Internship report

Tamar Huygen
tamar@huygen.nl

December 1, 2016

Abstract

Participation iGEM 2015 my contribution

Chapter 1

Introduction

General Introduction

-climate change -our solution -modelling

1.0.1 Sustainability (under construction)

The main returning theme in this paper is stability. The word stability has different meanings in different contexts. In this paper several definitions of stability will be discussed. We will start with one that has to do with one of the greatest challenges that mankind has to face in this time. The stability of our own population, humanity, on this planet.

We might think that we will always be able to live like we were used to on our planet, but considering our own population dynamics (see figure 1.1) and the accompanying demand for energy and food. We can safely assume that using resources we cannot replenish, will not be a stable strategy. Sooner or later one of the irreplaceable resources will run out. When that happens we have to see whether the other resources are able to fill in the gap. Until that happens we have to look for more sustainable ways of producing food and energy. Resources that cannot run out fast, such as sunlight or geothermal energy, have the advantage that they will be a reliable energy source for as long as humanity lives on this planet.

In the way we are using the resources of the planet now we also have multiple environmental issues. Pollution of the environment may give an (economic) advantage on the short term, but polluting strategies are certainly not feasible considering longer timespans. Then there is also the problem of global warming. Because of the increasing use of fossil fuels a lot more CO_2 is released in the atmosphere and CO_2 levels have been rising. Although there are gasses with higher global warming potential, a slight change in the balance of gases in the atmosphere, can have an effect on temperature. This slight effect can also have an effect on other sources of greenhouse gases, such as release of methane from permafrost tundras which are now thawing, enhancing the greenhouse effect. This gas-compositions shift and accompanying temperature shift will have major ecological and economical consequences on our planet. If we want to slow this process down it is of major importance that we reduce the emission of greenhouse gases and CO_2.

This is were the bio-based economy comes in. More and more entrepreneurs and scientists are looking for less polluting ways to produce, where possible also with less CO_2 emission. They can do this with the help of biotechnological advances. In biotechnology, there are also still a lot of challenges. In the past food crops have been used to create fuel in a bio-based manner. This caused the crop prices to rise in certain areas, which is seen as a great disadvantage. We have to find a way to produce in a sustainable manner, without competing with food. By choosing a non crop organism, which harvests light and CO_2, like Synechocystis, we might get

a small step closer to reach this goal.

In an ideal world we would have closed carbon cycles. That which we produce in CO_2 would also be consumed. In this way we would enhance the stability of the environment in which mankind seems to prevail.
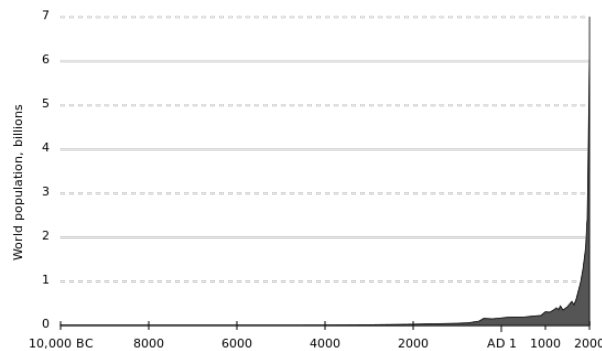

Figure 1.1: Our own population dynamics so far.

1.0.2 Stable bio-production

Bio-based production usually resolves around an organism producing a compound that is in some way useful to humanity. An easy way to make a certain organism produce a certain compound in a certain way is to transform the organism with the genes that are needed for the biosynthesis of the product. This means inserting genes However this is not a very stable form of production. Populations of organisms with

1.0.3 Stable consortium

1.0.4 The search for a sustainable solution and synthetic biology

A lot of different approaches in the search for sustainable production and energy harvest have been used.

1.0.5 The team iGEMAmsterdam 2015 Solution

The iGEM team of

iGem cyanobacteria, filipe overview literature, oview project, stability, my contribution When is something considered stable? Why a symbiose? Why Synechocystis? Why E. Coli? Why the algorithms? Why kinetic models? How does this module fit into the project?

Chapter 2

Algorithms

Abstract

2.1 Introduction

An important goal in the bio-industry is the production of certain carbon containing compounds. When we let micro-organisms produce these compounds, there is always a stability challenge. Usually it is better in terms of fitness for an organism use all carbon it can for the generation of biomass. So consider an organism which where we have inserted genes which forces the organism to produce a certain compound and transport it out of the cell, so that we may use it. A mutation may yield an organism which does no longer produce the compound. Organisms with this mutation likely have a fitness advantage over organisms without this mutation, because they can use all carbon compounds for generation of biomass and don't leak carbon compound to the environment. This means that soon after the appearance of the mutation the non-productive organisms will take over the entire population and the production stops and has to be re-initiated. We want to address this problem and have created an algorithm which searches for solutions that force an organism to produce a carbon compound, with a much smaller chance of rendering a non-productive population.

2.1.1 Models

When studying cells in silico one is always dependent on models. To study metabolism of cells and the connections to the genetic network, flux balance analysis (FBA) has become an important tool. (hier een stukje in appendix over fba??) FBA models consist of a set of differential equations that model the flux of mass through reactions as these reactions are assumed to be in steady state. The models simulate the more or less constant nett flux of mass converted by the reaction. It is important to note that the base of the model consists simulations of reactions. Genes play an important role in the models, but only in terms of the influence of the genes on the reactions. In reality there is a complex interaction between genes, gene regulation, transcription, translation and the reactions. In our models the genes are only encoded in the influence they eventually have on the flux through the reactions. The reactions present in the model sometimes have gene associations. These gene associations list genes, that when present and active, enable the reaction to happen. If there are multiple genes listed in the gene association they can have an *or* or an *and* relationship. In figure !! two reactions are shown with the genes that are listed in the gene association of these reactions. For reaction 1 to happen, either gene 1 or gene 2 are has to be active. For reaction 2 both gene 2 and gene 3 are have to be active. So when gene 2 is deactivated or deleted, while gene 1 and gene 3 stay active, reaction 2 is taken out, while reaction 1 is still active. What happens in the model to simulate the inactivity of gene 2 is that the flux of reaction 2 is set to zero. The models we used were not perfect. Especially the *and* and *or* relationships between the genes cannot be trusted, simply because we don't know these relationships most of the times. How can we then say anything about which genes we have to alter in order to have the cell create a certain compound? If we don't know the *and* and *or* relationships, we still have the gene associations, see figure !!. Assume we want to prevent a certain reaction from happening. Let us call this reaction the primary reaction. Now we assume our algorithm does the following:

1. Make a list of all genes associated to the primary reaction

2. Make a list of all reactions that have at least one the genes from the list in its gene associations. We will call these reactions secondary reactions.

3. Set the fluxes of all secondary reactions to zero

4. Analyse model

5. Check if cells still grow

6. Check if compound of interest is produced by the cell

7. If cell still grows and produces compound, return set of genes

This algorithm is in two aspects conservative in the solutions it returns. In reality we are not sure what will happen when only a subset of genes associated to a reaction gets deleted. We want to avoid the risk that a primary reaction still takes place when any of the associated genes is still present. We also don't want to render non-viable cells after we have deleted the genes. A secondary reaction might be essential for cell growth. We want to avoid the risk the algorithm returns a seemingly valid solution, where actually a secondary reaction that is essential for cell growth has been stopped. This is why we assume all secondary reactions are stopped, even though only a subset of the associated genes of these reactions is deleted.

2.1.2 Stable compound generator

The goal of the stable compound generator is to find genes which can be deleted in order to create an organism which has to produce a certain carbon compound. Consider the reaction scheme in figure **??**. Our goal is to produce compound B. Our initial plan to achieve this goal is to take out reaction 2. Assume reaction 2 is the only reaction which consumes B. By taking out the reaction we hope the cell will not be able to use B and accumulate and transport it out of the cell.

We have 2 prerequisites to be able to have a stable production of compound B. The first is that reaction 1 is essential for the production of biomass. If cells that don't use reaction 1 are still viable and produce biomass, they will most likely have a fitness advantage over cells that still use reaction 1, because they won't have to produce compound B and can use more carbon for biomass generation. But when cells that take out reaction 1 are non-viable the cells can never take over the population. The second requirement is that we are able to take out reaction 2 and still yield viable cells, so reaction 2 has to be non-essential for the production of biomass. It is reaction 2 which we want to find. So a first algorithm can look for all reactions that consume B and take all of them out and see if the cells still produce biomass. If so, it returns reaction 2 to us.

Combinatorial problems

The reaction schemes shown in figure !! are in some way all simplifications. Most reactions have a forward reaction and a reverse reaction. When a reaction is in chemical equilibrium, the forward reaction happens in the same rate as the reverse reaction. This means the net flux of mass through the reaction is zero. Potentially the rate of the reverse reaction can be higher than the rate of the forward reaction. This causes a negative flux through the reaction. Consider figure !!. Under normal circumstances B is consumed by reaction 2 and reaction 3 is in chemical equilibrium. Only when reaction 2 is taken out, reaction 3 is used to consume B and the forward flux becomes positive. In order for the cell to produce B reactions 2 and 3 have to be taken out simultaneously. An algorithm which searches for reactions that have a mass flux from B to another substance, will only find reaction 2, because initially the flux through reaction 3 is equal to zero. Then after it sets the flux through reaction 2 to zero, the flux through reaction 3 will become positive and the cell would still not produce B. To solve this problem the algorithm will have to search for all reactions which are potentially able to consume B. This means all reactions that list compound B as one of its products or as one of its reactants, independent of the mass flux. An algorithm that would search for these reactions would also list the reaction that initially produces the compound. In figure !! reaction 1 is

such a reaction. Earlier we assumed reaction 1 is essential for the production of biomass and will yield non-viable cells. When the algorithm would set the flux of all reactions that list compound B one of its products as well as all reactions that list B as one of its reactants to zero it would also set the flux of reaction 1 to zero. When it then analyzes the model it will find that the cell is not viable anymore. So in order to find the combination of reactions that we have to take out in order for B to be produced (in this case take out reaction 2 and reaction 3, but not reaction 1) we will have to test combinations of reactions. Another possibility is that there are multiple reactions such as reaction 1 which produce B, and are essential for biomass generation. This means that we cannot just find reaction 1 and only set the flux through all other compound B involving reactions to zero. So the algorithm has to list all reactions that potentially consume B, and find a subset of reactions which it can take out such that the cell is still viable and produces B. In order to find this subset we have to try out all possible subsets. In this scheme the possibilities the algorithm would create the following list of reactions: $1, 2, 3$. Then it will have to create a list of all combinations of these reactions: $1, 2, 3, 1, 2, 1, 3, 2, 3$ (the list doesn't contain $\emptyset$, nor $1, 2, 3$ because we don't have to try taking out no reactions, because then we wouldn't change the cell and we don't have to try taking out all reactions, because then we will also take out all reactions that produce B and we cannot produce B to begin with). The algorithm then has to try each of these subsets. In the example given by figure !! it is possible to try out all these possibilities within a reasonable amount of time. In all combinations where reaction 1 is taken out the cells are not viable, so such an algorithm won't return them and in cases where only reaction 2 or reaction 3 is taken out, B won't be produced, so the algorithm also won't return them, but the algorithm will return $2, 3$ given that reactions 2 and 3 are not essential for cell growth. However if n is the number of reactions where B is involved in, the amount of possible combinations increases with 2^n. This means the algorithm will exponentially become slower with the amount of reactions that are involved. In the implementation of the algorithm we will define an upper limit of combinations which are tried out before it stops. We will start with combinations in which only a small number of reactions has to be taken out, because this is also the easiest to achieve in vivo.

Completeness

Now consider the reaction scheme shown in figure !!. To produce B we might take out reaction 2, but another possibility might be to take out reaction 3. When C accumulates the flux through reaction 2 might become 0 and B might also be produced and transported out of the cell. Our algorithm won't find this possibility, because it only searches for reactions that potentially consume B. This means our algorithm is not complete and will not return all possible reactions which can be taken out in order to create an organism that produces B. We have shortly tried to find reactions further down the pathways, but usually this yields almost all reactions of a cell. This in combination with the earlier described combinatorial problems yields an unusable algorithm. So for now we accept that the algorithm is not complete and does not guarantee that it finds all possible reactions that can be taken out to produce B.

From reactions to genes

Up until now we've only spoken about cells in terms of reactions and our algorithm takes out reactions. In reality it is very hard to influence reactions in cells directly. More often we try to manipulate reactions that happen in cells through genes. We can stop reactions for example by taking out genes that encode enzymes which catalyze a reaction. Since we want to find out what genes we can take out instead of reactions directly, we also would want our algorithm to provide these genes for us. We can change the algorithm to do this. The algorithm then has to do the following:

- List all reactions that potentially consume B

- List all combinations of reactions that when taken out potentially can make the cell produce B

- For each combination of reactions just listed, for each reaction in the combination, list the genes associated with the reaction.

- For each listed gene association list the combination of genes that can be turned of in order to make the reaction no longer happen.

- For each combination of genes just listed, turn off the genes (this means the algorithm will find each reaction which has the gene in its gene association and determines which of these reactions is taken out when the genes are taken out of this genes and sets the flux through these reactions to 0)

- Analyze model and check if cell still produces biomass and if compound B is produced.

- Return successful combinations of genes.

How can we then say anything about which genes we have to alter in order to have the cell create a certain compound? To solve this problem we try to be conservative when to return a solution. We can let the algorithm list all genes associated with the reactions we want to take out. Then we list all reactions which have one of these genes listed in their gene associations. To take out a certain reaction, we take out all genes associated with it. To simulate this, we take out all reactions which have one of these genes in its gene association. We can see that there are now two contradictory assumptions in play. Consider In order to take out one reaction we first assume we have to take down all gene associated with it.

2.1.3   Auxotrophy sniper

Auxotrophes are organisms that have an auxotrophy. An auxotrophy is a dependence on a certain compound for a cell to grow. See previous sections. The general idea is that an auxotroph can be created by knocking out a combination of genes involved in the production of the compound we want to make it dependent on. The organism should not be able to grow if these genes are knocked out and there is no source for the compound in the extracellular space. Once again our algorithm searches for such genes. When we found a certain combination of genes we want to test, we deactivate them in the model, and then analyse the model to see if the organism is not able to grow anymore. When this is the case we create a source in the extracellular space (in this way we simulate adding compound to the medium) and once again analyse the model. Now the organism should be able to grow again. Here follows an overview of how the Auxotrophy Sniper works: The algorithm takes a list of compounds of which you want to make a synthesis deficiency in the organism of choice. This list can contain for example vitamins or amino acids. For each metabolite on the list it does the following: âŮŔ Set flux boundaries of source reaction in the extracellular space (if there is already one present) to zero. âŮŔ Find all primary reactions associated to the compound we want to make the organism dependent on. âŮŔ For each primary reaction, find genes which are associated to the primary reactions. âŮŔ For each gene find all reactions which have the gene in their gene association (secondary reactions). âŮŔ Make a list of possible combinations of genes which can be knocked out. âŮŔ Go over the combinations oneÂŋbyÂŋone, for each gene in a combination turn off primary and secondary reactions. âŮŔ Per combination, do a flux balance analysis and check for biomass formation (growth). âŮŔ If it does not form any biomass, a source reaction of the compound is added to the model âŮŔ Another flux balance analysis of the model is done
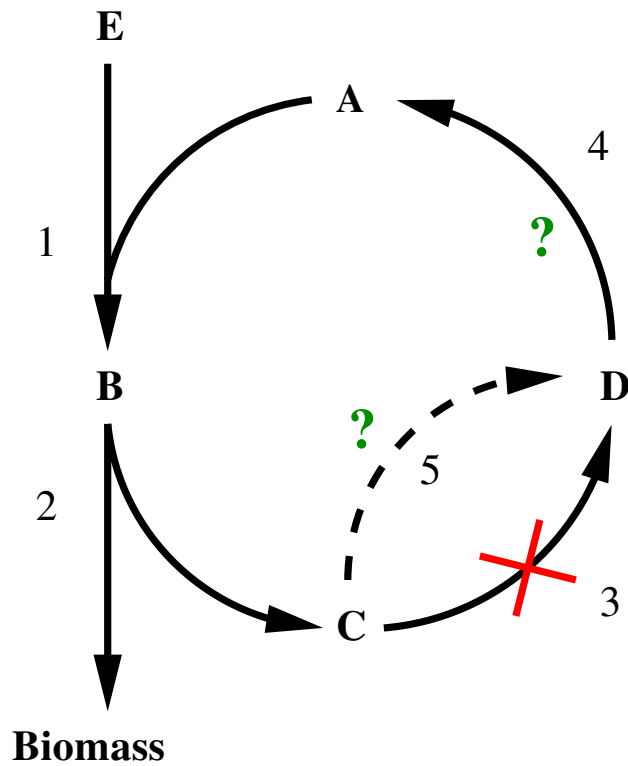
Figure 2.1: Scematic overview of a pathway in which product C is produced when reaction 3 is taken out. The algorithm finds reaction 3 and takes it out and checks if biomass is still produced and if so if C is produced. However, suppose there is another pathway which the cell can use to recycle substance C, (in this figure symbolized by reaction 5), then the cell might make use of this pathway to recycle C. Taking out all reactions which might use substance C will also take out reaction 2, since in reality most of the reactions are . When reaction 2 is taken out, C won't be produced.

and if there is biomass formation now, the combination of genes are good candidates. If this combination of genes is knocked out the organism will probably become an auxotroph.

2.2 Methods

2.2.1 Implementation

Genes

Because the models we use are not perfect, the and- and or- relations between the genes cannot be trusted. We still want the algorithm to return genes. Consider figure !!. Since we cannot trust the *and* and *or* relations, these are left out of the figure. Now if we want to take out reaction 1 we look at all genes associated to it (in this case gene 1 and gene 2). Now we turn off all genes associated to reaction 1. We turn these genes off by finding all reactions which have one of these genes in the gene association. The fluxes of all these reactions are set to 0. So in this case we find gene 1 and gene 2 which are associated with reaction 1. Now we find all reactions which have these genes in its gene association and find reaction 1 (gene 1 and gene 2) and reaction 2 (gene 2). In order to simulate taking out gene 2 we set the fluxes of reaction 1 and reaction 2 to zero. This seems contradictory: we say reaction 1 only stops when all genes associated to it are taken out. While when we just take out gene 2 we assume

2.3 Results

2.3.1 Stable compound generator

het algoritme (een paar relevante algoritmes) toepassing lijst met genen die mogelijk uit te schakelen zijn Toepassing van iGem

2.4 Discussion

completeness sinks Christine's experiment opbrengst

Algorithm 1 Old (simple) stable compound generator algorithm. We find reactions associated with a certain compound and take a combination of these reactions and switch them of. In reality we cannot switch off reactions, we have to remove genes. This is why this is just a simple version of the algorithm.

1: **procedure** FINDREACTIONS($compound$, $model$)
2:   $hits \leftarrow \{\ \}$
3:   $r1 \leftarrow [\ ]$
4:   **for** $r \in model.reactions$ **do**
   ▷ $model.reactions$ gives a list of all reactions in the model. r loops over all these reactions
5:    **if** $compound \in r.products \lor compound \in r.reactans$ **then**
    ▷ $r.products$ is a list of all products produced by reaction and $r.ractants$ is a list of all reactants of reaction r.
6:     $r1.append(r)$
7:    **end if**
8:   **end for**
9:   **if** $length(r1) > 1$ **then**
10:    $combinations \leftarrow [comb(r1)]$
    ▷ $combinations$ is a list of lists. Each list in combinations contains a combination of reactions that are in r1
11:    $genes \leftarrow \{\ \}$
12:    **for** $c \in combinations$ **do**
13:     **for** $r \in c$ **do**
14:      $r.flux \leftarrow 0$
15:     **end for**
16:     $analyseModel(model)$
17:     **if** $growth > 0$ **then**
18:      **if** $compound \in hits.keys()$ **then**
19:       $hits[compound].append(c)$
20:      **else**$hits[compound] \leftarrow [c]$
21:      **end if**
22:     **end if**
23:    **end for**
24:   **end if**
25: **end procedure**

Chapter 3

Kinetic models

One-way dependency

Unlimited cell growth is exponentially. The amount of biomass per time for an exponentially growing species can be given by the following different equation:

$$\frac{da}{dt} = \mu a \tag{3.1}$$

Herein is a the amount of biomass per liter and μ the growth rate normalized for biomass. One can easily verify that the solution of this differential equation is indeed exponential growth ($a = ce^{\mu t}$). Now from experimental data it has been shown that limited growth on a substrate is a bit different. The normalized growth rate is then dependent on the concentration of substrate. According to the monod equation, the μ is dependent on $[S]$ in the following way:

$$\mu = \mu_{max}\frac{[S]}{k_S + [S]} \tag{3.2}$$

Herein is μ_{max} the maximal growth rate (equal to the growth rate at unlimited growth), and $[S]$ the concentration of substrate. k_S is the concentration of $[S]$ at a rate $\frac{1}{2}\mu_{max}$. We also know this equation from enzyme kinetics as the Michaelis-Menten equation. In enzyme kinetics this equation is used to calculate the rates at which enzymes convert products. However the Michaelis-Menten equation is based on theoretical arguments, while Monod is based on experimental findings. According to Monod, the growth rate saturates as the concentration becomes higher (see figure 3.1).

In our consortium, Escherichia coli grows limited on acetate. So now we know how μ depends on the concentration of acetate. We now need to model the concentration of acetate. We assume the synthesis of acetate is growth coupled and depends linear on the growth of *Synechocystis*. We also know that the maximal uptake rate of acetate by *E. coli* is dependent $1/y$ herein is y the yield of *E. coli* on acetate in is in milligram Dry Weight *E. coli* per millimole acetate per liter. The uptake of substrate per unit time is also in a a saturable way dependent on the concentration of S, exactly the same way the growth speed is dependent on the the concentration of substrate. The synthesis of the substrate is growth coupled and is dependent on the amount of biomass of *Synechocystis* formed in time. The ch We now arrived at the following set of differential equations:

$$\frac{d\text{syn}}{dt} = \mu_{syn}\text{syn} \tag{3.3}$$

$$\frac{d\text{ec}}{dt} = \mu_{max,ec}\frac{[S]}{k_{s,ec} + [S]}\text{ec} \tag{3.4}$$

$$\frac{d[S]}{dt} = \frac{1}{y_{s,syn}}\mu_{syn}\text{syn} - \frac{1}{y_{s,ec}}\frac{[S]}{k_{s,ec} + [S]}\text{ec} \tag{3.5}$$
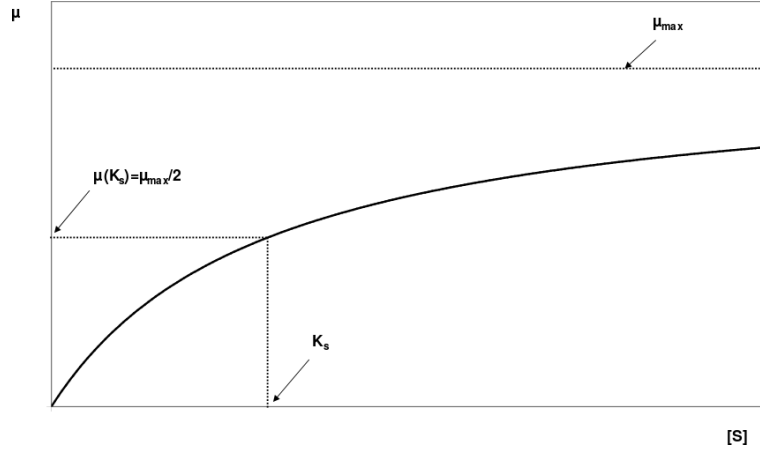
Figure 3.1: Limited growth on a substrate according to the Monod equation. μ is the normalized growth rate in units per hour μ_{max} is the maximal growth rate and $[S]$ is the substrate concentration. k_S is the concentration at the rate equal to $\frac{1}{2}\mu_{max}$.

Herein is syn the amount of biomass of *Synechocystis* and ec the amount of biomass of E. coli. y_s is the substrate yield of *Synechocystis*. Since in this model the substrate is only formed when *Synechocystis* forms biomass, per amount of biomass formed, there is a constant amount of substrate formed. The yield is usually expressed in gram dry weight mole substrate *used*. In this case we mean gram dry weight mole substrate *formed*. So to find the amount of substrate that is formed per amount of biomass that is formed we simply take $\frac{1}{y,syn}$.

It can be easily seen that such a relationship will not be stable if $\mu_{max,ec} << \mu_{max,cyn}$. In this case by stability we mean the convergence of the growth rate to the same value.

Substrate dependent

In this model, *Synechocystis* is not dependent on E. coli. There are two ways in which *Synechocystis* may be dependent on E. coli we have explored. Firstly, *E. coli* may produce a substrate *Synechocystis* grows on, as is the case with the auxotrophic *Synechocystis*. Secondly, E. coli may decrease the light intensity in the culture, in this way slowing down the growth of *Synechocystis*. The growth rate of *Synechocystis* can only be limited by one of these two processes and it will always be limited by the process that slows it the most. Either μ_{syn} is lower than $\mu_{max,syn}$ because there is a photon shortage, but then the amount of substrate available at that growth rate would be enough, or the amount of substrate is limiting, but then the amount of photons available would also be enough for that given growth rate. So actually the growth rate of *Synechocystis* would be

$$\min(\mu_{max,syn}\frac{[S_2]}{k_{s,syn} + [S_2]}, \mu_{max,syn}f(\text{syn}, \text{ec})) \tag{3.6}$$

Herein is $f(\text{syn}, \text{ec})$ a function which determines the factor of decrease in growth rate because of a photon shortage and it is a function of the amount of biomass per liter of *Synechocystis* as well as that of E. coli. If we now assume that the amount of substrate E. coli produces is going to be limiting we arrive at the following set of differential equations.

$$\frac{d\text{syn}}{dt} = \mu_{max,syn} \frac{[S_2]}{k_{s2,syn} + [S_2]}\text{syn} \tag{3.7}$$

$$\frac{d\text{ec}}{dt} = \mu_{max,ec} \frac{[S_1]}{k_{s1,ec} + [S_1]}\text{ec} \tag{3.8}$$

$$\frac{d[S_1]}{dt} = \frac{1}{y_{s,syn}} \mu_{max,syn} \frac{[S_2]}{k_{s2,syn} + [S_2]}\text{syn} - \frac{1}{y_{s1,ec}} \frac{[S]}{k_{s1,ec} + [S]}\text{ec} \tag{3.9}$$

$$\frac{d[S_2]}{dt} = Q_{p,ec}\text{ec} - \frac{1}{y_{s2,syn}} \frac{[S_2]}{k_{s2,syn} + [S_2]}\text{syn} \tag{3.10}$$

Herein is $Q_{p,ec}$ the amount of $[S_2]$ formed by E. coli per gram dry weight of E. coli. We assume E. coli doesn't produce in a growth coupled way, but has a constant production per amount of biomass.

$$\frac{d\text{syn}}{dt} = \mu_{max,syn} \frac{[S_2]}{k_{s2,syn} + [S_2]}\text{syn} \tag{3.11}$$

$$\frac{d\text{ec}}{dt} = \mu_{max,ec} \frac{[S_1]}{k_{s1,ec} + [S_1]}\text{ec} \tag{3.12}$$

$$\frac{d[S_1]}{dt} = \frac{1}{y_{s,syn}} \frac{d\text{syn}}{dt} - \frac{1}{y_{s1,ec}} \frac{d\text{ec}}{dt} \tag{3.13}$$

$$\frac{d[S_2]}{dt} = Q_{p,ec}\text{ec} - \frac{1}{y_{s2,syn}} \frac{d\text{syn}}{dt} \tag{3.14}$$

here

Light limited growth

Like stated in section 3, it is also possible for the photoautotroph to grow light limited instead of substrate limited. As the cell density of both organism increases, the available light for the photoautotroph decreases. In equation 3.6 it is already suggested that this light dependency might be a function which is dependent on the biomass of E. coli and the biomass of Synechocystis. But let's first take a look at light dependent growth. According to literature [?], the μ if a photoautotroph is dependent on the light intensity in the following way:

$$\mu(I) = \mu_{max,syn} \frac{I}{K_{s,syn} + I + \frac{I^2}{K_i}} \tag{3.15}$$

Herein is $\mu(I)$ the specific growth rate on light, $\mu_{max,syn}$ the maximal specific growth rate of Synechocystis, I the light intensity, $k_{s,syn}$ is the saturation constant on light and K_i is an inhibitory constant. Now the light intensity is not only dependent on the cell density of the photoautotroph and the chemoheterotroph, but it is also dependent on the place in the reactor. However, to not further complicate the model we made a simplification, by assuming the light intensity is linear dependent on the biomass concentration of Synechocystis and E. coli in the following way:

$$I(\text{syn}, \text{ec}) = \max(I_{max} - sh_s \cdot \text{syn} - sh_e \cdot \text{ec}, 0) \tag{3.16}$$

Herein are sh_s and sh_e 'shading coefficients' for Synechocystis and E. coli respectively. In this way the light intensity linear decreases with the amount of biomass of E. coli and Synechocystis,

but it can never be lower than zero. We then arrive at the following set of differential equations:

$$\frac{d\text{syn}}{dt} = \mu_{syn}\text{syn} \tag{3.17}$$

$$\frac{d\text{ec}}{dt} = \mu_{ec}\text{ec} \tag{3.18}$$

$$\frac{d[S]}{dt} = \frac{1}{y_{s,syn}}\mu_{syn}\text{syn} - \frac{1}{y_{s1,ec}}\mu_{ec}\text{ec} \tag{3.19}$$

$$\tag{3.20}$$

with:

$$\mu_{syn} = \mu_{max,syn}\frac{I}{I + k_{s,syn} + \frac{I^2}{k_i}} \tag{3.21}$$

$$\mu_{ec} = \mu_{max,ec}\frac{[S]}{[S] + k_{s,ec}} \tag{3.22}$$

With I the light intensity, dependent on the biomass or *E. coli* an *Synechocystis* as given in equation 3.16.

3.0.1 Turbidostat

If we want to model the consortium in a turbidostat, we have to account for the fact that both *Synechocystis* and E. coli are increasing the OD as they grow. This means that the dilution rate is dependent on the biomass of *Synechocystis* as well as that of E. coli. For simplicity we make the assumption that there is a constant flow through the system, instead of only diluting when the threshold is reached. To understand this we first look at the case of a single strain, called b. In a chemostat the growth rate of the organism would become equal to the dilution rate. In a turbidostat however, an organism can grow at its maximal growth rate, but the amount of biomass must still become constant. This means the following:

$$\frac{db}{dt} = \mu \cdot b - D \cdot b = 0 \tag{3.23}$$

Where b is the amount of biomass of the strain b, mu is the growth rate and D the dilution rate.

In a chemostat it would mean that $D < \mu_{max}$ is a chosen dilution rate and that μ becomes equal to D due to substrate limitation. In a turbidostat however, D becomes equal to μ_{max}, because the species in not growing limited.

In the case where there are two strains, strain a and b, that share a turbidostat, the differential equations that then describe the system looks like the following:

$$\frac{da}{dt} = f(a, b, t) - Da \tag{3.24}$$

$$\frac{db}{dt} = g(a, b, t) - Db \tag{3.25}$$

Herein are f and g functions that describe the growth of the organisms a and b respectively. D is again the dilution rate. Now in a shared turbidostat it holds that $a + b = k$, where k is a

constant amount of biomass. Then the following holds:

$$a + b = k \tag{3.26}$$

$$\implies \frac{da}{dt} + \frac{db}{dt} = 0 \tag{3.27}$$

$$\implies f - Da + g - Db = 0 \tag{3.28}$$

$$\implies Da + Db = g + f \tag{3.29}$$

$$\implies D = \frac{g + f}{a + b} \tag{3.30}$$

For the turbidostat we then arrive at the following set of differential equations:

$$\frac{d\text{syn}}{dt} = \mu_{syn}\text{syn} - D\text{syn} \tag{3.31}$$

$$\frac{d\text{ec}}{dt} = \mu_{ec}\text{ec} - D\text{ec} \tag{3.32}$$

$$\frac{d[S_1]}{dt} = \frac{1}{y_{syn/S1}}\mu_{syn}\text{syn} - \frac{1}{y_{ec/S1}}\mu_{ec}\text{ec} - D[S_1] \tag{3.33}$$

$$\frac{d[S_2]}{dt} = Qp_{S2/ec}\text{ec} - \frac{1}{y_{syn/S2}}\mu_{syn}\text{syn} - D[S_2] \tag{3.34}$$

where

$$D = \frac{\mu_{syn} + \mu_{ec}}{\text{syn} + \text{ec}} \tag{3.35}$$

$$\mu_{syn} = \mu_{max,syn}\frac{[S_2]}{k_{s2,syn} + [S_2]} \tag{3.36}$$

$$\mu_{ec} = \mu_{max,ec}\frac{S_1}{k_{s1,ec} + [S_1]} \tag{3.37}$$

3.1 Results

3.1.1 Batch

One-way dependency

First we simulated the one-way dependent model as depicted by equations 3.3. To find out whether the system is robust with respect to the initial conditions we did some simulations. Our expectations were that the biomass ratio of the photoautotroph and the chemoheterotroph in this model would converge irrespective to the initial conditions. To find out whether the biomass ratio of this system converges, we did some simulations. To be able to make simulations with the models, we have to estimate and measure the parameters. We tried to measure all parameters as accurate as possible. In table 3.1 the measured parameters are given.

It can be seen from this table, that $\mu_{max,ec} < \mu_{max,syn}$. It is very unlikely that this is correct. It might be caused by experimental Since μ of *Synechocystis* higher than the μ_{max} of *E. coli* and given that the growth of *Synechocystis* is independent, we can already expect this model with these parameter to be unstable. *Synechocystis* will keep growing faster than *E. coli*, because the growth rate of *Synechocystis* has a constant value that is higher than the maximal growth rate *E. coli* will ever achieve. If we simulate the model with these parameters find the biomass ratio given in figure 3.2. It can be seen that the ratio of biomass does not seem to converge. A more clear image can be seen if we look at the *mu*'s of *E. coli* and *Synechocystis* in figure 3.3.

$\mu_{max,syn}$ (h^{-1})	0.075
$\mu_{max,ec}$ (h^{-1})	0.08
$k_{s,ec}$ (mM)	0.02
$y_{\frac{s}{syn}}$ (mmol $\cdot$ gDW^{-1})	0.40
$y_{\frac{ec}{s}}$ (gDW $\cdot$ mmol^{-1})	0.031

Table 3.1: Measured parameters for the one-way dependency model. It can be seen that $\mu_{max,ec} <$ $\mu_{max,syn}$. This is highly unlikely the correct value for $\mu max, ec$
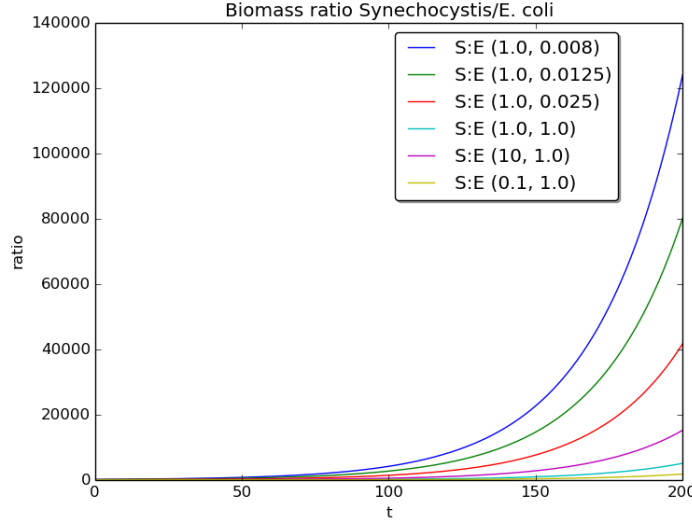

Figure 3.2: Boiomass ratio's of *Synechocystis*/*E. coli* in time. They seem to diverge, dependent on the initial conditions. This seems logical, since the growth rate of *E. coli* will never become as high as that of *Synechocystis*. In the legend the initial ratio of *Synechocystis* biomass : *E. coli* biomass is given. The model is given by equations 3.3. Parameters are given by table 3.1

Chemoheterotrophs are usually have a really high growth rate compared to photoautotrophs. Even though the maximal growth rate of *E. coli* on Acetate might be less than the maximal growth rate on glucose, it is very unlikely that the maximal growth rate is less than that of Synechocystis. So, the question remains, will the biomass ratio converge if we change the parameters such that $\mu_{max,ec} > \mu_{max,syn}$? To test this we have increased $\mu_{max,ec}$ tot 0.08, which is just higher than the growth rate of Synechocystis. In table 3.2 the new parameters are shown.

$\mu_{max,syn}$ (h^{-1})	0.075
$\mu_{max,ec}$ (h^{-1})	0.08
$k_{s,ec}$ (mM)	0.02
$y_{\frac{s}{syn}}$ (mmol $\cdot$ gDW^{-1})	0.40
$y_{\frac{ec}{s}}$ (gDW $\cdot$ mmol^{-1})	0.031

Table 3.2: Measured and estimated parameters for the one-way dependency model. It can be seen that now $\mu_{max,ec} > \mu_{max,syn}$.

In figure 3.4, the biomass ratio over time given several different initial conditions is shown. Now the biomass ratio does seem to converge. However, this ratio is more than ten times higher than what is found in the lab. This might suggest that we still use inaccurately measured
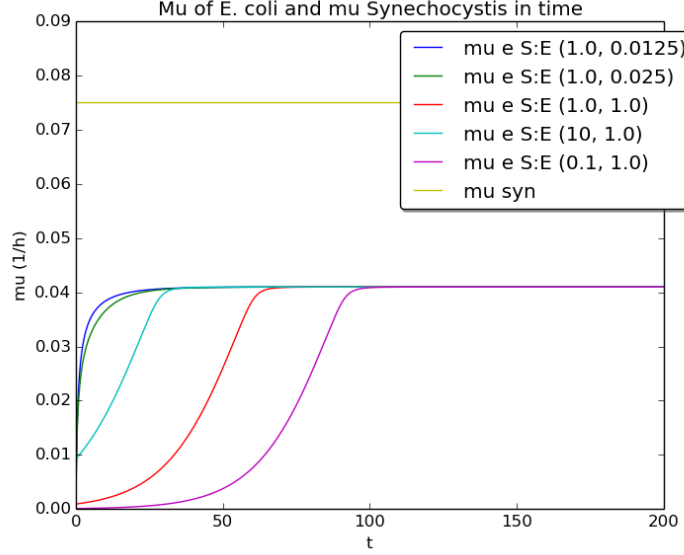
Figure 3.3: μ_{syn} and μ_{ec} under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$). μ_{ec} converges to $\mu_{max,ec}$, while μ_{syn} stays constant at the value of $\mu_{max,syn}$. The model is given by equations 3.3. Parameters are given by table 3.1
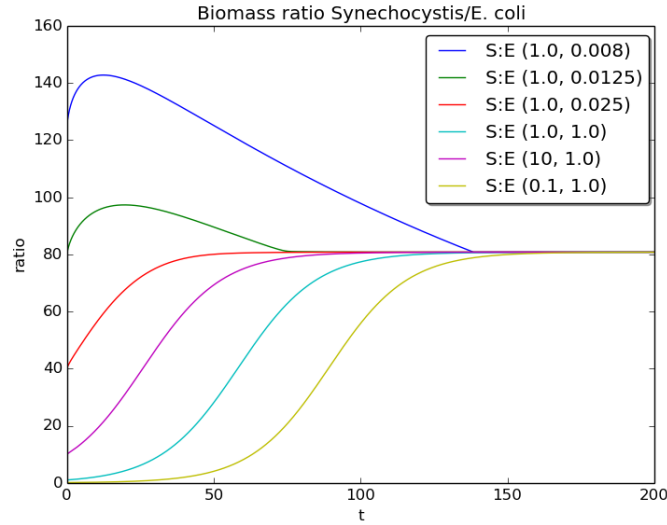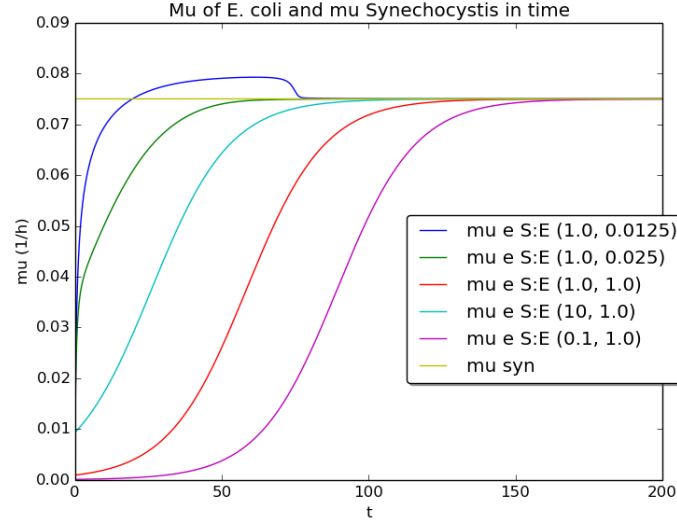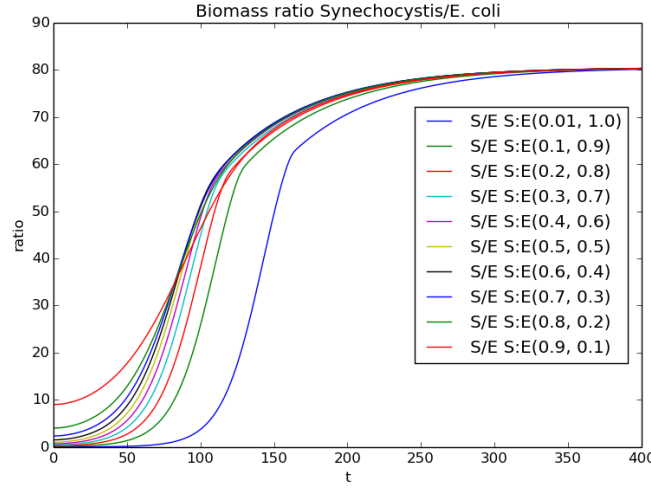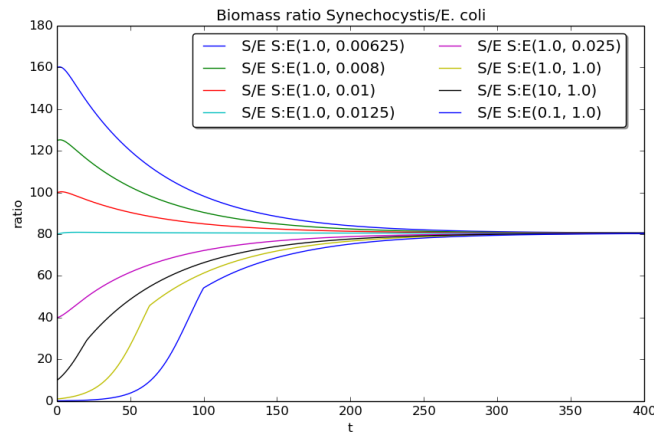
parameters.

Figure 3.4: The biomass ratio of *Synechocystis E. coli* under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$). The biomass ratio converges to a constant ratio. The model is given by equations 3.3. Parameters are given by table 3.2

If we look once again at the μ's over time in figure 3.5 we can see that the μ_{ec} converges to μ_{sys}. This is quite intuitive, since the growth rate of *E. coli* is limited by the production of substrate, which is dependent on the growth rate of *Synechocystis*. And the biomasses can only converge to a stable ratio, if the growth rates of both species converge to the same value. We can conclude that the system now is robust with respect to the initial values.

Figure 3.5: μ_{syn} and μ_{ec} under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$). The μ_{ec} converges to $\mu_{syn} = \mu_{max,syn}$. The model is given by equations 3.3. Parameters are given by table 3.2.

Substrate interdependency

To simulate the interdependency on substrates, we needed a lot more parameters. For some of the parameters we had no good measurements, so we estimated them. The general behavior stays generally the same under

$\mu_{max,syn}$ (h^{-1})	0.075
$\mu_{max,ec}$ (h^{-1})	0.08
$k_{s1,ec}$ (mM)	0.02
$k_{s2,syn}$ (mM)	0.02
$y_{\frac{s1}{syn}}$ (mmol $\cdot$ gDW^{-1})	0.40
$y_{\frac{ec}{s}}$ (gDW $\cdot$ mmol^{-1})	0.031
$y_{\frac{s2}{syn}}$ (mmol $\cdot$ gDW^{-1})	0.1
$Q_{p,ec}mmol$ (mmol $\cdot$ gDW$^{-1} \cdot$ h^{-1})	0.13

Table 3.3: Measured and estimated parameters for the substrate interdependent model.

Figure 3.6: The biomass ratio of *Synechocystis E. coli* under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$). The biomass ratio converges to a constant ratio. The model is given by equations 3.7. Parameters are given by table 3.4

Figure 3.7: The biomass ratio of *Synechocystis E. coli* under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$). The biomass ratio converges to a constant ratio. The model is given by equations 3.7. Parameters are given by table 3.4

3.1.2 Turbidostat

In a turbidostat you can assume the growth won't be light limited, since every time a threshold OD is measured, the culture is diluted. This means you can regulate the light intensity in such a way, it won't become limiting.

This is why we chose to mainly look at the consequences of growing a consortium with a substrate dependent photoautotroph in a turbidostat. So the model we use is described in equations 3.42 to 3.48 in section 3.2. The parameters used are given in table 3.4 In a turbidostat it is expected that the total amount of biomass, so that if *E. coli* and *Synechocystis* combined would stay constant. To test this, we have plotted the total amount of biomass under different initial conditions. The result is show in figure 3.16, and we can see that this is indeed constant. This however, does not mean that the amount of biomass of each species stays the same. In figure 3.15 several examples of the amount of biomass in time of *Synechocystis* and *E. coli* are given.
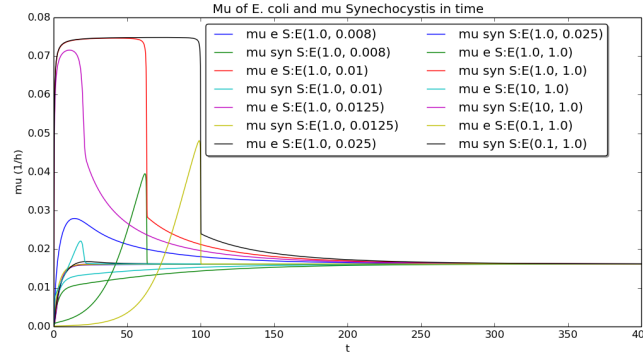
16

Figure 3.8: μ_{syn} and μ_{ec} under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$). The μ_{ec} and μ_{syn} converge both to the same value. The model is given by equations 3.7. Parameters are given by table 3.4.
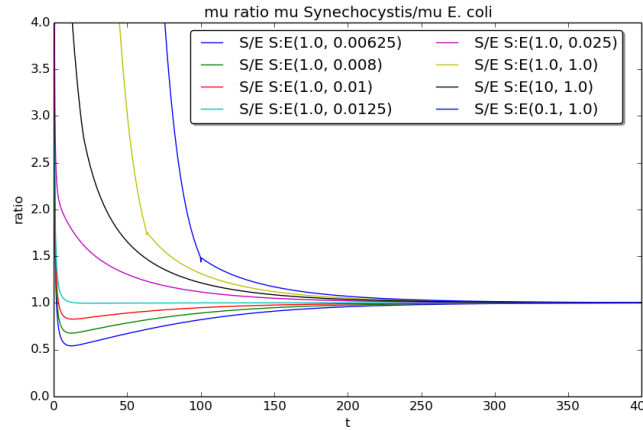

Figure 3.9: Ratio between μ_{syn} and μ_{ec} under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$). The μ_{ec} and μ_{syn} both converge to the same value, so the ratio converges to one. The model is given by equations 3.7. Parameters are given by table 3.4.

Now we can once again look at the biomass ratio to see if this converges. In figure 3.19 the biomass ratio in time is given under different initial conditions.

We can see that indeed the biomass ratio converges. Also if we look at the values for μ in figure 3.20 in time and the ratio of mu in time in figure 3.13, we can see that the growth rates converge. In a turbidostat the dilution rate will become equal to this final growth rate. We can conclude that this system is also robust with respect to the initial conditions.

The models for substrate dependencies are based on the Monod equation. We created models for a consortium with a single dependency of *E. coli* on a substrate *Synechocystis* produces, for a consortium with a substrate interdependency, in which *Synechocystis* is also dependent on a substrate *E. coli* produces, and for a consortium in which *Synechocystis* is growing light limited. For each of these consortia we created a model for a turbidostat environment and for batch cultures. In this handout we will show you the model we created for a substrate interdependent consortium in a turbidostat environment and we will show you the results.
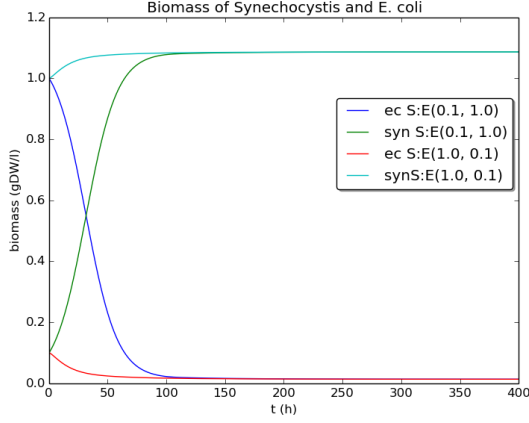
Figure 3.10: Biomass of *Synechocystis* and *E. coli* in time in a turbidostat. Even thought the total amount stays the same, the amount of each species varies in time.
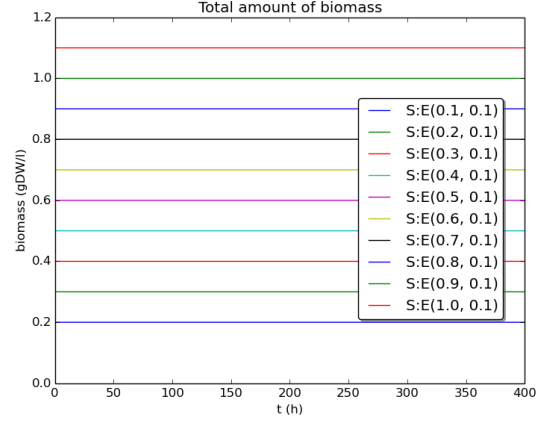
Figure 3.11: The added biomass of *Synechocystis* and *E. coli* in a turbidostat. Like expected this amount stays constant.
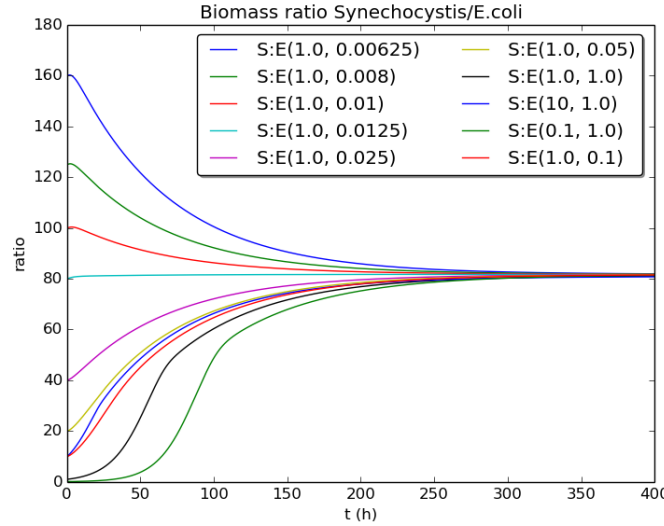

Figure 3.12: Ratio of the biomass of *Synechocystis* and *E. coli* under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$) in a turbidostat. The biomass ratio of *Synechocystis* and *E. coli* converge to the same value for each initial condition. The model is given by equations 3.7. Parameters are given by table 3.4.

3.2 Turbidostat model for a substrate interdependent consortium

If we want to model the consortium in a turbidostat, we have to account for the fact that both *Synechocystis* and E. coli are increasing the OD as they grow. This means that the dilution rate is dependent on the biomass of *Synechocystis* as well as that of E. coli. For simplicity we make the assumption that there is a constant flow through the system, instead of only diluting when the threshold is reached. To understand this we first look at the case of a single strain, called b. In a chemostat the growth rate of the organism would become equal to the dilution rate. In a turbidostat however, an organism can grow at its maximal growth rate, but the
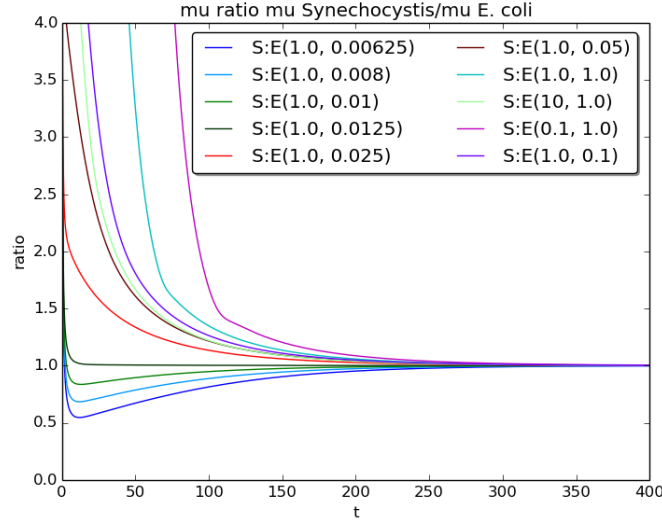
18

Figure 3.13: Ratio of μ_{syn} and μ_{ec} under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$) in a turbidostat. μ_{ec} and μ_{syn} both converge to the same value, so the ratio converges to one. The model is given by equations 3.7. Parameters are given by table 3.4.
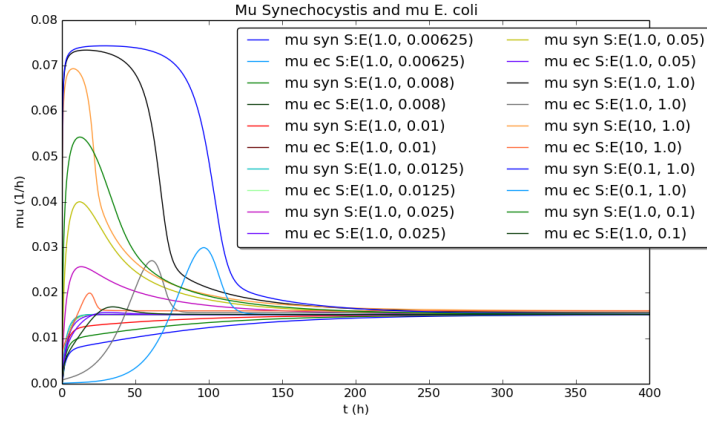

Figure 3.14: μ_{ec} and μ_{syn} in time in a turbidostat. μ_{ec} and μ_{syn} both converge to the same value, this value is equal to the dilution rate D. The model is given by equations 3.7. Parameters are given by table 3.4.

amount of biomass must still become constant. This means the following:

$$\frac{db}{dt} = \mu \cdot b - D \cdot b = 0 \tag{3.38}$$

Where b is the amount of biomass of the strain b, mu is the growth rate and D the dilution rate.

In a chemostat it would mean that $D < \mu_{max}$ is a chosen dilution rate and that μ becomes equal to D due to substrate limitation. In a turbidostat however, D becomes equal to μ_{max}, because the species in not growing limited.

In the case where there are two strains, strain a and b, that share a turbidostat, the differential

19

equations that then describe the system looks like the following:

$$\frac{da}{dt} = f(a, b, t) - Da \tag{3.39}$$

$$\frac{db}{dt} = g(a, b, t) - Db \tag{3.40}$$

Herein are f and g functions that describe the growth of the organisms a and b respectively. D is again the dilution rate. Now in a shared turbidostat it holds that $a + b = k$, where k is a constant amount of biomass. This implies:

$$D = \frac{g + f}{a + b} \tag{3.41}$$

For the turbidostat we then arrive at the following set of differential equations:

$$\frac{d\text{syn}}{dt} = \mu_{syn} \cdot \text{syn} - D \cdot \text{syn} \tag{3.42}$$

$$\frac{d\text{ec}}{dt} = \mu_{ec} \cdot \text{ec} - D \cdot \text{ec} \tag{3.43}$$

$$\frac{d[S_1]}{dt} = \frac{1}{y_{syn/S1}} \cdot \mu_{syn} \cdot \text{syn} - \frac{1}{y_{ec/S1}} \cdot \mu_{ec} \cdot \text{ec} - D \cdot [S_1] \tag{3.44}$$

$$\frac{d[S_2]}{dt} = Qp_{S2/ec} \cdot \text{ec} - \frac{1}{y_{syn/S2}} \cdot \mu_{syn} \cdot \text{syn} - D \cdot [S_2] \tag{3.45}$$

where

$$D = \frac{\mu_{syn} + \mu_{ec}}{\text{syn} + \text{ec}} \tag{3.46}$$

$$\mu_{syn} = \mu_{max,syn} \frac{[S_2]}{k_{s2,syn} + [S_2]} \tag{3.47}$$

$$\mu_{ec} = \mu_{max,ec} \frac{S_1}{k_{s1,ec} + [S_1]} \tag{3.48}$$

Herein is syn the amount of biomass of *Synechocystis* and ec the amount of biomass of *E. coli*. The μs are growth rates. μ_{max} is the maximal growth rate (equal to the growth rate at unlimited growth), and $[S]$ the concentration of substrate. k_S is the concentration of $[S]$ at a rate $\frac{1}{2}\mu_{max}$. In this case there are two μ and two μ_{max} values, one for *Synechocystis* and one for *E. coli*. There are also two substrates. $[S_1]$ can be seen as acetate. It is being produced by *Synechocystis* and consumed by *E. coli*. Its production is growth coupled. S_2 could be arginine, It is produced by *E. coli*, not in a growth coupled way, and consumed by *Synechocystis*. $y_{syn/S1}$ is the substrate yield of *Synechocystis*. Since in this model the substrate is only formed when *Synechocystis* forms biomass, per amount of biomass formed, there is a constant amount of substrate formed. The yield is usually expressed in gram dry weight mole substrate *used*. In this case we mean gram dry weight mole substrate *formed*. So to find the amount of substrate that is formed per amount of biomass that is formed we simply take $\frac{1}{y_{syn/S1}}$. $y_{ec/S1}$ and $y_{syn/S2}$ are biomass yields.

3.2.1   Parameters

To simulate the interdependency on substrates, we needed a lot more parameters. For some of the parameters we had no good measurements, so we estimated them. The general behavior stays generally the same under variations in parameters, although exact outcomes may differ.

$\mu_{max,syn}$ (h^{-1})	0.075
$\mu_{max,ec}$ (h^{-1})	0.08
$k_{s1,ec}$ (mM)	0.02
$k_{s2,syn}$ (mM)	0.02
$y_{\frac{s1}{syn}}$ (mmol $\cdot$ gDW^{-1})	0.40
$y_{\frac{ec}{s}}$ (gDW $\cdot$ mmol^{-1})	0.031
$y_{\frac{s2}{syn}}$ (mmol $\cdot$ gDW^{-1})	0.1
$Q_{p,ec}mmol$ (mmol $\cdot$ gDW^{-1} $\cdot$ h^{-1})	0.13

Table 3.4: Measured and estimated parameters for the substrate interdependent model.

3.3 Results

Here we will show you some results of the simulations of a consortium with a substrate dependent photoautotroph in a turbidostat. The model we used is described in equations 3.42 to 3.48 in section 3.2. The parameters used are given in table 3.4 In a turbidostat it is expected that the total amount of biomass, so that if *E. coli* and *Synechocystis* combined would stay constant. To test this, we have plotted the total amount of biomass under different initial conditions. The result is show in figure 3.16, and we can see that this is indeed constant.

This however, does not mean that the amount of biomass of each species stays the same. In figure 3.15 several examples of the amount of biomass in time of *Synechocystis* and *E. coli* are given.

Now we can once again look at the biomass ratio to see if this converges. In figure 3.19 the biomass ratio in time is given under different initial conditions.

We can see that indeed the biomass ratio converges. Also if we look at the values for μ in figure 3.20 in time, we can see that the growth rates converge. In a turbidostat the dilution rate will become equal to the final growth rate to which the growth rates converge. We can also conclude that this system stabilizes in time and that it is robust with respect to the initial conditions.
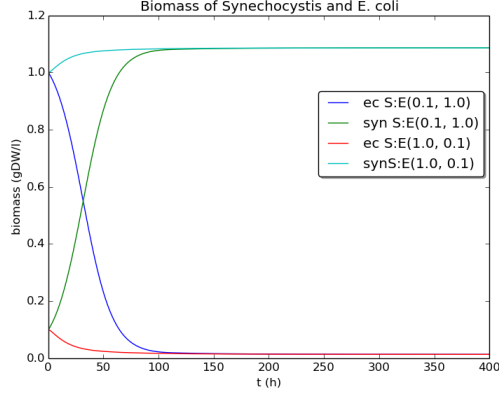
Figure 3.15: Biomass of *Synechocystis* and *E. coli* in time in a turbidostat. Even thought the total amount stays the same, the amount of each species varies in time.
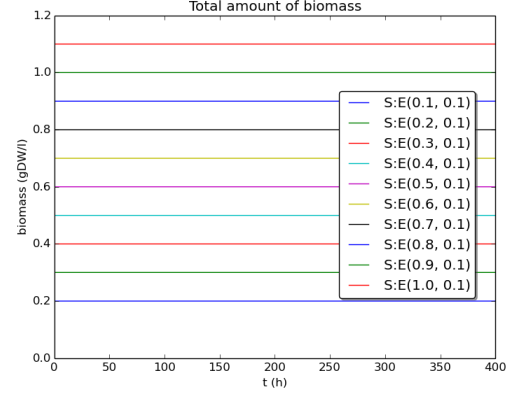

Figure 3.16: The added biomass of *Synechocystis* and *E. coli* in a turbidostat. Like expected this amount stays constant.
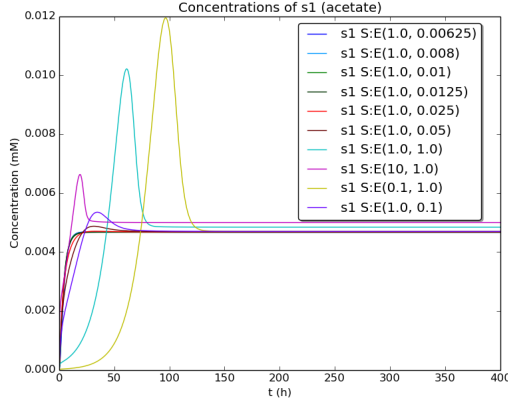

Figure 3.17: Concentration of acetate, or S_1 in the model in time under different initial conditions. After an initial spike it stabelizes.
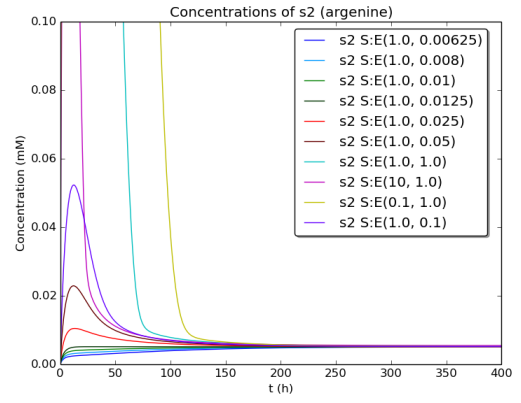

Figure 3.18: Concentration of argenine, or S_2 in the model in time under different initial conditions. After an initial spike it stabelizes
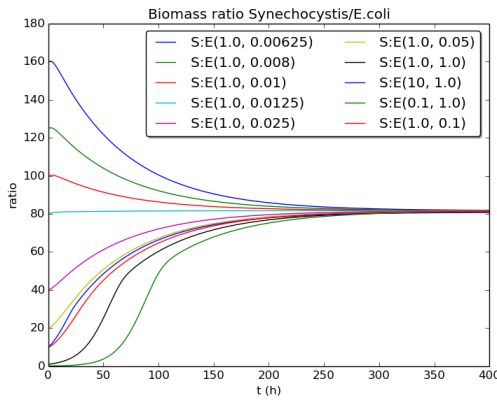

Figure 3.19: Ratio of the biomass of *Synechocystis* and *E. coli* under several different initial conditions (shown in legend in $gDW \cdot l^{-1}$) in a turbidostat. The biomass ratio of *Synechocystis* and *E. coli* converge to the same value for each initial condition. The model is given by equations 3.42 to 3.48. Parameters are given by table 3.4.
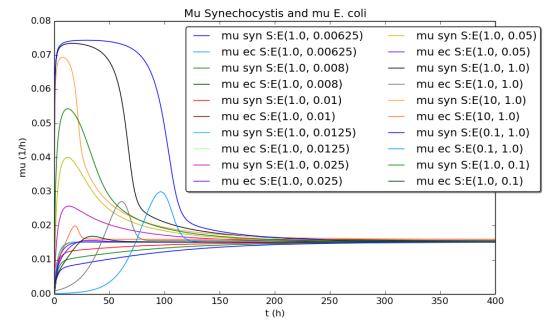

Figure 3.20: μ_{ec} and μ_{syn} in time in a turbidostat. μ_{ec} and μ_{syn} both converge to the same value, this value is equal to the dilution rate D. The model is given by equations 3.42 to 3.48. Parameters are given by table 3.4.

Abstract

3.4 Introduction

Why do we want to use kinetic models? Under what restrictions does a symbiosis converge. When is something considered stable?

3.5 Methods

type of models sets of differential equations software

3.6 Results

convergenge pictures toepassing (Stijn en Hugo) stability analysis

3.7 Discussion

Different values than Stijn's values maybe add delay to get oscillations in future research too much combinations of parameters too much parameters

Chapter 4

Discussion

Results in the light of the iGEM project Why a 2 organisms, why symbiose, complications with 2 organisms in a tank How to advance.

Chapter 5

Acknowledgement