathway in *Synechocystis* PCC 6803, not much is known about the components, and the interactions between the components, that are involved in phototaxis and chemotaxis. One thing that is known though, is that cyclic AMP (cAMP) is involved as a key regulator of phototaxis [26]. When exposed to blue light, the cAMP level in *Synechocystis* PCC 6803 increases, as well as the production of pili [27]. Cyanobacteria have the peculiarity among prokaryotes of containing both cAMP and cyclic GMP (cGMP), and there are indications that also cGMP

[1] Unpublished result from experiment performed by Jean Houmard, Ecole Normale Supérieur (ENS), Paris.

is involved in motility in *Synechocystis* PCC 6803. Both cAMP and cGMP are molecules involved in many signaling networks with various functions in different organisms. It is plausible that cAMP and cGMP are involved in several additional signaling pathways in *Synechocystis* PCC 6803, which makes the situation even more complex.

Modeling and simulation of the motility pathway in *Synechocystis* PCC 6803 is particularly difficult because, in addition to the limited amount of qualitative information about which components are involved and how they interact, almost no quantitative data, such as component concentrations, binding constants, and other kinetic constants, is available. Therefore, to simulate this network, most methods are impossible to apply since they usually require precise quantitative data. For this reason, a method based on piecewise-linear differential equations (PLDEs), which does not require quantitative data, has been used [11]. For this method it is sufficient to know certain parameter inequalities that can be determined from qualitative information about the biological system. The method has previously been applied to systems for which only qualitative information has been available [9, 24, 28]. However, in the case of the motility signal system in *Synechocystis* PCC 6803, there is less qualitative information available than there was for any of the systems that have been modeled and simulated with this method so far. Therefore, it is of additional interest to test whether the method is of any assistance when applied to the motility system in *Synechocystis* PCC 6803, where in addition to the almost complete absence of quantitative data, not very much qualitative information is available.

In this work, a small subsystem of the phototaxis and the putative chemotaxis network in *Synechocystis* PCC 6803, centered around the cyclic nucleotides cAMP and cGMP, has been studied. The aim of this study is twofold. Firstly, it is to use the qualitative modeling and simulation method previously described to understand how the overall behaviour arises from molecular interactions in this subsystem of the motility network in *Synechocystis* PCC 6803. Due to the scarce information about this system there are many possible hypotheses about the network structure and hence many different hypothetical models that describe the system. So the second aim is to use modeling and simulation to help propose experiments whose results compared to the simulation results allow confirming or rejecting hypotheses about the system.

The information that has been used to build possible networks comes from (1) the literature, (2) Jean Houmard, biologist and cyanobacteria specialist at Ecole Normale Supérieure (ENS) in Paris, and (3) the CyanoBase Internet site [6]. Due to the fact that not much research is done on the system under study, this information is incomplete. Therefore, four classes of plausible networks underlying the stimulation of motility in *Synechocystis* PCC 6803 could be derived from the available information after doing some underlying assumptions about the interactions. For all these networks, piecewise-linear differential equation models (PLDE models) were created. The models in class I contain 4 PLDEs each, one PLDE representing the light input and the other three describing the evolution of concentrations of 3 gene products in the system under study. The models in class II, III, and IV contain glucose as an additional input variable and the concentration of 2 more genes as compared to the models in class I, and hence these models contain 7 PLDEs. The PLDE models were investigated by Genetic Network Analyzer (GNA), software implementing the above-mentioned simulation method [10]. For each model, GNA provides all possible time evolutions for the qualitative concentration levels of the different components in the model.

Analysis of the results of simulations of the motility network by means of the different models has shown that it is possible to distinguish between some of the models by performing

certain experiments. The light influence on motility can be elucidated by studying the motility in response to light in two different mutants. That is, it can be determined which one of the models in class I is true. A second experiment can be performed in order to learn more about the glucose influence on the motility. More specifically, with this second experiment it would be possible to determine whether glucose affects motility directly via cGMP or whether cGMP acts via cAMP on the motility.

In the next section, the biological system under study will be presented, that is, more details about the motility inducing process in *Synechocystis* PCC 6803 and different hypotheses about the system will be given. In section 3, the qualitative modeling and simulation method that has been used will be explained. In sections 4 and 5, the derivation of the qualitative mathematical models and the simulation results will be given. The report ends with a discussion of the results and future work.

2 Regulation of motility by cyclic nucleotides in *Synechocystis* PCC 6803

When life arose on our planet, the first immotile organisms could only utilize the energy source that was available at the spot where they happened to live, and when exposed to toxics they were helpless. It is easy to imagine that motility would make a great improvement of the ability of many organisms to survive. Motility might, for example, help organisms to escape from toxics or move towards spots with a high concentration of nutrients. The 3.5 billion years old cyanobacteria might have been the first group of organisms that developed motility during the evolution.

Different solutions for the performance of motility in prokaryotes have evolved. Most common is the use of flagella, tail-like appendages on the outside of bacteria, connected to the cell surface via a huge protein complex, responsible for swimming motility. This advanced motor mechanism has been studied in much detail [3]. The cyanobacterium *Synechocystis* PCC 6803 exhibits a sort of sliding motility, called twisting motility, and for this it does not use flagella, but instead pili [1]. Pili are also a kind of tail-like cell surface structures, but shorter than flagella (see figure 1). In contrast to the flagella machinery, the pili apparatus has not been studied in any detail. Pili are known for their ability to stick the cell to surfaces or to other cells, it was only recently shown that some pili are used for motility. On the surface of *Synechocystis* PCC 6803 there are two types of pili, thin and thick, and it is the thick type, which also is known as type IV pili, that is responsible for the motility [1].

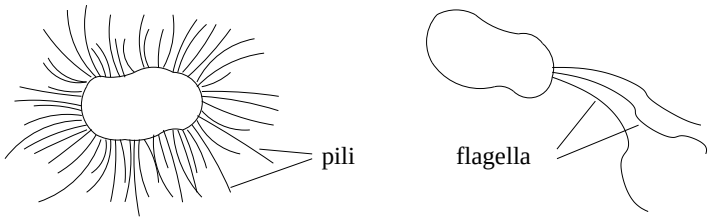

Figure 1: Bacterial cells with pili and flagella attached to the cell surfaces. Flagella are in general longer and less numerous than pili. Pili are distributed around the whole cell surface, while flagella may be either scattered around the entire cell surface (like pili) or concentrated at one or both ends of the cell (as in the picture) [3].

Synechocystis PCC 6803 is both positively and negatively phototactic, *i.e.* it can respond to light inputs by moving towards or away from it, respectively, depending on the type of light it is exposed to [20]. The positive chemotaxis is induced by low-intensity light, with peak spectral sensitivity in the red light spectrum. When exposed to high-intensity light, the bacterium is negatively phototactic, and this response is caused mainly by light in the blue light spectrum. It is also known that this latter response acts via changing the cAMP level [27]. This latter response is the one we refer to when we henceforth speak about the response to light, *i.e.* we will only regard the response to blue light and not to red light in this report. Since it is also known that the type IV pili are responsible for cell motility [1], it is plausible that one way to generate pili is via the cAMP signaling pathway as a response to high-intensity light.

There are indications that motility of *Synechocystis* PCC 6803 might be inhibited by glucose, *i.e.* the bacterium might exhibit chemotaxis in addition to phototaxis. What is

known is that *Synechocystis* PCC 6803 tends to aggregate in the presence of glucose (according to experiments performed by Jean Houmard, ENS, Paris). Also, in the presence of both light and glucose, the concentration of the cyclic nucleotide cGMP is increased to the double [16]. This suggests that glucose is inhibiting motility via the cGMP pathway, but as seen, the present evidence for this is inconclusive.

By computer analysis and experiments, many proteins have been proved or suggested to bind the cyclic nucleotides, indicating that the motility signaling may involve many complex interactions [4]. It is also shown by mutation experiments that a large number of different genes are necessary for motility [2], a fact that also suggests that the motility network is not trivial.

2.1 Division of the response into modules

As a first attempt to organize the sparse information about the motility pathway in *Synechocystis* PCC 6803, one might divide the process roughly into three main modules, which we will call the signal transduction module, the regulation module and the motility response module, respectively (see figure 2). We define the signal transduction module as the network including the molecules that receive the light or glucose signal on the cell surface and forwards this signal to the inside, and then onwards to the regulation module. The regulation module is centered around the cyclic nucleotides cAMP and cGMP and includes a well-defined set of molecules as described below. We guess that this module, as our name of it indicates, is responsible for the regulation of the entire motility process by integrating all signals that influence the cyclic nucleotides and taking the decision whether to increase or decrease the pili production. The motility response module contains all molecules regulated by the regulation module, including the ones that in the end perform the pili production by expression of proteins that are constituents of the motility machinery.

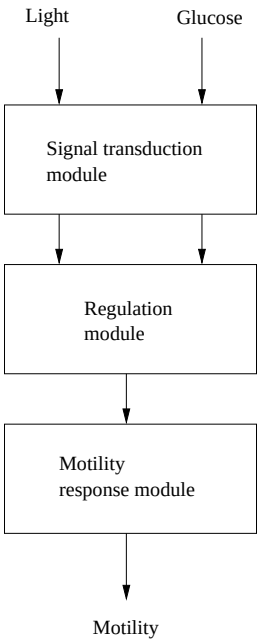

Figure 2: Division of the motility network into three modules.

Gene name used here	Name in CyanoBase
cya1	*slr1991*
sycrp1	*sll1371*
pdeB	*sll1624*
cya2	*sll0646*
sycrp2	*sll1924*
pdeA	*slr2100*

Table 1: Names of genes in the motility network used in CyanoBase [6].

The modeling done in this work is covering the middle module, *i.e.* the regulation module, and the other two modules are here regarded as input and output respectively. That is, the signal transduction module is collapsed into input signals representing light intensity and glucose concentration, and in a similar way, the motility response module is collapsed into an output signal representing the pili production rate.

2.2 Components included in the regulation module

Generally, in various types of organisms, cAMP and cGMP are produced out of ATP and GTP by adenylyl cyclases and guanylyl cyclases (together called nucleotide cyclases). The enzymatic hydrolysis of cAMP and cGMP to AMP and GMP is performed by phosphodiesterases. Further, cAMP and cGMP can bind to receptor proteins that, when in complex with the cyclic nucleotides, can directly regulate the expression of target genes. See figure 3(a) for an overview, in which the notation formalism suggested by Kohn [18] (see figure 4) is used to present the information. The regulation module considered in this work consists of the cyclic nucleotides and proteins expected to act as cyclases, phosphodiesterases, and cAMP- and cGMP-receptor proteins. In the remainder of this subsection, we will review information about all components included in the regulation module, first the ones centered around cAMP, then cGMP. All of the information in this and the next subsections refers to components in *Synechocystis* PCC 6803. The gene names used in the database CyanoBase [6] for the genes in the motility network are showed in table 1.

The protein Cya1, encoded by the gene *cya1*, shows high sequence similarity to adenylyl cyclases in other bacteria [26]. It has been shown that *cya1* mutants are immotile and contain almost no cAMP [26]. Also it has been shown that when exposed to blue light, the cAMP level in wild type cells is increased, while the cAMP level in a *cya1* mutant does not change [27]. All these facts constitute strong evidence for Cya1 being an adenylyl cyclase activated by light and involved in the motility pathway. The protein SYCRP1, encoded by *sycrp1*, functions as a cAMP receptor protein. It has been shown that SYCRP1 is cAMP-specific, *i.e.* it does not bind cGMP [29]. The complex between SYCRP1 and cAMP (SYCRP1∼cAMP) is able to bind to certain DNA sites via a DNA-binding helix-turn-helix motif and affects the transcription of genes downstream of these sites. Experiments have shown that the SYCRP1∼cAMP complex binds to the consensus sequence for CRP (cAMP receptor protein) in *E. coli* [22]. SYCRP1∼cAMP is known to be involved in motility since the *sycrp1* mutant is immotile [31]. Database screening and sequence analysis indicate that the protein PdeB, encoded by *pdeB*, might be a phosphodiesterase [8]. The protein contains a HD-domain, which is a domain related to the catalytic region of eukaryotic phosphodiesterases [8]. Also this protein has a domain which is homologous to cAMP-binding domains [4] and it has the

same modular organization as cAMP-specific PDE in *Dictyostelium* [8]. The connections between the genes and proteins mentioned so far are summarized in figure 3(b).

The protein Cya2, encoded by *cya2*, has been found by database screening and sequence analysis to be a putative nucleotide cyclase and to contain catalytic residues typical for nucleotide cyclases with specificity for cGMP [7]. The level of cAMP in a *cya2* mutant compared to the level in wild type cells was shown to be the same and to increase in the same way when exposed to light [7, 27]. However, the cGMP level in the mutant was only about one third of the cGMP level in wild type [7]. Taken together, these results strongly suggest that Cya2 is a guanylyl cyclase. A disruption mutant of *cya2* was still motile [26], showing that Cya2 by itself is not essential for motility. The protein SYCRP2, encoded by *sycrp2*, shows homology with cAMP receptor protein in *E. coli* [29]. Sequence analysis has also shown that the protein contains a cyclic nucleotide binding domain attached to a DNA-binding helix-turn-helix motif [8]. Because experiments have shown that the protein does not bind cAMP [29], it is considered a putative cGMP receptor protein. The protein PdeA, encoded by the gene *pdeA*, shows sequence similarity to phosphodiesterases [8]. Experiments have shown that PdeA binds cGMP but not cAMP[2], which suggests that this protein is a cGMP specific phosphodiesterase. These genes and proteins centered around cGMP are summarized in figure 3(c).

The six genes mentioned above and their corresponding proteins together with cAMP and cGMP constitute the regulation module, *i.e.* these are the components that have been analyzed in this work. It is likely that there are other proteins in the motility network that also might interact with the cyclic nucleotides. For example there are two proteins besides SYCRP1 and SYCRP2 that are also potential cAMP/cGMP receptor proteins. However, the information about these is vaguer than the information about the above mentioned components, so they have not been included in this study.

2.3 Other possible interactions in the regulation module

The facts mentioned so far about the six genes and their corresponding proteins cover their interactions with the cyclic nucleotides and the derivates of these, and for some of the components, how they are influenced by light and glucose. In order to regulate the system, it is probable that there might exist additional interactions with regulatory function between the genes and proteins. For example there might exist activations and inhibitions that constitute different kinds of feedback. Also, there might be interactions between the two subsystems centered around cAMP and cGMP, respectively. Based on the knowledge that cAMP and cGMP are known to be involved in complicated regulatory networks in other organisms, regulated by interactions between the different components, it is likely that regulatory interactions also exist here. Also, light might influence the system in other ways than via Cya1, and we do not know whether glucose has an effect on the behaviour of the system and, in that case, via which molecules it would influence the system. Here we will now discuss some potential regulatory interactions between the genes and proteins, as well as influences by light and glucose on the system. A more detailed account of the possibilities can be viewed in appendix A. The information in this subsection comes from three sources: (1) the literature, (2) interviews and discussions with the biologist and cyanobacteria specialist, Jean Houmard, researcher at Ecole Normale Supérieure (ENS) in Paris, and (3) the CyanoBase Internet site [6].

[2]Unpublished result from experiment performed by Jean Houmard, Ecole Normale Supérieure (ENS), Paris.

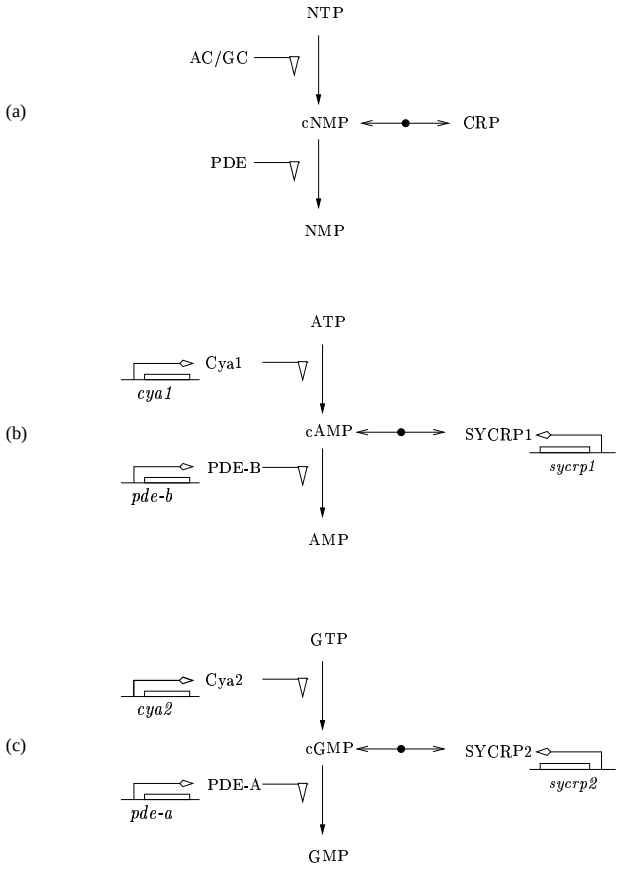

Figure 3: (a) The general network centered around the cyclic nucleotides in various organisms. Abbreviations: cNMP is a general notation for the cyclic nucleotides cAMP and cGMP, N standing for either A or G. Similar notation for NTP and NMP. AC/GC means adenylyl cyclase or guanylyl cyclase respectively, and PDE is phosphodiesterase. CRP is here an abbreviation for both cAMP and cGMP receptor proteins (not to mix up with CRP in *E. coli* which stands for cAMP receptor protein only). (b) and (c) show probable structures of the networks centered around cAMP and cGMP respectively in *Synechocystis PCC 6803*. The notation formalism used, proposed by Kohn [18], is shown in figure 4.

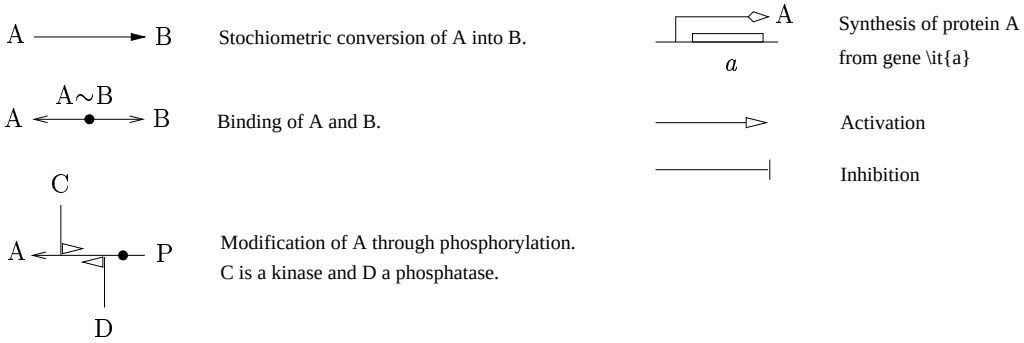

Figure 4: Graphical representation of cellular events, the notation used is the one suggested by Kohn [18].

The fact that the cAMP level in the wild type increases as a response to blue light, while a *cya1* mutant is insensitive to blue light [27], suggest that either the transcription of *cya1* or the activity of Cya1 or both must be activated by light. No experiments have been done yet on the change in cGMP level in a *cya2* mutant exposed to light, so it is not known whether *cya2* transcription or Cya2 activity is influenced by light, neither of these possibilities should be excluded. It is also conceivable that light influence the phosphodiesterases or the cAMP and cGMP receptor proteins.

The increase in cGMP as a response to glucose, might be due to a glucose inhibition of *pdeA* or PdeA or activation of *sycrp2* or SYCRP2. Whether cGMP is influencing any of the other genes and proteins in the network cannot be excluded, but there is no information on this.

As mentioned, SYCRP1 in complex with cAMP (SYCRP1∼cAMP) is capable to bind the consensus sequence of CRP in *E. coli* [22], indicating that this complex might regulate some of the genes in the network. Searches for sequences with homology to the consensus sequence of CRP in *E. coli* have revealed that such a sequence is situated upstream of *cya1* and *sycrp1* [8]. These facts make it plausible that *cya1* and *sycrp1* are regulated by the SYCRP1∼cAMP complex. Screening for target genes of the SYCRP1∼cAMP complex, by comparing microarrays for the wild type and a *sycrp1* mutant, have shown that all target genes found so far are upregulated by the complex [30], which suggests that it might be more likely to invoke activation rather than inhibition. On the other hand, CRP in *E. coli* is functioning both as an upregulator and a downregulator, and it has both positive and negative autoregulation [15, 21], which indicates that inhibition cannot be excluded, especially not in the case of the CRP homologue *sycrp1*.

SYCRP2, like SYCRP1, contains a helix-turn-helix motif [8]. The putative SYCRP2∼cGMP complex therefore likely possesses the ability to bind DNA and regulate genes in the system. This putative complex might either bind the same DNA sequence as the SYCRP1∼cAMP complex, and thereby also possibly regulate *cya1* and *sycrp1*, or it might recognize and bind to some other sequence.

It cannot be excluded that the cyclic nucleotides change the activity of the nucleotide cyclases or the phosphodiesterases. That would mean that the enzymes would be affected by their own product or substrate, respectively. These interactions are not very likely in this case, although it sometimes happens that enzyme activities are influenced by the binding of their own substrate or product.

All hypothetical interactions discussed above pose questions that would be interesting to investigate. The aim with this work is to be able to exclude or establish some of these or other interactions in the regulation module. For this, we have applied qualitative mathematical modeling and simulation to the network. Several different mathematical models, taking different hypothetical interactions into account, have been created and simulated. By comparing simulation results we will show in this report how performance of some selected experiments would discriminate between different hypothetical models.

3 Qualitative modeling and simulation of genetic regulatory networks

In order to study the biological network described in the last section, a modeling and simulation method that does not require too much data about the system is needed, since the information about the system is sparse. All since the idea of applying modeling and simulation on molecular networks was born in the sixties, most methods that have been developed need an extensive amount of quantitative information about the system. The recently developed qualitative modeling and simulation method used in this work [11], has the advantage that it does not require any quantitative information, but instead qualitative information is enough. The method is based on piecewise-linear differential equations (PLDEs) and it provides a coarse-grained description of genetic regulatory networks. In this section, a summary of the qualitative modeling and simulation method we have used will be given. Most part of the text in sections 3.1-3.5 comes from the report of Ropers *et al.* [24]. Slight modifications have been done in order to adapt the text to this work.

3.1 Piecewise-linear models of genetic regulatory networks

The dynamics of genetic regulatory networks can be modeled by a class of piecewise-linear (PL) differential equations originally proposed by Glass and Kauffman [13], and generalized by Mestl *et al.* [19]. The equations have the form

$$\dot{x}_i = f_i(\boldsymbol{x}) - g_i(\boldsymbol{x})\, x_i, \quad x_i \geq 0,\ 1 \leq i \leq n, \tag{1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)'$ is a vector of cellular protein concentrations. The *state equations* (1) define the rate of change of the concentration x_i as the difference of the rate of synthesis $f_i(\boldsymbol{x})$ and the rate of degradation $g_i(\boldsymbol{x})\, x_i$ of the protein.

The function $f_i : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ expresses how the rate of synthesis of the protein encoded by gene i depends on the concentrations $\boldsymbol{x}$ of proteins in the cell. It is defined as

$$f_i(\boldsymbol{x}) = \sum_{l \in L} \kappa_{il}\, b_{il}(\boldsymbol{x}), \tag{2}$$

where κ_{il} is a rate parameter ($\kappa_{il} > 0$), $b_{il} : \mathbb{R}^n_{\geq 0} \to \{0, 1\}$ a *regulation function*, and L a possibly empty set of indices of regulation functions. The function g_i describes the regulation of protein degradation. It is defined analogously to f_i, except that we demand that $g_i(\boldsymbol{x})$ is strictly positive. In addition, in order to formally distinguish degradation rates from synthesis rates, we will denote the former by γ instead of κ. Notice that with the above definitions of f_i and g_i, the state equations (1) are *piecewise-linear (PL)*.

A regulation function b_{il} describes the logic of gene regulation [25]. More precisely, it describes the conditions under which the protein encoded by gene i is synthesized (degraded) at a rate κ_{il} ($\gamma_{il}\, x_i$). These conditions are formulated as expressions of step functions $s^+, s^- : \mathbb{R}^2 \to \{0, 1\}$:

$$s^+(x_j, \theta_j) = \begin{cases} 1, & x_j > \theta_j, \\ 0, & x_j < \theta_j, \end{cases} \qquad \text{and} \qquad s^-(x_j, \theta_j) = 1 - s^+(x_j, \theta_j), \tag{3}$$

where x_j is an element of the state vector $\boldsymbol{x}$, and θ_j a constant denoting a threshold concentration ($\theta_j > 0$). Notice that step functions $s^+(x_j, \theta_j)$ and $s^-(x_j, \theta_j)$ are not defined for $x_j = \theta_j$,

so neither are the regulation functions in which they occur. We use regulation functions that are the arithmetic equivalent of logical functions, as described in [23].

The PL models can be extended to take into account *input variables* $\boldsymbol{u} = (u_1, \ldots, u_m)'$, representing the concentration of proteins and small molecules whose synthesis and degradation are regulated outside the system. This leads to models of the form:

$$\dot{x}_i = f_i(\boldsymbol{x}, \boldsymbol{u}) - g_i(\boldsymbol{x}, \boldsymbol{u})\, x_i, \quad x_i \geq 0,\ u_j \geq 0,\ 1 \leq i \leq n,\ 1 \leq j \leq m, \tag{4}$$

In what follows, we will assume that the input variables are constant, *i.e.*, $\dot{\boldsymbol{u}} = \boldsymbol{0}$. As a consequence, (4) can be reduced to (1) without loss of generality, by prior evaluation of the step-function expressions in which input variables occur.

3.2 Mathematical analysis of piecewise-linear models

The dynamical properties of the PL models can be analyzed in the n-dimensional phase space box $\Omega = \Omega_1 \times \ldots \times \Omega_n$, where every Ω_i, $1 \leq i \leq n$, is defined as

$$\Omega_i = \{x_i \in \mathbb{R}_{\geq 0} \mid 0 \leq x_i \leq max_i\}, \tag{5}$$

and max_i is a parameter denoting a maximum concentration for the protein.

Given that the protein encoded by gene i has p_i threshold concentrations, the $(n-1)$-dimensional threshold hyperplanes $x_i = \theta_i^{k_i}$, $1 \leq k_i \leq p_i$, partition Ω into hyperrectangular regions that are called *domains*. More precisely, a domain $D \subseteq \Omega$ is defined by $D = D_1 \times \ldots \times D_n$, where every D_i, $1 \leq i \leq n$, is given by one of the equations below:

$$\begin{aligned}
D_i =& \{x_i \mid 0 \leq x_i < \theta_i^1\},\\
D_i =& \{x_i \mid x_i = \theta_i^1\},\\
D_i =& \{x_i \mid \theta_i^1 < x_i < \theta_i^2\},\\
& \ldots\\
D_i =& \{x_i \mid \theta_i^{p_i} < x_i \leq max_i\}.
\end{aligned} \tag{6}$$

If for a domain D, there are some i, j, $1 \leq i \leq n$, $1 \leq j \leq p_i$, such that $D_i = \{x_i \mid x_i = \theta_i^j\}$, then D is called a *switching domain*. Otherwise, D is called a *regulatory domain*.

When evaluating the step-function expressions in (2) in a regulatory domain, $f_i(\boldsymbol{x})$ and $g_i(\boldsymbol{x})$ reduce to sums of rate constants. More precisely, in every regulatory domain $D \subseteq \Omega$, $f_i(\boldsymbol{x})$ reduces to some $\mu_i^D \in M_i \equiv \{f_i(\boldsymbol{x}) \mid \boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{max}\}$, and $g_i(\boldsymbol{x})$ to some $\nu_i^D \in N_i \equiv \{g_i(\boldsymbol{x}) \mid \boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{max}\}$. M_i and N_i collect the synthesis and degradation rates of the protein in different domains of Ω. It can be easily shown that all trajectories in D monotonically converge towards the so-called *target equilibrium*, lying at the intersection of the $(n-1)$-dimensional threshold hyperplanes $x_i = \mu_i^D/\nu_i^D$ $(1 \leq i \leq n)$ [13, 19, 25]. The target equilibrium level μ_i^D/ν_i^D of the protein concentration x_i gives an indication of the strength of gene expression in D. Call $\Psi(D) = \{(\mu_1^D/\nu_1^D, \ldots, \mu_n^D/\nu_n^D)'\}$ the *target equilibrium set* of D. If $\Psi(D) \cap D \neq \{\}$, then all solutions in D asymptotically approach the target equilibrium, which is then a stable equilibrium point of the system. If $\Psi(D) \cap D = \{\}$, the solutions will leave D at some point.

In switching domains, $f_i(\boldsymbol{x})$ and $g_i(\boldsymbol{x})$ are not defined in general, because some concentration variables assume a threshold value. Moreover, due to the use of step functions, $f_i(\boldsymbol{x})$

and $g_i(\boldsymbol{x})$ may be discontinuous at the $(n-k)$-dimensional threshold hyperplane in which the domain is contained ($1 \leq k \leq n$). In order to cope with this problem, the system of differential equations (1) is extended into a system of differential inclusions, following an approach inspired by Filippov [11, 12, 14]. Using this generalization, it can be shown that, in the case of a switching domain D, the solution trajectories either cross D instantaneously or converge towards a target equilibrium set $\Psi(D)$ located in the threshold hyperplane containing D. Here, $\Psi(D)$ is the smallest hyperrectangle containing the target equilibria of the regulatory domains that have D in their boundary, intersected with the threshold hyperplane containing D. In the case of switching domains, $\Psi(D)$ is generally not a single point. If $\Psi(D) \cap D = \{\}$ all solutions will leave D at some point. On the other hand, if $\Psi(D) \cap D \neq \{\}$, there exist solutions in D that reach or asymptotically approach the target equilibrium set $\Psi(D)$ as $t \to \infty$. Every $\boldsymbol{\psi} \in \Psi(D) \cap D$ is an equilibrium point of the system. Whether this equilibrium point is stable or unstable must be determined through further analysis [5].

3.3 Qualitative description of dynamics of piecewise-linear models

The mathematical framework presented in the previous section suggests an intuitive qualitative description of the dynamics of regulatory systems described by PL models (1). This description is based on a discrete abstraction of the state of the system, a so-called *qualitative state*, consisting of the domain D in which the system resides and the position of the target equilibrium set $\Psi(D)$ with respect to D. A qualitative state thus captures the local dynamics of the system. There exists a *transition* between two qualitative states QS and QS', corresponding to contiguous domains D and D', if some solution trajectories starting in D reach D', without passing through an intermediate domain. The sets of qualitative states and transitions between qualitative states define a *state transition graph*, summarizing the qualitative dynamics of the regulatory system.

The state transition graph may contain one or more *qualitative equilibrium states*, each of which corresponds to an equilibrium point of the system. A path in the state transition graph is called a *qualitative behavior*. It describes how the bounds on protein concentrations evolve over time, according to the sequence of transitions between qualitative states. A cyclic qualitative behavior, a so-called *qualitative cycle*, may correspond to a limit cycle or to trajectories spiraling towards or from an equilibrium point. The set of qualitative states from which a qualitative equilibrium state or qualitative cycle is reachable form its *attraction set*. The qualitative nature of the state transition graph is well-adapted to measurement techniques in genomics, which currently have limited quantitative precision, but are able to detect qualitative changes in gene expression over time.

3.4 Qualitative piecewise-linear models

Most of the time, precise numerical values for the threshold and rate parameters in a PL model are not available. However, instead of specifying precise numerical values, it is often possible to supplement the state equations with inequality constraints on the parameter values. The inequality constraints express weak, but reliable information about the regulatory interactions that can be inferred from biological data. The resulting, so-called *qualitative PL model*, subsumes a set of quantitative PL models, the qualitative dynamics of each of which can be described by means of a state transition graph.

The first type of constraint, the *threshold inequalities*, are obtained by ordering the p_i threshold concentrations of the protein encoded by gene i, *i.e.*,

$$0 < \theta_i^1 < \ldots < \theta_i^{p_i} < max_i. \tag{7}$$

The threshold inequalities determine the partitioning of Ω into regulatory and switching domains.

The second type of constraint, the *equilibrium inequalities*, order the possible target equilibrium levels of x_i in different regulatory domains D with respect to the threshold concentrations. Biologically speaking, the equilibrium inequalities define the strength of gene expression in the domain in a qualitative way, on the scale of ordered threshold concentrations. More precisely, for every $\mu_i \in M_i$, $\nu_i \in N_i$, and $\mu_i, \nu_i \neq 0$, we specify one of the following pairs of inequalities:

$$\begin{aligned} &0 < \mu_i/\nu_i < \theta_i^1, \\ &\theta_i^1 < \mu_i/\nu_i < \theta_i^2, \\ &\ldots \\ &\theta_i^{p_i} < \mu_i/\nu_i < max_i. \end{aligned} \tag{8}$$

The equilibrium inequalities constrain the relative position of D and its target equilibrium set $\Psi(D)$.

The models of genetic regulatory networks treated by the simulation method consist of state equations (1), supplemented by parameter inequalities (7) and (8). Every such *qualitative* PL model corresponds to a set of *quantitative* PL models consisting of state equations (1) and a particular combination of numerical parameter values consistent with the parameter inequalities. It has been shown that in the region of the parameter space defined by the inequalities in the qualitative PL model, all quantitative PL models yield the same state transition graph [11]. This graph can be efficiently computed from the inequality constraints by symbolic instead of numerical means.

3.5 Qualitative simulation

A state transition graph may become exceedingly large, as the number of domains, and hence qualitative states, grows exponentially with the dimension of the system. For many purposes, it is sufficient to know which qualitative states are reachable from a given initial qualitative state, that is, which qualitative behaviors the system can exhibit when initially being in this state. The algorithm for *qualitative simulation* described in [11] generates the reachable part of a state transition graph from a qualitative PL model and an initial domain.

The state transition graph generated through qualitative simulation is a prediction of the qualitative dynamics of the system. This state transition graph has been shown to cover all qualitative behaviors permitted by any one of the quantitative PL models subsumed by the qualitative PL models [11]. That is, whatever the exact numerical values for the parameters may be, if these values are consistent with the threshold and equilibrium inequalities specified in the qualitative PL model, the qualitative shape of the solution is described by a sequence of states in the state transition graph. The inverse is not true: the state transition graph resulting from a qualitative simulation may contain qualitative behaviors that are not permitted by some quantitative PL model subsumed by the qualitative PL model.

The simulation method has been implemented in Java 1.4, in a program called *GNA (Genetic Network Analyzer)* [10].[3] Given input data, *i.e* the state equations, inequality constraints, and the initial domain, the program produces a state transition graph. To analyze this graph, GNA provides a number of user-friendly tools. For example it is possible to select a part of the graph and study the temporal evolution of qualitative concentration levels for this part in more detail.

[3]GNA is available for non-profit academic research purposes at `http://www-helix.inrialpes.fr/gna`.

4 Qualitative models of motility regulation network

In this section, the regulation module of the motility network in *Synechocystis* PCC 6803, described in section 2, will be modeled with the method described in the previous section. Since the information on the biological system is incomplete, we have to make assumptions in order to create the models. Since many different assumptions can be made, there are many models that might describe the phototaxis and chemotaxis of *Synechocystis* PCC 6803. Therefore we will start with simple models, built on the best established facts, and then make more complex alternative models by adding less certain facts. In the first subsection, the assumptions that are made in *all* of the models will be presented. In the following subsections, four classes of models, in order of increasing complexity, will be presented. For some models it will be described in detail how the state equations and inequality constraints are derived from the biological properties of the system, while for others only parts of the mathematical models are explained. The complete models (state equations and inequality constraints) are given in appendix B. In all interaction schemata in this section, the notation proposed by Kohn [18] is used, see figure 4.

4.1 Assumptions made in all models

Based on the facts in section 2 it is possible to make some assumptions that we are more or less sure about and that will be used in all models. Here we present these assumptions.

Cya1, which is known to be an adenylyl cyclase [26], is assumed to have no guanylyl cyclase activity, *i.e.* it is assumed to not be able to use GTP as a substrate. Cya2 is assumed to be a guanylyl cyclase and not to be able to use ATP as a substrate [7]. Further, we assume that SYCRP2 is cGMP-binding and not cAMP-binding, and that the SYCRP2∼cGMP complex is a transcription factor, even though this has not been conclusively demonstrated (see section 2).

The phosphodiesterases PdeA and PdeB are not included in the mathematical models for two reasons. Firstly, they are not as well studied as the above-mentioned proteins. Secondly, adding these proteins would have the same mathematical effect as decreasing the adenylyl/guanylyl activity. More specifically, an activation of a phosphodiesterase would have the same effect as an inhibition of an adenylyl/guanylyl cyclase and *vice versa* and therefore it would not be possible to differentiate between these two events, given currently available data.

The light and glucose inputs, as well as the motility output, are modeled as either on or off. This level of abstraction is well-adapted to currently available data from the experiments performed by Jean Houmard and from the articles [27] and [20] (see also section 2).

When light is present and glucose is absent, the motility is modeled to be at its high level, according to the facts given in section 2. Since light is necessary for motility, the motility is modeled as low when light is absent, no matter whether glucose is present or not. We also model the motility as low when both glucose and light are present. The reason for this is that the cGMP level is high under these conditions. However, the connection between cGMP and motility is unclear (see explanation in section 2 of experiment performed by Jean Houmard and [16]). The behaviour of the bacterium in response to different light and glucose conditions is summarized in table 2. These responses are design criteria for model building, *i.e.*, all models with both light and glucose inputs have been constructed with the aim to satisfy these constraints. In some models, only the light input has been included, and in these, the glucose level is assumed to be on a constant low level. Therefore, these models are

Motility	light on	light off
glucose on	low	low
glucose off	high	low

Table 2: The motility level (high or low) under different light and glucose conditions.

expected to give high motility output when light is present, and low when absent.

The last and least supported assumption is that light and glucose are acting on the Cya1 and Cya2 proteins, respectively, and not on the genes. Our motivation is that in order for the light and glucose signals to stimulate a quick response, it would be more effective for the signals to act on the proteins, since protein interactions generally occur on a much faster time scale than gene expression. This is supported by the way the glucose signal in *E. coli* [24] and the starvation signal in *B. subtilis* [9] are acting.[4]

4.2 Class I models: only light influence

Since there is not much information on the glucose influence on motility, models that take only the light influence into account will be considered here. According to the assumptions above, these models include only the genes *cya1* and *sycrp1*. Because there are sequences homologous to the *E. coli* CRP consensus sequence upstream of both these genes (see section 2 and appendix A comment 16 and 20), they are for this class of models assumed to be autoregulated by SYCRP1∼cAMP, either positively or negatively. This gives us four possible models in this class.

4.2.1 Model I1

Since we have some indications that the regulation of *sycrp1* by SYCRP1∼cAMP is more likely negative than positive (see appendix A, comment 20), this will be assumed in the first model, called model I1 (I for class one and 1 for being the first model in this class). In this model, the regulation of *cya1* is also negative, by analogy to *cya* in *E. coli* (see appendix A, comment 16). This means that the regulation schema looks as in figure 5.

Because the synthesis and degradation of cAMP, as well as its association with and dissociation from SYCRP1, can be assumed to take place on a much faster time-scale than the synthesis and degradation of Cya1 and SYCRP1, the former processes can be regarded to be in quasi-equilibrium with respect to the latter. How this can be used to make a rigorous description of the kinetics of the corresponding system in *E. coli* is given in the research report of Ropers *et al.* [24]. In this report it is shown how the activity of the gene *crp* in *E. coli* (homologous to *sycrp1*), depends on the concentration of CRP and Cya (the latter is homologous to Cya1 and Cya2 in *Synechocystis* PCC 6803) and of the presence of a signal, see figure 6. It is also explained how this function can be approximated by a step function expression.

In analogy to the corresponding system in *E. coli*, the activity of *sycrp1* can be approximated by a similar step function expression as the one for CRP. Let x_{s1} and x_{c1} be the

[4]Alternative models taking into account that the signals influence gene expression have been tested. The simulations have shown the same qualitative effect on motility as the models that assume the signals to influence protein activity.

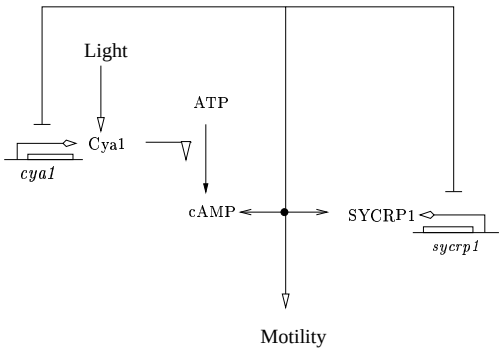

Figure 5: Interaction schema for model I1. When Cya1 is activated by light, it catalyzes the conversion of ATP into cAMP. cAMP binds to SYCRP1 and the complex activates genes that induce motility, and inhibits the genes *cya1* and *sycrp1*.

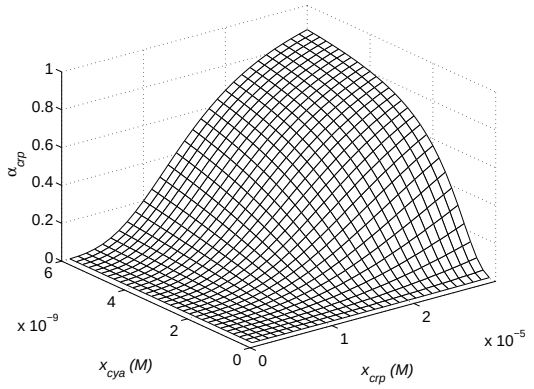

Figure 6: Activity of the gene *crp* (α_{crp}) in the presence of a signal, as derived in [24]. The activity depends on the concentration of CRP (x_{crp}) and Cya (x_{cya}) and is normalized on the scale 0 to 1. In *Synechocystis* PCC 6803 we assume that the activity of *sycrp1* depends on the concentrations of Cya1 and SYCRP1 in a similar way to how the activity of *crp* depends on the concentrations of CRP and Cya.

concentrations of SYCRP1 and Cya1 respectively, and let the input variable u_l denote the light signal. The step function $s^+(u_l, \theta_l)$ evaluates to 1 if light is present and 0 otherwise. From the considerations in [24] we get that the step function expression

$$s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)$$

evaluates to 1 if the concentration of the SYCRP1∼cAMP complex is above a certain threshold, and 0 otherwise. This means that the concentrations of SYCRP1 and Cya1 have to be above certain thresholds, θ^2_{s1} and θ^2_{c1}, and light has to be present, in order for the concentration of the complex to be above its threshold.

Given the meaning of the above expression and supposing that *cya1* is expressed at the rate κ^1_{c1} when SYCRP1∼cAMP is present, and at the rate $\kappa^1_{c1} + \kappa^2_{c1}$ when the complex is absent, we can express the time derivative of the concentration of Cya1 as

$$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1}(1 - s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)) - \gamma_{c1}x_{c1}$$

where $\gamma_{c1}x_{c1}$ is the degradation rate of Cya1. Similarly, the time derivative of the concentration of SYCRP1 can be expressed as

$$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1}(1 - s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)) - \gamma_{s1}x_{s1}$$

where $\gamma_{s1}x_{s1}$ is the degradation rate of SYCRP1, κ_{s1} is the expression rate when the SYCRP1∼cAMP complex is present, and $\kappa^1_{s1} + \kappa^2_{s1}$ is the expression rate when the complex is absent.

Note that in this model, both genes *cya1* and *sycrp1* are repressed by the complex when its concentration is above the *same* threshold. The reasons for this are that we do not know what the real thresholds are and also it does not change the motility as long as these thresholds both are above the threshold for activating the genes in the motility response module. The reason that this latter threshold should be lower than the threshold for activation of the genes, is that otherwise the negative autoregulation would curb the concentration of the complex so that its concentration would never be high enough to activate the genes in the motility response module, and we would not get motility in response to light. This latter threshold, which needs to be reached in order for motility to be turned on, is exceeded when the expression

$$s^+(x_{s1}, \theta^1_{s1})s^+(x_{c1}, \theta^1_{c1})s^+(u_l, \theta_l)$$

evaluates to 1. Here, $\theta^1_{c1} < \theta^2_{c1}$ and $\theta^1_{s1} < \theta^2_{s1}$ are the corresponding thresholds for Cya1 and SYCRP1, and light can still only be on its two previous levels.

Let m be a gene in the motility module which is a target of SYCRP1∼cAMP. Let us regard motility as turned on when the concentration of the product M of gene m is above the threshold θ_m, and off otherwise. Let the time derivative of the concentration of M be expressed by

$$\dot{x}_m = \kappa_m s^+(x_{s1}, \theta^1_{s1})s^+(x_{c1}, \theta^1_{c1})s^+(u_l, \theta_l) - \gamma_m x_m$$

where $\kappa_m > 0$ and $\gamma_m x_m$ describes the degradation of the protein M.

Now it remains to specify the equilibrium inequalities. When the genes *cya1* and *sycrp1* are expressed at their low level, they are expected to induce motility, which means that their target equilibrium values have to be above the threshold values θ^1_{c1} and θ^1_{s1}, respectively, in order to be able to activate motility. At the low level of expression, the complex cannot regulate the genes, since otherwise the gene expression would always be repressed. Therefore, the target equilibrium values for the genes have to be below the higher thresholds. More exactly we have:

$$\theta^1_{c1} < \kappa^1_{c1}/\gamma_{c1} < \theta^2_{c1} \text{ and } \theta^1_{s1} < \kappa^1_{s1}/\gamma_{s1} < \theta^2_{s1}$$

When the genes are expressed at the high level, the concentration of the complex is high enough to both induce motility and to negatively autoregulate the genes. Therefore we get

$$\theta^2_{c1} < (\kappa^1_{c1} + \kappa^1_{c1})/\gamma_{c1} < max_{c1} \text{ and } \theta^2_{s1} < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < max_{s1}$$

Also, in order for the motility variable x_m to be able to cross its threshold and thereby turn on motility, we need to have

$$\theta_m < \kappa_m/\gamma_m < max_m$$

Thus all state equations and inequality constraints have been defined for model I1.

4.2.2 Other class I models: I2-I4

In model I2 (see figure 7) we keep the regulation of *sycrp1* negative but change the regulation of *cya1* to positive, since we have a little bit more information about the regulation of the former gene (compare comments 16 and 20 in appendix B). In this case, *cya1* is activated by

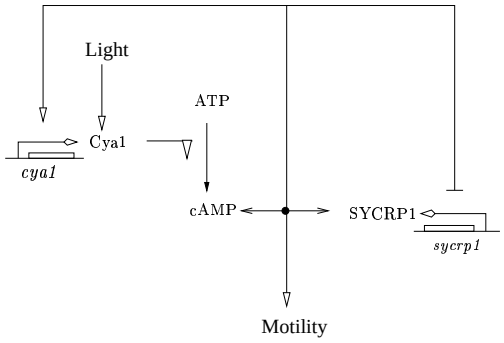

Figure 7: Interaction schema for model I2. The SYCRP1∼cAMP complex activates *cya1* and inhibits *sycrp1*.

the SYCRP1∼cAMP complex already at a low concentration of the complex. This concentration level is not high enough for either turning on motility or inhibition of *sycrp1*. When the complex concentration exceeds a second level, it turns on motility, but still there is no inhibition of *sycrp1*. Finally, when the complex concentration exceeds a third level, we also have inhibition of *sycrp1*. Hence, in this case we need three thresholds for each gene:

$0 < \theta^1_{c1} < \theta^2_{c1} < \theta^3_{c1} < max_{c1}$ and $0 < \theta^1_{s1} < \theta^2_{s1} < \theta^3_{s1} < max_{s1}$

The expression of *cya1* is here at a low level κ^1_{c1} if the complex is absent, and at a high level, $\kappa^1_{c1} + \kappa^2_{c1}$ if the complex is present. This yields a different state equation for *cya1* as compared to the previous model:

$$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} s^+(x_{s1}, \theta^1_{s1}) s^+(x_{c1}, \theta^1_{c1}) s^+(u_l, \theta_l) - g_{c1} x_{c1}$$

The other state equations are only slightly different from the ones for model I1 (for details see appendix B).

The regulatory interactions in model I3 are analogous to the ones in *E. coli*. That is, in this model *cya1* is inhibited and *sycrp1* is activated by the SYCRP1∼cAMP complex, see figure 8.[5] Since there exist both up- and downregulation by the complex also in this module, three thresholds for the genes are needed. The state equations and inequality constraints, shown in appendix B, are derived by similar reasoning as for the previous models.

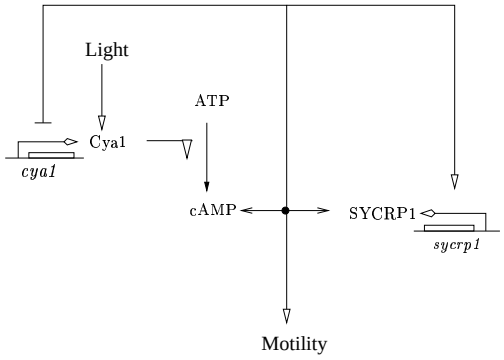

Figure 8: Interaction schema for model I3. The SYCRP1∼cAMP complex inhibits *cya1* and activates *sycrp1*.

In the fourth and last possible model in this class, model I4, SYCRP1∼cAMP activates both genes, see figure 9. In this case, the thresholds for activation of the genes need to be lower than the thresholds for turning on motility, since otherwise the motility would be on even at low concentrations of the complex, no matter if there were autoregulation or not, which means that the positive autoregulation would have no meaning.

4.3 Class II models: Light influence and simple glucose influence

Glucose is known to increase the cGMP level and decrease motility. We also know that Cya2 is a guanylyl cyclase and SYCRP2 probably is a cGMP receptor protein. From this sparse information, the simplest network that can be built does not have any regulation by SYCRP2∼cGMP except inhibition of the motility genes. That is, we use one of the models in class I, and add a glucose influence in the way shown in figure 10. In this figure, the network is based on model I1 from class I, and we call this extended model II1. The models with the same type of glucose influence, but based on models I2, I3, and I4, are called II2, II3, and II4

[5] To be more exact, in *E. coli* the gene *crp* can actually be negatively and positively regulated by the CRP∼cAMP complex, but the positive interaction seems more important [24].

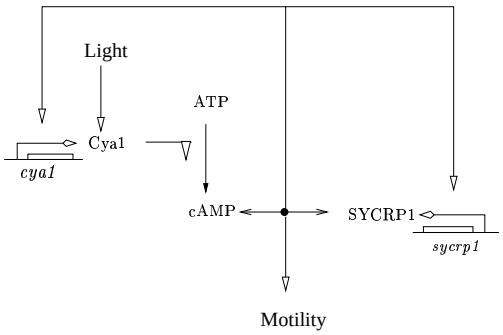

Figure 9: Interaction schema for model I4. The SYCRP1∼cAMP complex activates both *cya1* and *sycrp1*.

respectively. Here, we will explain how the state equation for the gene m is changed in the case of model II1.

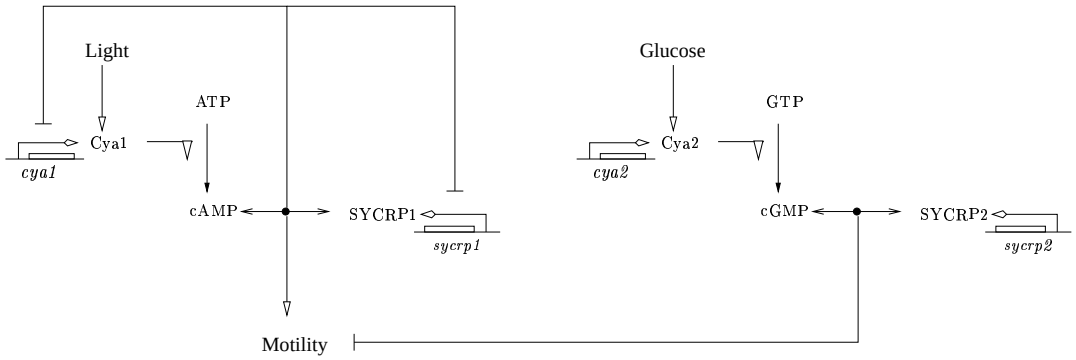

Figure 10: Interaction schema for model II1. To model I1, a negative effect from SYCRP2∼cGMP on the genes that induces motility, is added. When Cya2 is activated by glucose, it can convert GTP into cGMP and cGMP forms a complex with SYCRP2. No regulations by SYCRP2∼cGMP, except from on the motility genes, are taken into account in the class II models.

Let x_{c2} and x_{s2} be the concentrations of Cya2 and SYCRP2 respectively, and let the input variable u_g denote the glucose signal. The step function $s^+(u_g, \theta_g)$ evaluates to 1 if glucose is present and 0 otherwise. From the considerations in [24] we have for the SYCRP2∼cGMP complex, analogously to the SYCRP1∼cAMP complex, that

$$s^+(x_{c2}, \theta_{c2})s^+(x_{s2}, \theta_{s2})s^+(u_g, \theta_g)$$

evaluates to 1 if the concentration of SYCRP2∼cGMP is above a certain threshold, and 0 when it is below this threshold. θ_{c2} and θ_{s2} are the thresholds that the concentrations of Cya2 and SYCRP2 have to exceed in order for SYCRP2∼cGMP to be above its threshold. The assumption that both the presence of light as well as the absence of glucose are necessary for motility, gives a logical 'and' expression in the state equation for the motility gene m:

$$\dot{x}_m = \kappa_m s^+(x_{s1}, \theta^1_{s1})s^+(x_{c1}, \theta^1_{c1})s^+(u_l, \theta_l)(1 - s^+(x_{c2}, \theta_{c2})s^+(x_{s2}, \theta_{s2})s^+(u_g, \theta_g)) - \gamma_m x_m$$

The expression rate of the gene m will be κ_m if and only if both the conditions 'light is present' and 'glucose is absent' are fulfilled, and 0 otherwise. For the other state equations and inequality constraints, see appendix B.

4.4 Class III models: Parallel light and glucose influence

The next level of complexity is obtained by adding autoregulations to the glucose part of the network. Since we have no information about binding sites for SYCRP2∼cGMP (we do not even know if it is DNA-binding), one reasonable assumption is that a 'copy and paste' process has taken place in the evolution. That is, the genes from the light part of the network might have been copied, and the copy slightly changed, so that it now responds to a glucose signal. With this assumption it is reasonable to assume that the glucose part of the network is like a mirror of the light part. For example, as in model III1 (see figure 11), if SYCRP1∼cAMP negatively autoregulates *cya1* and *sycrp1*, we assume that SYCRP2∼cGMP analogously negatively autoregulates *cya2* and *sycrp2*. However, we have to assume that the

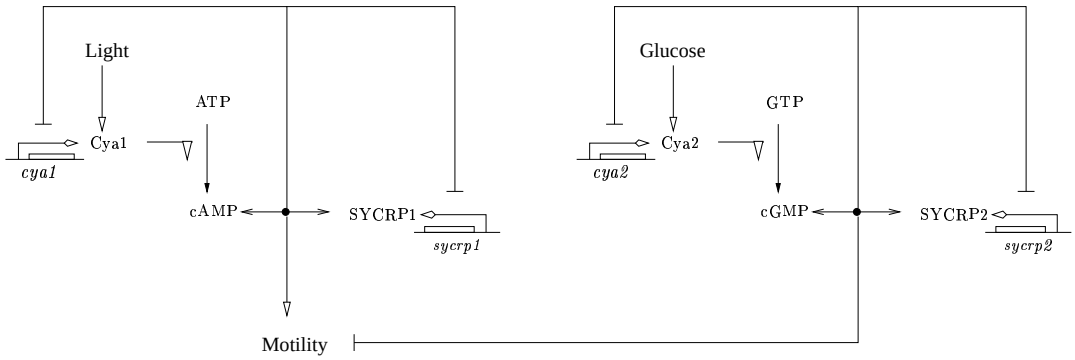

Figure 11: Interaction schema for model III1. To model II1 is added negative regulation from SYCRP2∼cGMP on *cya2* and *sycrp2*. The regulations on the right side of the network are a copy of the regulations on the left side in the class III models.

influence of SYCRP2∼cGMP on motility is negative in order to get an inhibitory effect of glucose on motility. The state equations and inequality constraints are derived in the same way as above and can be seen in appendix B.

4.5 Class IV models: cGMP acts via cAMP on the motility

A fourth possibility is to add regulations from SYCRP2∼cGMP on the *cya1* or *sycrp1* genes, having only SYCRP1∼cAMP (and not SYCRP2∼cGMP) act directly on the motility response genes. Models with these types of interactions constitute class IV (see figure 12 and 13 for examples).

The opposite (SYCRP1∼cAMP act on the *cya2* or *sycrp2* genes, and SYCRP2∼cGMP act directly on the motility response module) is not a possible model, since it is known that a *cya2* mutant retains motility when exposed to light [26]. Other possibilities, such as having both SYCRP1∼cAMP and SYCRP2∼cGMP regulating genes in both the cAMP and the cGMP part of the network, are also conceivable. But since sites homologous to the CRP consensus sequence have been found upstream of *cya1* and *sycrp1*, but not upstream of *cya2*

and *sycrp2*, and we want to regard the simplest models, we limit this class to contain the interactions mentioned above.

Since glucose is expected to have a negative influence on motility, it is reasonable to assume negative regulation of *cya1* and *sycrp1* by SYCRP2∼cGMP. In the network for model IV1, shown in figure 12, there are negative regulations of *cya1* and *sycrp1* from both complexes on *cya1* and *sycrp1*.

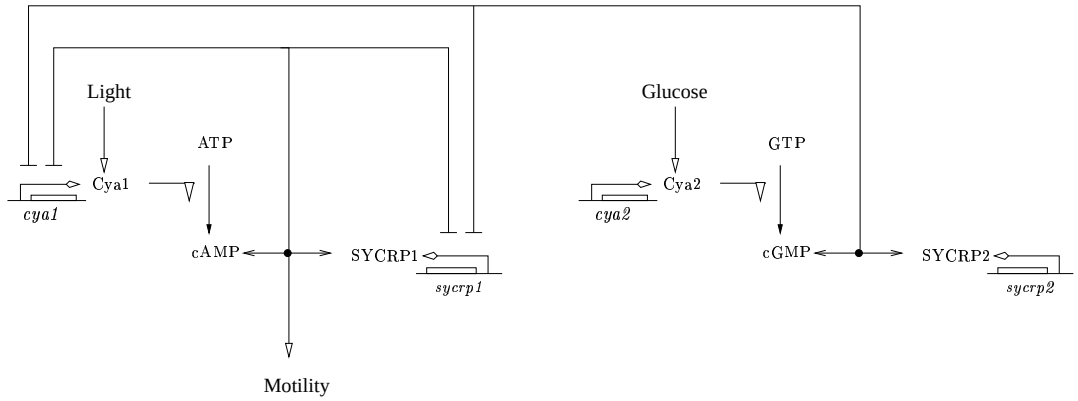

Figure 12: Interaction schema for model IV1. To model I1 is added an influence from glucose. By increasing the level of SYCRP2∼cGMP, which in turn inhibit the genes *cya1* and textitsycrp1, glucose decreases motility indirectly. The class IV models all act indirectly on motility through the SYCRP1∼cAMP pathway.

Another possibility is that the two complexes SYCRP1∼cAMP and SYCRP2∼cGMP compete for the same binding sites upstream of *cya1* and *sycrp1*. SYCRP1∼cAMP activates gene expression but SYCRP2∼cGMP binds to the site, without activating the genes, and thus blocks the site from binding the other complex (see figure 13. We call this model IV4, since it is an extension of model I4 (which also contain positive regulation from SYCRP1∼cAMP on *cya1* and *sycrp1*). For details about the mathematical models, see appendix B.

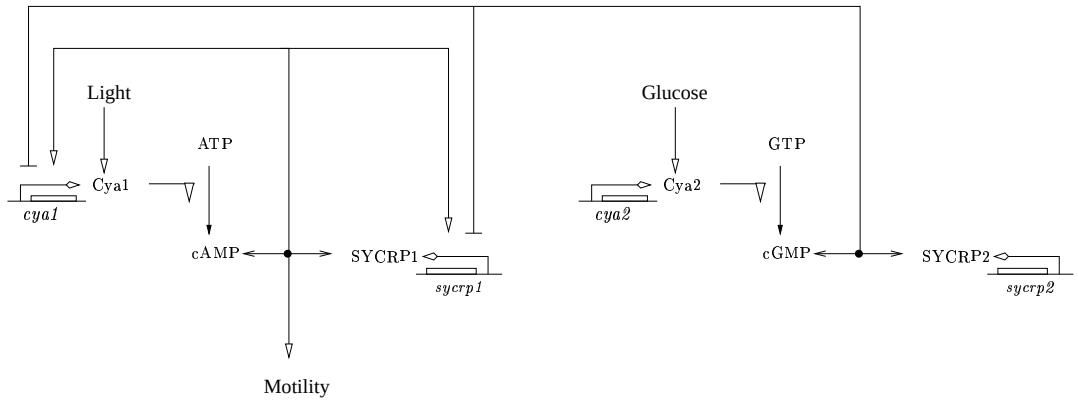

Figure 13: Interaction schema for model IV4. The SYCRP1∼cAMP and the SYCRP2∼cGMP complexes compete about the binding sites upstream of *cya1* and *sycrp1*. The gene is activated only when SYCRP1∼cAMP is bound, and SYCRP2∼cGMP blocks the binding of SYCRP1∼cAMP.

5 Simulation of motility regulation

As seen in the previous section, there are many possible models of the genetic regulatory network underlying the motility response in *Synechocystis* PCC 6803. Here we will first present the result of the qualitative simulations of these models with the method presented in section 3. The state equations and inequality constraints have been entered in GNA, and running the program results in a state transition graph, showing all states that are reachable from the initial state(s). It is possible to chose a sequence of states in the state transition graph and obtain a set of plots showing the time evolution of the different variables in the model. For an example of a state transition graph and time evolution graphs, see figure 14.

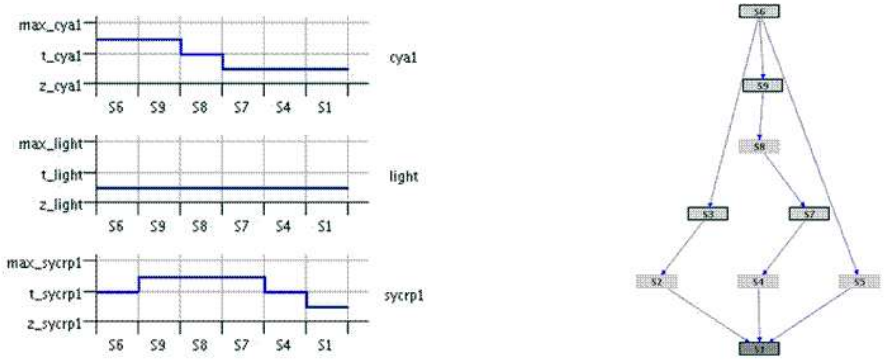

Figure 14: Screenshots from GNA. To the left: graph of the time evolution of the state variables *cya1*, light and *sycrp1*. To the right: state transition graph.

In subsection 5.2, following the results of the simulations of models derived in section 4, some models have been modified in order to mimic bacteria mutants. Simulations of the modified models have shown how it would be possible to discriminate between some groups of models by performing selected experiments.

5.1 Simulating the motility network

When *Synechocystis* PCC 6803 is exposed to light it responds by moving away from it. Also, we know that glucose induces aggregation of the bacteria, which is an indication of decreased motility. A more detailed description of what we know about the motility as a response to light and glucose is given in section 2. The assumptions we have made about the expected motility output, based on the facts in section 2, are described in section 4.1. As mentioned in the latter (section 4.1), the models of the motility network containing both light and glucose input have been constructed in such a way that they are expected to give the motility output shown in table 2. In the models in class I, having only light as input variable, the glucose level is assumed to be on a constant low level, so these models are expected to give high motility output when light is present, and low when absent.

When run in GNA, all models in section 4 have been shown to behave as expected, *i.e.* they give the expected motility outputs explained above. By these results, none of the constructed models can be rejected.

How the motility output arises from the molecular interactions according to the different models, can be understood by studying the state transition graphs produced by GNA. For simple models, as most of our models are, it is also possible to understand the behaviour by logic reasoning about the interactions without using GNA. Here follow two illustrative examples of the simulations and the behaviour observed by GNA for the genetic regulatory networks described by models I1 and model I4, respectively. The dynamics of the other models will be explained briefly.

5.1.1 Simulation of model I1

Model I1 was first simulated in the absence of light. The simulations were performed from all possible initial states, and all of them ended in one single equilibrium state associated with the regulatory domain

$\theta_{c1}^2 < x_{c1} \leq max_{c1}$
$0 < u_l < \theta_l$
$0 < x_m < \theta_m$
$\theta_{s1}^2 < x_m \leq max_{s1}$

The motility is low in this state, $x_m < \theta_m$, as expected when light is off and the model contains only light input. The underlying interactions giving this response, observed by GNA, are the following. Because of the absence of light, Cya1 is inactive and hence cAMP cannot be produced. The resulting low concentration of cAMP leads to a low concentration of the SYCRP1∼cAMP complex, that is, a concentration that is not high enough to induce motility. Also, because of the low complex concentration, there is no repression of the genes *cya1* and *sycrp1*, and therefore the concentrations of Cya1 and SYCRP1 becomes high, *i.e.* $\theta_{c1}^2 < x_{c1} \leq max_{c1}$ and $\theta_{s1}^2 < x_m \leq max_{s1}$, respectively.

Starting in the equilibrium state given above, and switching on light, will immediately give a high concentration of complex and thereby the motility will be turned on. The reasons for the increase in complex concentration is that there are already high concentrations of SYCRP1 and Cya1, and the synthesis of cAMP (which occurs when light activates Cya1) and the binding of cAMP to SYCRP1, are supposed to be almost instantaneous processes on the time-scale of gene expression. The concentration of Cya1 and SYCRP1 will then initially decrease, since the SYCRP1∼cAMP complex represses *cya1* and *sycrp1*. The decrease of these proteins then results in a lower concentration of the complex and thereby the repression stops, following which the concentration of these proteins goes up again. In this way these proteins will keep fluctuating around some equilibrium concentrations, analogously to the function of a thermostat. By GNA, the new equilibrium is somewhere in the following part of the phase space:

$\theta_{c1}^2 \leq x_{c1} \leq max_{c1}$
$\theta_l < u_l \leq max_l$
$\theta_m < x_m \leq max_m$
$\theta_{s1}^2 \leq x_{s1} \leq max_{s1}$

To sum up, both the genes *cya1* and *sycrp1* can be said to regulate themselves with negative feedback processes. That is, when the concentrations of the gene products are high

enough, the complex downregulates the genes (provided that light is present).

5.1.2 Simulation of model I4 and other models

Simulation of model I4 in the absence of light ends in a state with low motility and low concentrations of the proteins:

$$\theta_{c1}^1 < x_{c1} < \theta_{c1}^2$$
$$0 < u_l < \theta_l$$
$$0 < x_m < \theta_m$$
$$\theta_{s1}^1 < x_m < \theta_{s1}^2$$

Since the complex cannot be formed in the absence of light, the genes cannot be activated, which results in the low concentrations of the proteins. When light is turned on, the complex activates the genes and the concentration of the complex then increases. Motility is induced when the concentration of the complex exceeds its higher threshold. The new equilibrium state is associated with the regulatory domain:

$$\theta_{c1}^2 < x_{c1} \leq max_{c1}$$
$$\theta_l < u_l \leq max_l$$
$$\theta_m < x_m \leq max_m$$
$$\theta_{s1}^2 < x_m \leq max_{s1}$$

The models I2 and I3 behave as a mixture of models I1 and I4, with one protein concentration at high level and one at low level when they have reached their equilibrium states in darkness.

In models of class II, the light part of the network behaves in exactly the same way as the class I models. The behaviour of models from class II differ from that of class I models, by having low motility as soon as glucose is at a high level, no matter whether light is absent or not. When glucose is added in this model, Cya2 is activated, cGMP is synthesized and the SYCRP2∼cGMP complex is formed. As soon as this complex is present above its threshold, motility is inhibited, no matter what signal there is from the light part of the network. This is a result of the logic of the state equation for x_m which contains a logical 'and' rule, *i.e.* light has to be present AND glucose has to be absent in order for the motility to become high (see section 4.3).

The light part of the models in class III is also identical to the class I models. In models of class III, the dynamics in the glucose part of the network is like a mirror of the dynamics in the class I models. Also here, as in class II, the state equation for x_m is determined by an 'and' rule.

In models IV1 and IV2, the addition of glucose results in increased activity of Cya2, hence cGMP is synthesized and the SYCRP2∼cGMP complex is formed. The complex then inhibits the expression of *cya1* and *sycrp1* to such an extent that the SYCRP1∼cAMP decreases to a level low enough to abolish the motility.

5.2 Discriminatory experiments

By modifying a mathematical model it is possible to simulate how a mutant would behave if the model is correct. Such modified models will here be called *in silico* mutants in order to

distinguish them from mutants of real bacteria. Depending on the model, the corresponding *in silico* mutants behave differently when exposed to light and glucose. Therefore, by performing experiments with mutants, changing the light and glucose concentrations, it would be possible to distinguish between the models. Here will be explained how the *in silico* mutants behave when simulated in GNA, and by means of which experiments it would be possible to distinguish between some of the models. Note that these results rely on the assumptions made in section 4.1.

5.2.1 Distinguish between models I1, I2, I3, and I4

Firstly, we consider the models from class I, not containing any glucose influence. We will show how experiments with two mutants can be performed in order to distinguish between the four possible models in this class, I1-I4. The mutants should have the properties that one of them is defective in the binding sequence upstream of *cya1*, and the other one defective in the binding sequence upstream of *sycrp1*, so that the SYCRP1∼cAMP is unable to bind to any of the mutated binding sequences.

To change model I1 in order to mimic the mutant defective in the binding sequence upstream of *cya1* can be done by changing the state equation for Cya1 into

$$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} - \gamma_{c1} x_{c1}$$

As seen, the synthesis of Cya1 is then constantly at a high level, $\kappa^1_{c1} + \kappa^2_{c1}$. This describes that the complex cannot bind and is thereby unable to repress the expression.

In model I4 the same mutation is made by changing the state equation into

$$\dot{x}_{c1} = \kappa^1_{c1} - \gamma_{c1} x_{c1}$$

Here, the synthesis of the complex will be constantly at a low level, κ^1_{c1}, because when the complex cannot bind, gene expression will not be activated. Model I2 was modified in the same manner as model I4, and model I3 in the same way as model I1. The mutants defective in the binding sequence upstream of *sycrp1*, were constructed in a similar way as above. For details about the state equations in the mutants, see appendix B.

When simulating these models in the presence of light, they all showed different behaviours, see table 3. If model I1 is true, the motility will still be high in the presence of light in both mutants. If model I2 is true, the motility will be low in the former mutant and high in the latter, and the opposite would be true for model I3. If model I4 is right, both the mutants would be deficient in motility. These results show that performance of experiments with the mutants mentioned, would reveal the interactions in the genetic regulatory network controlling the phototactic response of *Synechocystis* PCC 6803, if the assumptions in section 4.1 are true.

Observe that even if there is a glucose influence on the motility, *i.e.* one of the models in class II-IV is true, this experiment is still meaningful in order to determine the light influence, (the glucose level has to be kept on a constant low level).

Model	*cya1*	*sycrp1*
I1	high	high
I2	low	high
I3	high	low
I4	low	low

Table 3: The table shows whether the motility is high or low in the presence of light in mutants defective in the binding sequences for SYCRP1∼cAMP upstream of *cya1* and *sycrp1*, respectively.

Model	light	light + glucose
II1	high	low
III1	high	low
IV1	high	high

Table 4: The table shows whether the motility is high or low in a mutant defective in the binding sequences for SYCRP1∼cAMP upstream of both *cya1* and *sycrp1* in the presence of only light and after addition of glucose.

5.2.2 Distinguish between models II1, III1, and IV1

How to determine the glucose influence will depend on the results from the experiment mentioned above. Here we will show how this could be done if the experiment above would show that model I1 is true.

Supposing that model I1 is true for the light influence, gives us three main possibilities for the glucose influence: model II1, III1 or IV1. With this second experiment it would be possible to discriminate between IV1 versus II1 or III1. That is, it would be possible to say whether the SYCRP2∼cGMP complex acts on motility directly or via the cAMP pathway (compare figures 10, 11 and 12).

In silico mutants of models II1, III1, and IV1, defective in the binding sequences for SYCRP1∼cAMP upstream of both *cya1* and *sycrp1* were constructed (for details about their state equations, see appendix B). Simulations showed that, in the presence of both light and glucose, the motility was high in the *in silico* mutant of model IV1, and low in the *in silico* mutants of models II1 and III1, see table 4. Therefore, if adding glucose in the presence of light to a mutant defective in the binding sequences upstream of both *cya1* and *sycrp1* would decrease the motility, model IV1 would be the right one. On the other hand, if the motility after addition of glucose would be the same as it was in presence of only light, this would infer that either model II1 or III1 is true.

6 Discussion

In this work, a modeling and simulation approach has been used to investigate the genetic regulatory network underlying chemotaxis and phototaxis in the cyanobacterium *Synechocystis* PCC 6803. Since these processes are not very well-studied, the information is sparse, and hence there are many different possible underlying networks, taking different interactions into account. Studying literature and making assumptions has enabled the construction of four classes of plausible networks describing the phototaxis and chemotaxis in *Synechocystis* PCC 6803. These networks have been modeled and simulated by a qualitative modeling and simulation method based on PLDEs. The advantage with this method is that it does not require any quantitative information, which is a necessary property in our case since we have almost no such information about the system under study. Instead the method employs constrains in the form of algebraic inequalities that can be obtained from the literature. The aim has been to investigate the behaviour of the possible networks, and to propose experiments that enable discrimination between networks. The latter can be done by comparing the outcome from the experiments with the outcome from simulations of a network, and if the results do not agree, the network can be rejected. Also we wanted to know whether the qualitative modeling and simulation method is useful when analyzing a genetic regulatory network about which so little is known as the one in *Synechocystis* PCC 6803.

We have shown that it is possible to reject some networks by performing certain experiments, if the assumptions that were made when constructing the models are true. In model class I, which contains only models with light input and no glucose input, it is possible to select one of the four models by performing experiments with two mutants defective in the genes *cya1* and *sycrp1*, respectively. The experiment would give four different results depending on which one of the four possible networks is true. After selecting a model with only the light effect, it remains to determine the glucose influence. We have shown that it is possible to discriminate between two groups of models including both light and glucose input. This can be done by investigating a mutant defective in both *cya1* and *sycrp1*. The experiment would give two different results depending on which of the two groups the true model belongs to.

Because of the little information available on the system, it is of interest to perform sequence analysis in order to get some more information on the network. We have already started work on two sequence analysis tasks and this will be continued by people at INRIA. The first issue is to check whether the SYCRP2 protein is cGMP-binding, since in the model classes II-IV it is assumed that SYCRP2 binds cGMP and in model classes III-IV, it is also assumed that the SYCRP2$\sim$cGMP complex is able to bind DNA and therefore might regulate genes. The idea is to test whether the protein would be classified as a cAMP- or a cGMP-binding protein by a learning system trained with cAMP- and cGMP-binding proteins respectively. Since we know from experiments that SYCRP2 is not cAMP-binding, it is reasonable to assume that SYCRP2 does not bind cGMP either, if it would be classified as cAMP-binding. If SYCRP2 would fit into neither the cAMP-binding nor the cGMP-binding class, the same conclusion would be plausible. On the contrary, if it would show much higher similarity with the proteins in the cGMP-binding class than the ones in the cAMP-binding class, the underlying assumption that SYCRP2 is cGMP-binding is plausible. What has been done so far is a database search on the Internet for proteins that are known to bind cAMP and cGMP respectively. Also, Eric Coissac at INRIA has started the work on training a learning

system.

The second sequence analysis issue we have started to work on is the investigation of whether SYCRP1∼cAMP might regulate other genes than *sycrp1* and *cya1* in the genetic regulatory network under study. That is, we want to know whether there are sequences upstream of genes in the network that the complex can bind to, since this would indicate that the gene might be regulated by the complex. Especially we are interested in binding sites upstream of genes in the networks that we have investigated. Since SYCRP1∼cAMP is known to bind to the *E. coli* CRP consensus sequence, DNA sequences that *E. coli* CRP is shown to bind to can be used to make a learning system that can be used to find similar sequences. Therefore, DNA sequences that *E. coli* CRP binds to have been collected in a database. Additional information, such as nucleotide sequences up- and downstream of the binding sequence, and information about whether the gene downstream of the binding sequence is activated or inhibited is also stored in the database. This data will be used to construct a learning system in order to search for sequences in the *Synechocystis* PCC 6803 genome that the SYCRP1∼cAMP complex might bind to.

As mentioned, we were interested in finding out how much needs to be known about a system in order for GNA to be useful for its study. There are two factors that play a role for the usefulness of GNA when studying a network about which there is not much information. Firstly, if the knowledge about the connections between the components in the network is deficient, the number of possible models becomes exceedingly large, and hence using GNA would be an inefficient way of investigating the system. In these cases, it is sometimes possible to make several models out of the same network structure, by putting the inequalities in different ways. This might lead to difficulties in finding discrimination experiments, since the different models of the same network have different dynamical behaviour. Therefore, it is in this case probably of more use to first apply conventional experimental methods in order to decrease the number of alternative networks before applying GNA to the network. Secondly, when the size and complexity of the network is relatively low, it is possible to understand the behaviour of the system without using GNA.

Because of the incomplete information about the motility network, we had a big number of alternative networks despite several assumptions that were made about the system. In addition, we also faced the second difficulty, dealing with small networks. In most models of the motility module we had 4 genes, 2 input parameters and 1 output, so in total they had 7 variables. The models with only light input had 2 genes, so these had in total 4 variables. This number is quite low as compared to *e.g.* the 11 variables in the sporulation network, modeled by the same method [9]. In addition to few variables, there are many alternative networks because of the incomplete information, although many assumptions have been made.

If the underlying assumptions about the motility module (in section 4.1) would turn out to be wrong, this work or parts of it would not be useful. The information about the connections between the variables is thus in this case a bit below the amount needed for a useful study. Especially, there is one crucial fact which is of particular importance, and that is the information about how the output depend on the input. Since the models are designed in order to give the assumed output for a given input, all models would be wrong if this assumption would be wrong. In our case, the motility dependence on the light and glucose is not exhaustively studied.

If the underlying assumptions in section 4.1 already had been verified, there are still a huge amount of possible models. However, in spite of this big amount of models, it has been

possible to suggest a couple of experiments in order to distinguish between them. Thus, if the assumptions would be right, the amount of information about connections would have been just above the limit of the lowest amount of information needed.

The question is then whether the components involved in the network are so few that it would be possible to understand the network behaviour without GNA. The behaviour of most of the networks of the motility module could be understood by logic reasoning, without using GNA. On the other hand, GNA enables one to run a huge amount of models and different *in silico* mutants quickly and see how they behave, a process that would be more time-consuming and error-prone if doing it without GNA. We have investigated some networks with only two genes, one input and one output, which is probably the lowest amount of variables of a network which one might be interested in, since a 'network' with only one gene would not contain any interesting interactions. This shows that there seems to be no lowest number of components in order for GNA to be helpful, since even for the analysis of networks of the smallest possible size, GNA has been useful (but of course the usefulness becomes more obvious for larger networks).

In conclusion, for GNA to be useful, the number of variables involved seems to be less important than the fact that there is sufficient information about the connections in the network, so that no (or at least just a few well-supported) assumptions underlying all of the models, such as those in section 4.1, have to be made. If too many assumptions of this kind have to be made, it is probably more efficient to start with experiments in order to test these assumptions before applying GNA to the network.

Acknowledgements

I would like to thank my supervisor Hidde de Jong for enabling me to work in his group at INRIA Rhône-Alpes with this interesting project. Thanks also to Jean Houmard for his contribution to this project by providing information on the biological system and for proofreading the report. Further, I would like to thank Delphine Ropers for her contribution to the main part of section 3 of this report, and for patiently spending time answering my questions. Grégory Batt has also spent much time on patiently answering questions and helping me with computer difficulties, which I am tremendously thankful for. Finally, I would like to thank everybody in the GNA team for bringing about a nice atmosphere.

References

[1] D. Bhaya, N.R Bianco, D. Bryant, and A. Grossman. Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp. PCC6803. *Mol. Microbiol.*, 37(4):941–51, 2000.

[2] D. Bhaya, A. Takahashi, P. Shahi, and A.R. Grossman. Novel motility mutants of *Synechocystis* strain PCC 6803 generated by in vitro transposon mutagenesis. *J. Bacteriol.*, 183(20):6140–3, 2001.

[3] N.A. Campbell, J.B. Reece, and L.G. Mitchell. *Biology.* J. Green, fifth edition, 1999.

[4] M.J. Cann. Signalling through cyclic nucleotide monophosphates in cyanobacteria. *New Phytol.*, 161:23–34, 2003.

[5] R. Casey, H. de Jong, and J.-L. Gouzé. Piecewise-linear models of genetic regulatory networks: Equilibria and their stability. Technical Report RR-5353, INRIA, 2004.

[6] CyanoBase. The genome database for cyanobacteria, Dec 2004. http://www.kazusa.or.jp/cyano/.

[7] J.A.G. Ochoa de Alda, G. Ajlani, and J. Houmard. *Synechocystis* strain PCC 6803 *cya2*, a procaryotic gene that encodes a guanylyl cyclase. *J. Bacteriol.*, 182(13):3839–42, 2000.

[8] J.A.G. Ochoa de Alda and J. Houmard. Genomic survey of cAMP and cGMP signalling components in the cyanobacterium *Synechocystis* PCC 6803. *Microbiology*, 146(12):3183–94, 2000.

[9] H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in *Bacillus subtilis*. *Bull. Math. Biol.*, 66(2):261–99, 2004.

[10] H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic Network Analyzer: qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19(3):336–44, 2003.

[11] H. de Jong, J-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull. Math. Biol.*, 66(2):301–40, 2004.

[12] A.F. Filippov. *Differential equations with discontinuous righthand sides.* Kluwer Academic Publishers, Dordrecht, 1988.

[13] L. Glass and S.A. Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.*, 39(1):103–29, 1973.

[14] J.-L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dynam. Syst.*, 17:299–316, 2003.

[15] A. Hanamura and H. Aiba. Molecular mechanism of negative autoregulation of *Escherichia coli crp* gene. *Nucleic Acids Res.*, 19(16):4413–19, 1991.

[16] M. Herdman and K. Elmorjani. Cyclic nucleotides. *Methods Enzymol.*, 167:584–91, 1988.

[17] T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, and S. Tabata. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and of potential protein-coding regions. *DNA Res.*, 3(3):109–36, 1996.

[18] K.W. Kohn. Molecular interaction maps as information organizers and simulation guides. *Chaos*, 11(1):84–97, 2001.

[19] T. Mestl, E. Plahte, and S.W. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.*, 176(2):291–300, 1995.

[20] W.-O. Ng, A.R. Grossman, and D. Bhaya. Multiple light inputs control phototaxis in *Synechocystis* sp. strain PCC6803. *J. Bacteriol.*, 185(5):1599–607, 2003.

[21] K. Okamoto and M. Freundlich. Mechanism for the autogenous control of the *crp* operon: Transcriptional inhibition by a divergent RNA transcript. *Proc. Natl. Acad. Sci. USA*, 83(14):5000–4, 1986.

[22] K. Omagari, H. Yoshimura, M. Takano, D. Hao, M.Ohmori, A. Sarai, and A. Suyama. Systematic single base-pair substitution analysis of DNA binding by the cAMP receptor protein in cyanobacterium *Synechocystis* sp. PCC 6803. *FEBS Lett.*, 563(1-3):55–8, 2004.

[23] E. Plahte, T. Mestl, and S.W. Omholt. A methodological basis for description and analysis of systems with complex switch-like interactions. *J. Math. Biol.*, 36(4):321–48, 1998.

[24] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann. Qualitative simulation of the nutritional stress response in *Escherichia coli*. Technical Report RR-5412, INRIA, 2004.

[25] E.H. Snoussi. Qualitative dynamics of piece-linear differential equations: a discrete mapping approach. *Dyn. Stability Syst.*, 4(3-4):189–207, 1989.

[26] K. Terauchi and M. Ohmori. An adenylate cyclase, Cya1, regulates cell motility in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant. Cell. Pysiol.*, 40(2):248–51, 1999.

[27] K. Terauchi and M. Ohmori. Blue light stimulates cyanobacterial motility via a cAMP signal transduction system. *Mol. Microbiol.*, 52(1):303–9, 2004.

[28] A.U. Viretta and M. Fussenegger. Modeling the quorum sensing regulatory network of human-pathogenic *Pseudomonas aeruginosa*. *Biotechnol. Prog.*, 20(3):670–8, 2004.

[29] H. Yoshimura, T. Hisabori, S. Yanagisawa, and M. Ohmori. Identification and caracterization of a novel cAMP receptor protein in the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Biol. Chem.*, 275(9):6241–5, 2000.

[30] H. Yoshimura, S. Yanagisawa, M. Kanehisa, and M. Ohmori. Screening for the target gene of cyanobacterial cAMP receptor protein SYCRP1. *Mol. Microbiol.*, 43(4):843–53, 2002.

[31] H. Yoshimura, S. Yoshihara, S. Okamoto, M. Ikeuchi, and M. Ohmori. A cAMP receptor protein, SYCRP1, is responsible for the cell motility of *Synechocystis* sp. PCC 6803. *Plant. Cell. Physiol.*, 43(4):460–3, 2002.

Appendices

A Available data on interactions in the regulation module

The regulation module, which is the network we investigate in this study (see section 2.1 and figure 2), is constituted by a number of components (more precisely: light, glucose, *cya1*, *cya2*, Cya1, Cya2, *pdeB*, *pdeA*, PdeA, PdeB, *sycrp1*, *sycrp2*, SYCRP1, SYCRP2, cAMP and cGMP), and the interactions between these. In this appendix we will try to organize the information about the interactions in the regulation module. We do this by considering the influence that every component in the network might have on every other component in the network. In order to make an overview of these possible influences we have summarized them in table 5, where we represent the possible influences as either activation, inhibition or no interaction. Often there is not enough knowledge to tell for sure what kind of influence there is, in these cases the possible influences are ordered in the table according to how likely they are. To specify the knowledge about the interactions more exactly, all information in the table has references to comments in the text. In the comments, a detailed description of what is known about the influence is given and it is explained what experiments or arguments the knowledge about these influences is based on. All information in this appendix comes from three kinds of sources: (1) the literature, (2) interrogations and discussions with the biologist and cyanobacteria specialist, Jean Houmard, researcher at Ecole Normale Supérieure (ENS) in Paris, and (3) the CyanoBase Internet site [6].

Since it would be non-sense to talk about the influence that a component would have on light, light has been taken away among 'the components that are influenced' in table 5. Similarly, we took away glucose among 'the components that are influenced', since we regard the glucose level as an input of the system. Also, since the genes in the model do not influence any other component except from via their proteins, we have taken away all genes among 'the components that exercise an influence on other components'. Further, instead of SYCRP1 and SYCRP2, we include the two complexes SYCRP1∼cAMP and SYCRP2∼cGMP in the table, since the former ones probably do not interact with the other components in the model except when forming complexes with cyclic nucleotides. Additionally, since the interactions between the cyclic nucleotides and the other components in the network are described in section 2, the cyclic nucleotides are not included in the table.

Comments about the influences in the regulation module

Here are the comments to the information given in table 5.

1. It is known that the cAMP level in wild type increases in response to blue light, while the cAMP level in a *cya1* mutant is insensitive to blue light [27]. These results show that either the transcription of *cya1* or the activity of Cya1 or both must be activated by light. Since Cya1 does not have any photoreceptor domain and since the expression of a gene cannot be directly activated by light, this influence must be indirect. Some protein able to receive photons, might activate Cya1 or the expression of *cya1*, or there might be a chain of steps between the light signal and the activation of Cya1/*cya1*. These reaction steps are by our definition included in the signal transduction module (see figure 2).

	Light	Glucose	Cya1	Cya2	PdeB	PdeA	SYCRP1∼cAMP	SYCRP2∼cGMP
cya1	+ 0 [1]	0 [8]	0 [9]	0 [10]	0 [11]	0 [11]	+/- (0) [16]	0 +/- [23]
cya2	0 + (-) [2]	0 + [7]	0 [9]	0 [10]	0 [11]	0 [11]	0 +/- [17]	0 + - [24]
Cya1	+ 0 [1]	0 [7]	N	0 [10]	0 [12]	0 [12]	0 [18]	0 [18]
Cya2	0 + (-) [2]	0 +[7]	0 [9]	N	0 [13]	0 [13]	0 [18]	0 [18]
pdeB	- 0 (+) [3]	0 [7]	0 [9]	0 [10]	0 [11]	0 [11]	0 + (-) [17]	0 + - [24]
pdeA	- 0 (+) [4]	0 - [7]	0 [9]	0 [10]	0 [11]	0 [11]	0 + (-) [17]	0 + - [24]
PdeB	0 (-) (+) [3]	0 [7]	0 [9]	0 [10]	N	0 [14]	0 [19]	0 [19]
PdeA	0 (-) (+) [4]	0 - [7]	0 [9]	0 [10]	0 [14]	N	0 [19]	0 [19]
sycrp1	0 (+) (-) [5]	0 [8]	0 [9]	0 [10]	0 [11]	0 [11]	- + (0) [20]	0 (-) (+) [21]
sycrp2	0 (+) (-) [5]	0 [8]	0 [9]	0 [10]	0 [11]	0 [11]	0 - + [21]	0 (-) (+) [21]
SYCRP1∼cAMP	0 [6]	0 [8]	0 [9]	0 [10]	0 [15]	0 [15]	N	0 [22]
SYCRP2∼cGMP	0 [6]	0 [8]	0 [9]	0 [10]	0 [15]	0 [15]	0 [22]	N

Table 5: Influences between components in the regulation module. The top row contains the components that exercise an influence on the components in the leftmost column. So in each square, the possible influences that the component straight above has on the component straight to the left are shown. The influences are abbreviated as follows: + stands for activation, - for inhibition, 0 for no interaction and N for non-sense influence. When it is indicated that a gene is activated or inhibited, this means that the transcription of it is activated or inhibited respectively. When the same is indicated about a protein, this refers to the activity of the protein. In cases when several different hypotheses about the influence are conceivable, these are ordered with the most probable hypothesis first. When the symbol for the influence is written in parentheses, this indicates that the influence is very unlikely but not possible to exclude completely. The numbers in parentheses refer to comments in the text.

2. No experiments have been done yet on cGMP level change in a *cya2* mutant upon a light shift, so it is not known whether *cya2* transcription or Cya2 activity is influenced by light, but these possibilities should not be excluded. Because of the similarity between *cya1* and *cya2*, and between Cya1 and Cya2, there is reason to believe that Cya2 or *cya2* or both are activated by light in the same way as in comment (1) with possibly several steps between light and activation. Therefore, activation has been ordered as more probable than inactivation. Nothing indicates that inhibition is present but this possibility cannot be completely excluded.

3. It is shown that the cAMP level increases in response to light [27], and since phosphodiesterases are degrading cAMP, this suggests that *pdeB* or PdeB might be inhibited by light. However, activation cannot be excluded, since it is not impossible that the main effect of the light influence on the cAMP level is done by other genes or proteins such as *cya1* or Cya1. If there is an influence by light on either PdeB or *pdeB*, it is more likely on the gene than on the protein level. Since PdeB does not have any photoreceptor domain, and since the expression of a gene cannot be directly activated by light, the influence (if it exists) must be indirect. Some protein able to receive photons, might activate PdeB or the expression of *pdeB*, or there might be a chain of steps between the light signal and the activation of PdeB/*pdeB*. These reaction steps are by our definition included in the signal transduction module (see figure 2).

4. Because of the similarity between PdeA and PdeB and between *pdeA* and *pdeB* there is reason to believe that PdeA and *pdeA* are influenced by light in the same way as PdeB and *pdeB*, see comment (3).

5. There is no information about whether light has any influence on the genes *sycrp1* and *sycrp2*. If there would be an influence it is reasonable to think that it is activation rather than inhibition, since it is known that the cAMP level increases and motility is induced in response to light [27], and since the SYCRP1∼cAMP complex is known to be needed for this motility [31].

6. This complex is already in activated form and to inactivate it, the complex must dissociate into cyclic nucleotide receptor protein and cyclic nucleotide. There is no indication that this dissociation is influenced by light.

7. The fact that indicates that glucose might be involved in motility is, as mentioned, that it causes *Synechocystis* PCC 6803 to aggregate in the presence of glucose, which suggests that the motility is downregulated. It is shown that in cells growing under light and with glucose, the cGMP level is doubled as compared to cells growing under light alone [69]. Therefore, *cya2* or Cya2 might be activated by glucose, or *pdeA* or PdeA might be inhibited by glucose or any combination of these influences.

8. Since there is no reason to believe that these genes or proteins are influenced by glucose, we regard these possibilities enough unlikely to neglect them in our analysis. (Compare with comment (7).)

9. There is nothing that indicates that Cya1 would regulate any genes or proteins.

10. Cya2 is most likely a transmembrane protein [7] and therefore it does not regulate any genes or proteins.

11. PdeB and PdeA do not regulate any genes because they lack DNA-binding domains.

12. No indications of influence from PdeA or PdeB on Cya1.

13. Interaction between these proteins is impossible since they do not contain domains that can bind to each other.

14. PdeA and PdeB do not bind to each other and can therefore not exercise any influence on each other.

15. PdeA and PdeB bind neither SYCRP1∼cAMP nor SYCRP2∼cGMP and do therefore not exercise any influence on any of these complexes.

16. There exists a DNA sequence 314 nucleotides upstream from the coding region of *cya1* that is similar to the consensus sequence for CRP in *E. coli* [8]. This site upstream of *cya1* has, in the right position and with the appropriate spacing, the most crucial residues for the binding of SYCRP1∼cAMP to DNA [22]. All the eighteen targets for the SYCRP1∼cAMP complex that have been identified so far are upregulated by the complex [30], which suggests that the complex more likely would function as an activator than a repressor also in this case. On the other hand, in *E. coli*, more than 80% of the over hundred genes that are known to be regulated by CRP∼cAMP, are activated, and if we have a similar situation in *Synechocystis* PCC 6803, it could be a coincidence that only activations have been found so far. We also know that the CRP∼cAMP complex in *E. coli* is inhibiting the gene encoding the adenylyl cyclase *cya* in *E. coli* and, in fact, the complex is both activating and inhibiting *crp* in *E. coli*. Taken together, these facts indicate that activation and inhibition are equally probable for SYCRP1∼cAMP.

speices	CRP binding site	CRP binding sequence	sim cons	sim *E. coli*
S. 6803	*sycrp1* site I	GCAAGTGAAAAAACTCACTTTC	12/22	12/22
E. coli	*crp* site I	GTATGCAAAGGACGTCACATTA	12/12	
S. 6803	*sycrp1* site II	GACTGTGATAAATGTCTTTCAA	10/22	10/22
E. coli	*crp* site II	GAAGGCGACCTGGGTCATGCTG	12/12	
E. coli	CRP consensus	AAATGTGATCTAGATCACATTT		

Table 6: The similarity between sequences compared pairwise is given as the ratio between the number of positions with identical bases and the total number of positions. The abbreviation 'sim cons' stands for the similarity to the CRP consensus sequence in *E. coli*. The column 'sim *E. coli*' contains the similarities to the corresponding CRP binding site in *E. coli*, *i.e.* site I in *Synechocystis* PCC 6803 is compared to site I in *E. coli* and site II in *Synechocystis* PCC 6803 is compared to site II in *E. coli*. *S.* 6803 is an abbreviation for *Synechocystis* PCC 6803. In *E. coli*, site I is the downstream repressing site and site II is the upstream activating site.

17. Might work analogous to SYCRP1∼cAMP regulation of *cya1*, see comment (16), but this has not been investigated yet.

18. Interaction can be excluded, since the proteins do not bind to each other.

19. Influence from the cyclic nucleotide receptor proteins on the phosphodiesterases is very unlikely since the proteins do not contain domains that likely can bind to each other.

20. There are two sequences upstream from the coding region of *sycrp1* that are homologous to the consensus sequence for CRP in *E. coli*. These sequences also show similarity to the two binding sites of CRP upstream of the gene *crp* in *E. coli* [8]. In *E. coli*, the CRP∼cAMP complex is activating the gene *crp* by binding to the upstream site, and repressing the gene by binding to the downstream site. Since the similarity between the downstream elements of *crp* and *sycrp1* is higher than the similarity between the upstream sites (see table 6), it is slightly more likely that *sycrp1* is downregulated than upregulated by the SYCRP1∼cAMP complex.

21. Since we do not have any information about whether there is a binding site for the complex the first hypothesis is that there is no regulation. Negative regulation has been ordered as more probable than positive regulation by analogy with the SYCRP1∼cAMP regulation of *sycrp1* (see comment (20)). However, it is not known whether SYCRP2∼cGMP binds to the *E. coli* consensus sequence, just like SYCRP1∼cAMP. Also, it is not known whether there is an *E. coli* CRP binding site upstream of *sycrp2*.

22. There is nothing that indicates that these two complexes would interact.

23. When wild type *Synechocystis* is exposed to UV-B radiation, the concentration of *cya1* mRNAs decreases to half of the normal level after 30 min, while the concentration of *sycrp2* mRNAs does not change. Also, the concentration of cGMP dropped to about half of the normal level after 3-10 min under UV-B. After 30 minutes under UV-B, the cGMP level has returned to its normal level. The cAMP level did not change under the UV-B treatment. These facts (given by Jean Houmard, biologist at ENS in Paris)

shows that when the cGMP level goes down, the expression of *cya1* also decreases, which suggests that *cya1* might be activated by the SYCRP2∼cGMP complex. On the other hand, the expression of *cya1* also decreases by 50% in a *pdeA* mutant which has a higher cGMP concentration than the wild type (these facts also given by Jean Houmard). This suggests the contrary: that the SYCRP2∼cGMP complex is inhibiting *cya1*.

24. There might be an influence because of the homology between SYCRP1 and SYCRP2 and since SYCRP1∼cAMP is regulating *sycrp1*, see comment (16). However, in this case, it has not been shown whether SYCRP2∼cGMP binds the consensus sequence, and we do not know whether this complex is usually upregulating its targets like SYCRP1∼cAMP does.

Type 1 mutant:	Defect in the binding site upstream of *cya1*, precluding binding of SYCRP1∼cAMP and SYCRP2∼cGMP to this binding site.
Type 2 mutant:	Defect in the binding site upstream of *sycrp1*, precluding binding of SYCRP1∼cAMP and SYCRP2∼cGMP to this binding site.
Type 3 mutant:	Defect in the binding sites upstream of both *cya1* and *sycrp1*, precluding binding of SYCRP1∼cAMP and SYCRP2∼cGMP to these binding sites.

Table 7: The kind of defects in the three types of mutants that are modeled.

B Possible PLDE models for the regulation module

In section 4 it is explained how the mathematical models are derived. In the following tables of models (table 8 to 15), complete information about all the models in that section is shown. For each model, all state equations and threshold and equilibrium inequalities are given. Following the complete model, information about the *in silico* mutants derived from the model is given. The *in silico* mutants have been made by doing small modifications in the models, as explained in section 5.2. Only the state equations that differ between the *in silico* mutant and its original model are shown in the tables. All the inequality constraints for the mutant of a model are the same as in its original model. The defects in real mutants that are modeled are listed in table 7, and in the tables of models only the number of the type of mutant is given.

$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1}(1 - s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)) - \gamma_{c1}x_{c1}$ $0 < \theta^1_{c1} < \theta^2_{c1} < max_{c1}$ $\theta^1_{c1} < \kappa^1_{c1}/\gamma_{c1} < \theta^2_{c1},\ \theta^2_{c1} < (\kappa^1_{c1} + \kappa^2_{c1})/\gamma_{c1} < max_{c1}$
$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1}(1 - s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)) - \gamma_{s1}x_{s1}$ $0 < \theta^1_{s1} < \theta^2_{c1} < max_{s1}$ $\theta^1_{s1} < \kappa^1_{s1}/\gamma_{s1} < \theta^2_{s1},\ \theta^2_{s1} < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < max_{s1}$
$\dot{x}_m = \kappa_m s^+(x_{s1}, \theta^1_{s1})s^+(x_{c1}, \theta^1_{c1})s^+(u_l, \theta_l) - \gamma_m x_m$ $0 < \theta_m < max_m$ $\theta_m < \kappa_m/\gamma_m < max_m$
$\dot{u}_l = 0$ $0 < \theta_l < max_l$
Type 1 mutant: $\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} - \gamma_{c1}x_{c1}$
Type 2 mutant: $\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1} - \gamma_{s1}x_{s1}$
Type 3 mutant: $\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} - \gamma_{c1}x_{c1}$ $\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1} - \gamma_{s1}x_{s1}$

Table 8: Model I1

$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} s^+(x_{s1}, \theta^1_{s1}) s^+(x_{c1}, \theta^1_{c1}) s^+(u_l, \theta_l)) - \gamma_{c1} x_{c1}$ $0 < \theta^1_{c1} < \theta^2_{c1} < \theta^3_{c1} < max_{c1}$ $\theta^1_{c1} < \kappa^1_{c1}/\gamma_{c1} < \theta^2_{c1},\ \theta^3_{c1} < (\kappa^1_{c1} + \kappa^2_{c1})/\gamma_{c1} < max_{c1}$
$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1}(1 - s^+(x_{s1}, \theta^3_{s1}) s^+(x_{c1}, \theta^3_{c1}) s^+(u_l, \theta_l)) - \gamma_{s1} x_{s1}$ $0 < \theta^1_{s1} < \theta^2_{c1} < \theta^3_{c1} < max_{s1}$ $\theta^1_{s1} < \kappa^1_{s1}/\gamma_{s1} < \theta^2_{s1},\ \theta^3_{s1} < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < max_{s1}$
$\dot{x}_m = \kappa_m s^+(x_{s1}, \theta^2_{s1}) s^+(x_{c1}, \theta^2_{c1}) s^+(u_l, \theta_l) - \gamma_m x_m$ $0 < \theta_m < max_m$ $\theta_m < \kappa_m/\gamma_m < max_m$
$\dot{u}_l = 0$ $0 < \theta_l < max_l$
Type 1 mutant: $\dot{x}_{c1} = \kappa^1_{c1} - \gamma_{c1} x_{c1}$
Type 2 mutant: $\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1} - \gamma_{s1} x_{s1}$

Table 9: Model I2

$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1}(1 - s^+(x_{s1}, \theta^3_{s1})s^+(x_{c1}, \theta^3_{c1})s^+(u_l, \theta_l)) - \gamma_{c1}x_{c1}$ $0 < \theta^1_{c1} < \theta^2_{c1} < \theta^3_{c1} < max_{c1}$ $\theta^1_{c1} < \kappa^1_{c1}/\gamma_{c1} < \theta^2_{c1},\ \theta^3_{c1} < (\kappa^1_{c1} + \kappa^2_{c1})/\gamma_{c1} < max_{c1}$
$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1}s^+(x_{s1}, \theta^1_{s1})s^+(x_{c1}, \theta^1_{c1})s^+(u_l, \theta_l)) - \gamma_{s1}x_{s1}$ $0 < \theta^1_{s1} < \theta^2_{c1} < \theta^3_{c1} < max_{s1}$ $\theta^1_{s1} < \kappa^1_{s1}/\gamma_{s1} < \theta^2_{s1},\ \theta^3_{s1} < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < max_{s1}$
$\dot{x}_m = \kappa_m s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l) - \gamma_m x_m$ $0 < \theta_m < max_m$ $\theta_m < \kappa_m/\gamma_m < max_m$
$\dot{u}_l = 0$ $0 < \theta_l < max_l$
Type 1 mutant: $\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} - \gamma_{c1}x_{c1}$
Type 2 mutant: $\dot{x}_{s1} = \kappa^1_{s1} - \gamma_{s1}x_{s1}$

Table 10: Model I3

$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} s^+(x_{s1}, \theta^1_{s1}) s^+(x_{c1}, \theta^1_{c1}) s^+(u_l, \theta_l)) - \gamma_{c1} x_{c1}$ $0 < \theta^1_{c1} < \theta^2_{c1} < max_{c1}$ $\theta^1_{c1} < \kappa^1_{c1}/\gamma_{c1} < \theta^2_{c1},\ \theta^2_{c1} < (\kappa^1_{c1} + \kappa^2_{c1})/\gamma_{c1} < max_{c1}$
$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1} s^+(x_{s1}, \theta^1_{s1}) s^+(x_{c1}, \theta^1_{c1}) s^+(u_l, \theta_l)) - \gamma_{s1} x_{s1}$ $0 < \theta^1_{s1} < \theta^2_{c1} < max_{s1}$ $\theta^1_{s1} < \kappa^1_{s1}/\gamma_{s1} < \theta^2_{s1},\ \theta^2_{s1} < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < max_{s1}$
$\dot{x}_m = \kappa_m s^+(x_{s1}, \theta^2_{s1}) s^+(x_{c1}, \theta^2_{c1}) s^+(u_l, \theta_l) - \gamma_m x_m$ $0 < \theta_m < max_m$ $\theta_m < \kappa_m/\gamma_m < max_m$
$\dot{u}_l = 0$ $0 < \theta_l < max_l$
Type 1 mutant: $\dot{x}_{c1} = \kappa^1_{c1} - \gamma_{c1} x_{c1}$
Type 2 mutant: $\dot{x}_{s1} = \kappa^1_{s1} - \gamma_{s1} x_{s1}$

Table 11: Model I4

$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1}(1 - s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)) - \gamma_{c1}x_{c1}$ $0 < \theta^1_{c1} < \theta^2_{c1} < max_{c1}$ $\theta^1_{c1} < \kappa^1_{c1}/\gamma_{c1} < \theta^2_{c1},\ \theta^2_{c1} < (\kappa^1_{c1} + \kappa^2_{c1})/\gamma_{c1} < max_{c1}$
$\dot{x}_{c2} = \kappa_{c2} - \gamma_{c2}x_{c2}$ $0 < \theta_{c2} < max_{c2}$ $\theta_{c2} < \kappa_m/\gamma_m < max_{c2}$
$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1}(1 - s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)) - \gamma_{s1}x_{s1}$ $0 < \theta^1_{s1} < \theta^2_{c1} < max_{s1}$ $\theta^1_{s1} < \kappa^1_{s1}/\gamma_{s1} < \theta^2_{s1},\ \theta^2_{s1} < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < max_{s1}$
$\dot{x}_{s2} = \kappa_{s2} - \gamma_{s2}x_{s2}$ $0 < \theta_{s2} < max_{s2}$ $\theta_{s2} < \kappa_m/\gamma_m < max_{s2}$
$\dot{x}_m = \kappa_m s^+(x_{s1}, \theta^1_{s1})s^+(x_{c1}, \theta^1_{c1})s^+(u_l, \theta_l)(1 - s^+(x_{c2}, \theta_{c2})s^+(x_{s2}, \theta_{s2})s^+(u_g, \theta_g)) - \gamma_m x_m$ $0 < \theta_m < max_m$ $\theta_m < \kappa_m/\gamma_m < max_m$
$\dot{u}_l = 0$ $0 < \theta_l < max_l$
$\dot{u}_g = 0$ $0 < \theta_g < max_g$
Type 3 mutant: $\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} - \gamma_{c1}x_{c1}$ $\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1} - \gamma_{s1}x_{s1}$

Table 12: Model II1

$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1}(1 - s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)) - \gamma_{c1}x_{c1}$ $0 < \theta^1_{c1} < \theta^2_{c1} < max_{c1}$ $\theta^1_{c1} < \kappa^1_{c1}/\gamma_{c1} < \theta^2_{c1},\ \theta^2_{c1} < (\kappa^1_{c1} + \kappa^2_{c1})/\gamma_{c1} < max_{c1}$
$\dot{x}_{c2} = \kappa^1_{c2} + \kappa^2_{c2}(1 - s^+(x_{s2}, \theta^2_{s2})s^+(x_{c2}, \theta^2_{c2})s^+(u_g, \theta_g)) - \gamma_{c2}x_{c2}$ $0 < \theta^1_{c2} < \theta^2_{c2} < max_{c2}$ $\theta^1_{c2} < \kappa^1_{c2}/\gamma_{c2} < \theta^2_{c2},\ \theta^2_{c2} < (\kappa^1_{c2} + \kappa^2_{c2})/\gamma_{c2} < max_{c2}$
$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1}(1 - s^+(x_{s1}, \theta^2_{s1})s^+(x_{c1}, \theta^2_{c1})s^+(u_l, \theta_l)) - \gamma_{s1}x_{s1}$ $0 < \theta^1_{s1} < \theta^2_{s1} < max_{s1}$ $\theta^1_{s1} < \kappa^1_{s1}/\gamma_{s1} < \theta^2_{s1},\ \theta^2_{s1} < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < max_{s1}$
$\dot{x}_{s2} = \kappa^1_{s2} + \kappa^2_{s2}(1 - s^+(x_{s2}, \theta^2_{s2})s^+(x_{c2}, \theta^2_{c2})s^+(u_g, \theta_g)) - \gamma_{s2}x_{s2}$ $0 < \theta^1_{s2} < \theta^2_{s2} < max_{s2}$ $\theta^1_{s2} < \kappa^1_{s2}/\gamma_{s2} < \theta^2_{s2},\ \theta^2_{s2} < (\kappa^1_{s2} + \kappa^2_{s2})/\gamma_{s2} < max_{s2}$
$\dot{x}_m = \kappa_m s^+(x_{s1}, \theta^1_{s1})s^+(x_{c1}, \theta^1_{c1})s^+(u_l, \theta_l)(1 - s^+(x_{c2}, \theta^1_{c2})s^+(x_{s2}, \theta^1_{s2})s^+(u_g, \theta_g)) - \gamma_m x_m$ $0 < \theta_m < max_m$ $\theta_m < \kappa_m/\gamma_m < max_m$
$\dot{u}_l = 0$ $0 < \theta_l < max_l$
$\dot{u}_g = 0$ $0 < \theta_g < max_g$
Type 3 mutant: $\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} - \gamma_{c1}x_{c1}$ $\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1} - \gamma_{s1}x_{s1}$

Table 13: Model III1

$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1}(1 - s^+(x_{c1}, \theta^2_{c1})s^+(x_{s1}, \theta^2_{s1})s^+(u_l, \theta_l)) + \kappa^3_{c1}(1 - s^+(x_{c2}, \theta_{c2})s^+(x_{s2}, \theta_{s2})s^+(u_g, \theta_g)) - \gamma_{c1}x_{c1}$ $0 < \theta^1_{c1} < \theta^2_{c1} < max_{c1}$ $0 < \kappa^1_{c1}/\gamma_{c1} < \theta^1_{c1}$ $0 < (\kappa^1_{c1} + \kappa^2_{c1})/\gamma_{c1} < \theta^1_{c1}$ $\theta^1_{c1} < (\kappa^1_{c1} + \kappa^3_{c1})/\gamma_{c1} < \theta^2_{c1}$ $\theta^2_{c1} < (\kappa^1_{c1} + \kappa^2_{c1} + \kappa^3_{c1})/\gamma_{c1} < max_{c1}$
$\dot{x}_{c2} = \kappa_{c2} - \gamma_{c2}x_{c2}$ $0 < \theta_{c2} < max_{c2}$ $\theta_{c2} < \kappa_{c2}/\gamma_{c2} < max_{c2}$
$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1}(1 - s^+(x_{c1}, \theta^2_{c1})s^+(x_{s1}, \theta^2_{s1})s^+(u_l, \theta_l)) + \kappa^3_{s1}(1 - s^+(x_{c2}, \theta_{c2})s^+(x_{s2}, \theta_{s2})s^+(u_g, \theta_g)) - \gamma_{s1}x_{s1}$ $0 < \theta^1_{s1} < \theta^2_{s1} < max_{s1}$ $0 < \kappa^1_{s1}/\gamma_{s1} < \theta^1_{s1}$ $0 < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < \theta^1_{s1}$ $\theta^1_{s1} < (\kappa^1_{s1} + \kappa^3_{s1})/\gamma_{s1} < \theta^2_{s1}$ $\theta^2_{s1} < (\kappa^1_{s1} + \kappa^2_{s1} + \kappa^3_{s1})/\gamma_{s1} < max_{s1}$
$\dot{x}_{s2} = \kappa_{s2} - \gamma_{s2}x_{s2}$ $0 < \theta_{s2} < max_{s2}$ $\theta_{s2} < \kappa_{s2}\gamma_{s2} < max_{s2}$
$\dot{x}_m = \kappa_m s^+(x_{c1}, \theta^1_{c1})s^+(x_{s1}, \theta^1_{s1})s^+(u_l, \theta_l) - \gamma_m x_m$ $0 < \theta_m < max_m$ $\theta_m < \kappa_m/\gamma_m < max_m$
$\dot{u}_l = 0$ $0 < \theta_l < max_l$
$\dot{u}_g = 0$ $0 < \theta_g < max_g$
Type 3 mutant: $\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} + \kappa^3_{c1} - \gamma_{c1}x_{c1}$ $\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1} + \kappa^3_{s1} - \gamma_{s1}x_{s1}$

Table 14: Model IV1

$\dot{x}_{c1} = \kappa^1_{c1} + \kappa^2_{c1} s^+(x_{c1}, \theta^1_{c1}) s^+(x_{s1}, \theta^1_{s1}) s^+(u_l, \theta_l)(1 - s^+(x_{c2}, \theta_{c2}) s^+(x_{s2}, \theta_{s2}) s^+(u_g, \theta_g)) - \gamma_{c1} x_{c1}$ $0 < \theta^1_{c1} < \theta^2_{c1} < max_{c1}$ $\theta^1_{c1} < \kappa^1_{c1}/\gamma_{c1} < \theta^2_{c1}$ $\theta^2_{c1} < (\kappa^1_{c1} + \kappa^2_{c1}/\gamma_{c1} < max_{c1}$
$\dot{x}_{c2} = \kappa_{c2} - \gamma_{c2} x_{c2}$ $0 < \theta_{c2} < max_{c2}$ $\theta_{c2} < \kappa_{c2}/\gamma_{c2} < max_{c2}$
$\dot{x}_{s1} = \kappa^1_{s1} + \kappa^2_{s1} s^+(x_{c1}, \theta^1_{c1}) s^+(x_{s1}, \theta^1_{s1}) s^+(u_l, \theta_l)(1 - s^+(x_{c2}, \theta_{c2}) s^+(x_{s2}, \theta_{s2}) s^+(u_g, \theta_g)) - \gamma_{s1} x_{s1}$ $0 < \theta^1_{s1} < \theta^2_{s1} < max_{s1}$ $\theta^1_{s1} < \kappa^1_{s1}/\gamma_{s1} < \theta^2_{s1}$ $\theta^2_{s1} < (\kappa^1_{s1} + \kappa^2_{s1})/\gamma_{s1} < max_{s1}$
$\dot{x}_{s2} = \kappa_{s2} - \gamma_{s2} x_{s2}$ $0 < \theta_{s2} < max_{s2}$ $\theta_{s2} < \kappa_{s2}\gamma_{s2} < max_{s2}$
$\dot{x}_m = \kappa_m s^+(x_{c1}, \theta^2_{c1}) s^+(x_{s1}, \theta^2_{s1}) s^+(u_l, \theta_l) - \gamma_m x_m$ $0 < \theta_m < max_m$ $\theta_m < \kappa_m/\gamma_m < max_m$
$\dot{u}_l = 0$ $0 < \theta_l < max_l$
$\dot{u}_g = 0$ $0 < \theta_g < max_g$
Type 3 mutant: $\dot{x}_{c1} = \kappa^1_{c1} - \gamma_{c1} x_{c1}$ $\dot{x}_{s1} = \kappa^1_{s1} - \gamma_{s1} x_{s1}$

Table 15: Model IV4