University of West Bohemia
Faculty of Applied Sciences
Department of Cybernetics

BACHELOR THESIS

Systematic modelling and analysis of the interaction network in bacterial strains of Escherichia coli

Pilsen, 2013 Tereza Puchrová

Prohlášení

Předkládám tímto k posouzení a obhajobě bakalářskou práci zpracovanou na závěr studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

V Plzni dne 16. května 2013

. . . . . .

Declaration

I hereby declare that this bachelor thesis is completely my own work and that I used only the cited sources.

Acknowledgement

Foremost, I would like to express my sincere gratitude to my advisor M.Sc. Daniel Georgiev, PhD for the support and motivation. My sincere thanks also goes to Martin Leba who has been helping with the experiments.

Abstrakt

Bakterie Escherichia coli si vyvinula pozoruhodný autoregulační mechanismus, který reguluje proces jejího dělení. Min systém je systém tří proteinů, které v tomto procesu hrají stěžejní roli. Vzájemnou spoluprací tyto proteiny ovlivňují správnou volbu místa dělení a vyloučí tak nesymetrické rozdělení buňky tím, že donutí buňku rozdělit se přímo uprostřed. Byl navržen jednoduchý deterministický model dynamického chování Min systému a porovnán s výsledky experimentů *in vivo*. Správné pochopení skutečného chování tohoto systému je stěžejní pro budoucí užití jeho principů v buněčné regulaci.

Klíčová slova: matematické modelování, shluková analýza, regulace, buněčné dělení, E. coli, Min systém

Abstract

Escherichia coli bacterium developed a remarkable self-regulatory mechanism that regulates the process of its division. Min system is a system of three proteins that play the main part in this process. These proteins together effect the placement of division septum and prevent formation of unequal daughter cells by directing it to the centre of the cell. A simple deterministic model of the Min system dynamics was developed and validated with *in vivo* experiments. Understanding the system's dynamics is crucial for further use of its principles for regulation in cells.

Keywords: mathematical modelling, cluster analysis, regulation, cell division, E. coli, Min system

Contents

1 Introduction

Synthetic biology is a quite new and quickly spreading field bringing together biologists, chemists and engineers for the common purpose of creating functional organisms. Synthetic biological systems are mainly implemented at the DNA level, which houses the cellular program that instructs the organisms operations. Thanks to new technologies, changing present or even creating new DNA sequences has became a routine.

Common engineered systems are made up of electromechanical parts, e.g., robots, machines etc. Systems in synthetic biology are made up of various molecular species interconnected through complex interaction networks, e.g., protein interaction networks, transcriptional networks, etc. While the underlying technology is different, biological organisms (synthetic or wild types) implement many of the same systems design principles (e.g., to design signal transduction and feedback mechanisms). Hence, knowledge of systems theory is important towards designing new functional biological behaviours.

In this work, a system comprising a protein interaction network is presented. A mathematical model of its dynamical behaviour is introduced and further analysed. This model is the first step in understanding the systems regulatory principles and harnessing them for other functions.

2 Biological background

2.1 Cell division

Cell division is a process in which a parent cell divides into two daughter cells with the same genetic material. Cell division is also a way of reproduction of unicellular organisms such as bacteria and yeast. Binary fission is the simplest type of cell division. It is a form of asexual reproduction typical for prokaryotes. The whole process of division includes three steps: DNA replication, chromosome segregation and cytokinesis.

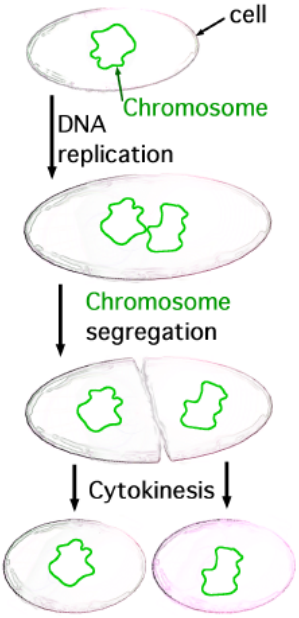

Figure 2.1: During the binary fission, a cell duplicates its genetic information (chromosome) and divides it between two newly-formed daughter cells. Proteins are mainly separated at random. Source: http://simple.wikipedia.org/wiki/Binary_fission

2.1.1 Binary fission

- **DNA replication:** In this process, a copy of parent DNA is made. The double helix structure of DNA is unwound and each of the two strands serves as a templet for synthesis of the new complemental strand. The result is two double helix-shaped DNA molecules (chromosomes) which are identical and so the genetic information is preserved.
- **Segregation:** The two chromosomes are separated into the opposite cell halves so that the cell could, ideally, start dividing between them and so each daughter cell obtains one chromosome.
- **Cytokinesis:** Cytokinesis is a complex process driven by the whole mechanism of proteins and genes. It starts with the selection of a division site which is determined by formation of Z ring. Z ring is a ring-shaped assembly of FtsZ proteins on the cell membrane, it starts getting smaller and smaller and forms a septum that divides the cell by constricting the cell membrane.

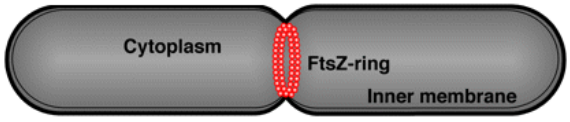

Figure 2.2: FtsZ ring is formed on the cell membrane at the centre of the cell and division is initiated. The constriction of Z ring divides the cell in two. Placement of the FtsZ ring is guided by the Min system. Source: [1]

2.1.2 Septum placement

Division placement is the key part that effects whether the cell divides equally or not. Precise division is important for the viability of daughter cells because unequal division may lead to minicelling, production of daughter cells through asymmetric septum placement generating wrong chromosome copy numbers and leading to the inability to replicate. To avoid minicelling the cell must divide at it's center between the two segregated chromosomes. How a cell directs the division right to the center of its long axis is a task for a system

of proteins called the Min system and investigating it is the main purpose of this paper.

2.2 The Min system in E. coli

E. coli has developed a self-regulatory system that ensures equal division by situating septum at the centre of the cell. The system consists of three proteins: MinE, MinD, and MinC, which cooperate all together and coordinate the FtsZ assembly and further equilibration of Min proteins in both halves of the dividing cell.

2.2.1 The Min system components

MinD is located in the cytoplasm in both the phosphorylated (MinD-ATP) and dephosphorylated (MinD-ADP) forms. It binds to the cell membrane at the cell poles, but only in the phosphorylated form. MinD proteins bound to the membrane and create a layer that has a shape of a test tube at the pole.

MinE proteins bind to the membrane bound MinDs and form MinDE complex. Once MinE is bound, MinDE complex dissociates from the membrane and MinD-ADP and MinE is released to cytoplasm where MinD-ADP is phosphorylated again. MinE proteins binding to MinDs on the edge of the MinD tube form ring-like structure that moves towards the cell pole releasing MinDs on the way. The released MinDs diffuse towards the opposite cell pole and again bind to the membrane. The whole process then repeats generating a spatial-temporal oscillation pattern alternating between the cell poles.

MinC is not really required for the oscillation to appear [2], it just binds to MinD and oscillates with it. But this protein actually is the one which affects Z ring formation and thus the location of the division site.

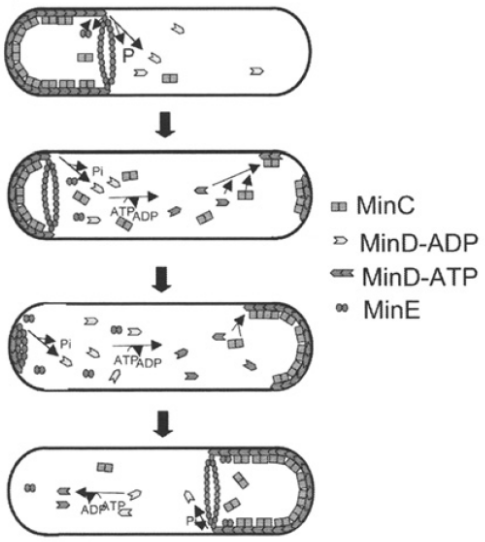

Figure 2.3: MinD proteins bind repetitively at the opposite cell poles followed by MinE and MinC proteins. This oscillating behaviour of Min proteins along the cell's long axis is required for equal division. Source: [3]

2.2.2 The effect on division placement

The role of MinC is that it inhibits assembly of FtsZ and so the formation of the Z ring. MinC concentrates at the cell poles because of its binding to MinD. This means MinC is only present at the poles and is absent in the centre, where FtsZ can assemble. This fact directs the division site right to the centre of the cell and prevents potential division sites at the poles.

2.2.3 Spatial-temporal behaviour

The Min system shows periodic oscillations in non-dividing cells but the periodicity is corrupted once the cell starts to divide. In dividing cell, the shape of the narrow membrane region near the septum resembles the pole regions and is also a new location where MinD proteins begin to aggregate. Irregular oscillation with mid-cell pausing ensue. In certain stage of constriction, the oscillations split and independent oscillations appear in both halves of

nearly divided cell. In this phase, levels of Min proteins on both sides of the constricting septum equilibrate. The result is that both daughters obtain an equal amount of Min proteins and the oscillations persist [4], [5].

This is an example of developed self-regulation which prevents possibility of one daughter cell obtaining an insufficient amount of Min proteins disrupting their ability to oscillate from end-to-end.

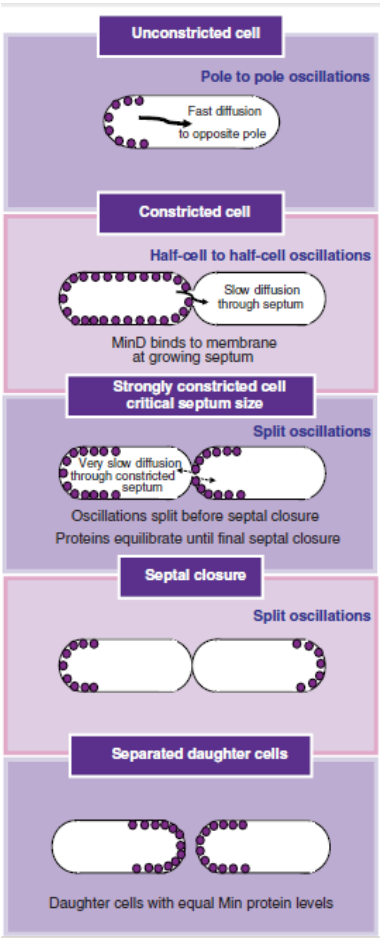

Figure 2.4: The oscillating pattern changes through the phases of the cell cycle. Source: [5]

2.3 Experimental systems

For the purpose of this work, for investigating Min system in vivo, genetically modified E. coli strain WM1264[1] was used. It is a strain with a chromosomaly integrated lac operator driving expression of GFP-MinD + MinE (GFP corresponds with green fluorescent protein).

2.3.1 Green fluorescent protein

GFP is a protein that fluoresces green when illuminated by blue light. It is often used for protein visualization. Fluorescent proteins fused to proteins of interest enable measuring relative amount of these proteins by measuring the fluorescence intensity.

The bacterial chromosome of WM1264 strain has been modified for the purpose of observing Min proteins through GFP fluorescence. The cell synthesizes MinD proteins fused to GFPs allowing the motion of MinD to be monitored.

2.3.2 The lac operon and IPTG induction

The amount of MinD can not only be monitored but also changed. Common way of changing protein levels in synthetic biology is to use lac operator. The lac operator is a gene sequence taken from the lac operon (a collection of native bacterial genes), where it plays a key role in controlling E. coli's lactose metabolism. When lactose is absent there is no need to produce the enzyme that lyses lactose so the transcription of the lac operon is inhibited by a regulatory protein binding to the lac operator. In presence of lactose, the regulatory protein is released and transcription is reactivated.

This mechanism can be influenced using IPTG (isopropyl β-D-1-thiogalactopyranoside), a small molecule that mimics lactose and activates transcription repressed using lac operator. Adding IPTG to a cell culture can be used to turn on expression of any protein of interest. In the WM1264 chromosome, the expression of GFP-MinD and MinE is controlled by the lac oper-

[1] obtained thanks to kind cooperation of prof. William Margolin, Ph.D., Department of Microbiology & Molecular Genetics, University of Texas-Houston Medical School

ator. Thereby, the concentration of GFP-MinD (and MinE) increases with increased concentrations of IPTG in the nutritional medium.

2.3.3 Fluorescence microscopy

Thanks to its fusion with GFP, monitoring of MinD proteins in cells is possible using fluorescence microscopy. According to known molecular excitation and emission parameters, specific filters are appended to standard light microscopes and used for illuminating and later observing the motion and concentrations of the fluorescent proteins. When illuminated with blue light, GFP radiates green light.

Photobleaching

Repeated illumination of the fluorescent proteins leads to a photobleaching effect that must be considered when interpreting the results. Photobleaching is the degradation of fluorescent proteins after the light exposure decreasing the fluorescence intensity for a fixed protein concentration.

2.3.4 Time-lapse microscopy

It is used mostly for long-term live cell imaging. Through repeated image capture at equidistant time intervals system dynamic behavior is visualized and quantified. Various techniques must be used to optimize the exposure time to minimize photobleaching and to extend the effective time duration.

In observing MinD oscillations, the sampling frequency must be sufficiently high relative to the oscillation period limiting the overall visualization period.

3 Theory

3.1 Biological models

For creating a model of Min system's behaviour, both biological and mathematical theory must be combined. Precise description of the biological system is essential and a mathematical model depends on it directly.

Min system can be described by chemical reactions in which six species are considered:
d_p.....phosphorylated MinD-ATP
d.....dephosphorylated MinD-ADP
d_m.....MinD bound to the cell membrane
d_e.....MinDE complex
e.....MinE
m.....cell membrane

The basics of Min system's behaviour is described by following reactions frequently described in the literature [6].

Cytoplasmic MinD-ATP binds to the cell membrane:

$$d_p + m \xrightarrow{k_m} d_m$$

Cytoplasmic MinE binds to membrane-bound MinD and creates MinDE complex:

$$d_m + e \xrightarrow{k_e} d_e$$

MinDE complex dissociates and releases MinE and MinD-ADP into cytoplasm:

$$d_e \xrightarrow{k_r} d + m + e$$

MinD-ADP is in cytoplasm again phosphorylated:

$$d \xrightarrow{k_p} d_p$$

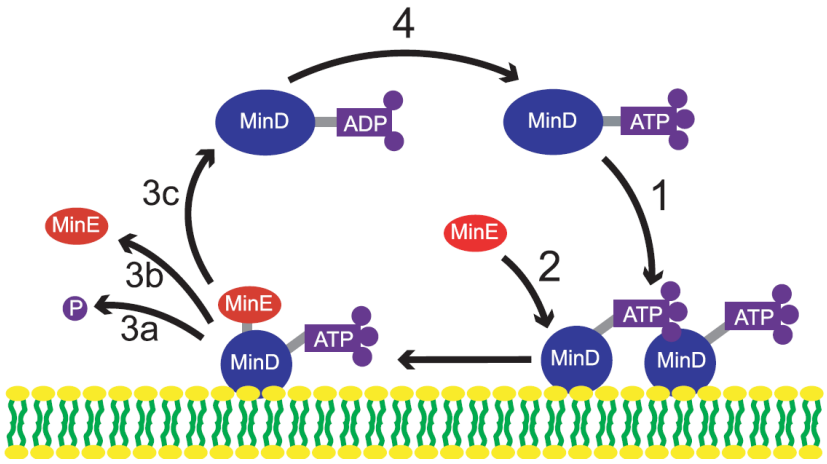

Figure 3.1: The cycle of reactions of Min system consists of MinD-ATP binding to the membrane (1), formation of MinDE comlplex (2) and its dissociation (3), and repeated phosphorylation od MinD-ADP (4). Source: [6]

3.2 Mathematical models

3.2.1 Notation

Vectors are denoted by lower case and matrices by capital letters, both in parenthesis. Sets are in curly braces, ranges and closed intervals in square brackets, open intervals in parenthesis. $\mathbb{R}^n$ denotes n-dimensional set of real numbers.

3.2.2 Mass action kinetics

The mathematical model was constructed from the chemical processes that actually take place in the cell. Chemical reactions can be transformed into a mathematical model using the Law of mass action.

Mass action kinetics are used to describe dynamic behaviour of reaction networks. Systems of differential equations describe how the concentrations

of reaction species change with time. The basic rule for applying the mass action law is as follows:

With reaction

$$A + B \xrightarrow{k} C \tag{3.1}$$

correspond ordinary differential equations, that represent changes in concentration of A, B and C in time:

$$\begin{aligned} \frac{dC}{dt} &= kAB \\ \frac{dA}{dt} = \frac{dB}{dt} &= -kC, \end{aligned}$$

where k is reaction rate. The resulting solutions always remain nonnegative and obey mass conservation laws.

3.2.3 Mathematical model of Min system

Application of the mass action kinetics gives a mathematical model for the Min system:

$$\begin{aligned} \dot{d_p} &= k_p d - k_m d_p m \\ \dot{d_m} &= k_m d_p m - k_e d_m e \\ \dot{d_e} &= k_e d_m e - k_r d_e \\ \dot{d} &= k_r d_e - k_p d \\ \dot{m} &= k_r d_e - k_m d_p m \\ \dot{e} &= k_r d_e - k_e d_m e \end{aligned}$$

If the total concentrations associated with the complexes of the different species are considered, the mass conservation law can be applied to reduce the number of equations.

Assuming d_T, e_T, m_T are total concentrations of MinD, MinE, and places on membrane respectively, the conservation laws are the following:

$$\begin{aligned} d_T &= d + d_m + d_e + d_p \\ m_T &= m + d_m + d_e \\ e_T &= e + d_e \end{aligned}$$

Choosing the states to be d, m, e, the other variables d_p, d_m, and d_e can be eliminated using the following relations:

$$\begin{pmatrix} d_p \\ d_m \\ d_e \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} d_T - d \\ m_T - m \\ e_T - e \end{pmatrix}$$

A system of only three nonlinear equations is obtained, with the state vector $x = (d, m, e)$ and the parameter vector $p = (k_p, k_m, k_r, k_e, e_T, d_T, m_T)$:

$$\dot{x} = f(x) : \tag{3.2}$$

$$\begin{aligned} \dot{d} &= k_r(e_T - e) - k_p d \\ \dot{m} &= k_r(e_T - e) - k_m m(d_T - d + m - m_T) \\ \dot{e} &= k_r(e_T - e) - k_e e(m_T - m + e - e_T) \end{aligned}$$

3.2.4 Nonlinear system analysis

This model comprises a single compartment based on the previous description of the Min protein interactions. It is expected that the concentrations of reaction species will change periodically with time. The binding and unbinding of Min proteins from membrane results in periodical filling and vacating of the membrane spaces. Not all parameter combinations, however, generate such cyclical behaviour. Parameter values, i.e., the reaction rates k_p, k_m, k_r, k_e and total concentrations d_T, m_T, e_T, for which the equation 3.2 has an oscillating solution need to be found.

The described behaviour can be achieved by finding a system with a stable limit cycle [7]:

Definition 1. *Limit cycle is an isolated periodic orbit in the phase space manifested by periodic oscillations in time. Limit cycle with property that all trajectories in the vicinity of the limit cycle ultimately tend toward the limit cycle as $t \to \infty$ is classically known as stable limit cycle.*

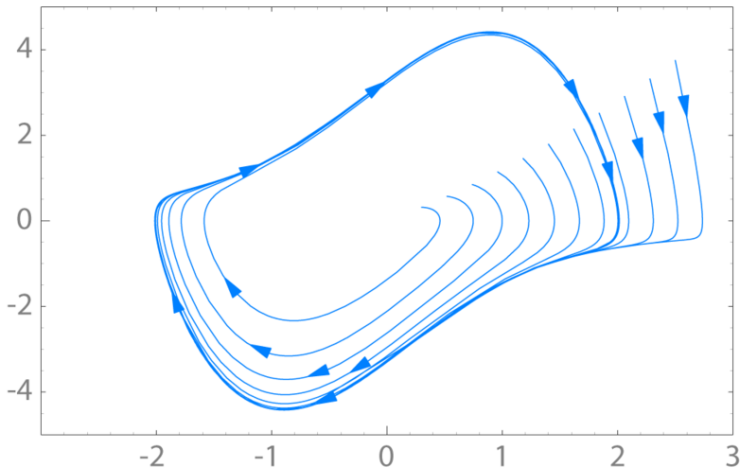

Figure 3.2: All the system trajectories tend towards stable limit cycle. Source: http://en.wikipedia.org/wiki/File:VanDerPolPhaseSpace.png

Such system can be found simply by using the sufficient condition for existence of a stable limit cycle:

Lemma 1. ***(Poincaré-Bendixson Criterion)*** *Consider second-order autonomous system $\dot{x} = f(x)$ and let M be a closed bounded subset of the plane such that*

- *M contains no equilibrium points*
- *Every trajectory starting in M stays in M for all future time.*

Then, M contains a periodic orbit of the system.

Note, the theorem also applies for regions M that contain unstable and isolated equilibrium points. Simply a sufficiently small ball is drawn around each equilibrium point and the interior region is removed from M. For higher order systems, greater than 2, the theorem does not apply due to possible chaotic-like behaviours. Below, we apply the criterion to a third order system assuming that such behaviours do not occur.

First step thus is to find all the system's equilibria for some parameter vector p. Letting

$$\dot{x} = 0$$

this requires solving a system of three nonlinear algebraic equations:

$$\begin{aligned} k_r(e_T - e) - k_p d &= 0 \\ k_r(e_T - e) - k_m m(d_T - d + m - m_T) &= 0 \\ k_r(e_T - e) - k_e e(m_T - m + e - e_T) &= 0 \end{aligned}$$

The process of solving this problem is described in Section 4.2.1.

Stability of the equilibria is determined through linearisation. The nonlinear function $f(x)$ is linearised about the equilibrium x_e giving the following linearised system's approximation:

$$\dot{x} = f(x) \approx Ax, \text{ where } A = \left.\frac{\partial f}{\partial x}\right|_{x=x_e}$$

A is called the Jacobian matrix.

The equilibrium x_e is unstable if $Re\{\lambda_i\} > 0$ for at least one eigenvalue $\{\lambda_i\}$ of A [7].

Parameter values for which all nonnegative equilibria are unstable are deemed admissible and retained for further analysis.

3.3 Parameter space analysis

In finding possible system representations, both the physical system properties and the admissible parameter clustering are considered. The areas of the parameter space where the system more robustly satisfies the conditions described in Section 3.2.4 are more likely to include the actual parameter values. If the seven parameters represent a point in a $\mathbb{R}^7$ space, this would appear as an area with higher density of admissible points.

Problem of analysing a huge amount of data is called data mining. Data mining is the computational process of discovering patterns in large data sets often involving methods at the intersection of artificial intelligence, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use [8].

One type of data mining is clustering, or cluster analysis, which is the task of grouping a set of objects in such a way that objects in the same group

(called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters) [8]. There are many clustering methods which corresponds to various notions of a cluster. Further, clustering on the basis of distances is introduced.

When representing these object by points in $\mathbb{R}^n$, cluster is a group of points that are close to each other, closer than to other points in the data set. Clustering algorithm goes point by point through the data set and computes the relative distances to other points in the set. Sufficiently close points are called neighbours. On the basis of distances the data are divided into clusters, some point also may not belong to any of them. Each point should belong to one cluster only. A cluster thus represents data with some common attributes that are specific for each cluster.

Generally, clustering methods can be divided into two classes:

- Methods for predetermined number of groups
- Methods without predetermined number of groups

The clusters are identified with the areas of parameter space described at the beginning of this section. The goal is to find at least one cluster of points among the earlier generated data that will represent a system with robust oscillating behaviour. It is not known whether such a cluster exists or what is the maximum number of clusters. It means some method without predetermined number of clusters should be used for this purpose.

3.4 Experimental validation

Validation of the model base in investigating the behaviour of MinD oscillations from these three aspects:

- Period
- Magnitude
- Shape

Investigating the **period** of oscillation should reveal in which circumstances the process of binding and unbinding of MinD to membrane accelerates or decelerates, whether its very high or low concentration can cause lost of the ability to oscillate between the cell poles. Varying **magnitude** indicates that the relative amount of MinD proteins that actually bind the membrane is also dependent on its concentration in cell. Finally, the actual **shape** of the oscillation testifies about the character of binding and unbinding of the membrane. It shows whether all the proteins quickly bind first, and then more slowly dissociate into cytoplasm or vice versa.

While using the bacterial strain WM1264 for the experiments, it is not possible to influence transcription of MinD and MinE independently. As a result, the model can only be validated with respect to equal changes in the concentrations of MinD and MinE.

Experiments with cell cultures of four different IPTG inductions are performed (the mechanism of induction is described in Section 2.3.2). As many fluorescent pictures as possible are taken during time-lapse observations, but the number of pictures suitable for further analysis is limited by photobleaching. At least 20-30 cells from each induced culture should be analysed to obtain statistically significant data sets.

Using special software, the fluorescent images are analysed. Fluorescence intensity is measured for both cell halves and also for the bacteria as a unit.

4 Results

The basis is to randomly generate vectors of parameters and then decide whether the system with these specific parameters shows oscillations. First, the ranges for each of the parameters values was established. Values that had been already published in articles that present a deterministic model of Min system [5] were followed . The published values for the reaction rates are the following:

$$\begin{aligned} k_p &= 0.5\ \mathrm{s}^{-1} \\ k_m &= 0.0125\ \mathrm{mm\ s}^{-1} \\ k_e &= 5.56 \cdot 10^7\ \mathrm{M}^{-1}\ \mathrm{mm\ s}^{-1} \\ k_r &= 0.7\ \mathrm{s}^{-1} \end{aligned}$$

If these are converted into concentrations of molecules per a cell volume, the values approximately belong into the following intervals:

$$\begin{aligned} k_p, k_m, k_e, k_r &\in (0, 10) \\ d_T, m_T, e_T &\in (0, 100) \end{aligned}$$

All the following procedures build on parameters generated in these ranges.

4.1 Biological and mathematical modelling

After generating the first set of random parameter values, none of them complied with system with a stable limit cycle. The system exhibited only stable equilibria. An example of the time response is shown in Figure 4.1.

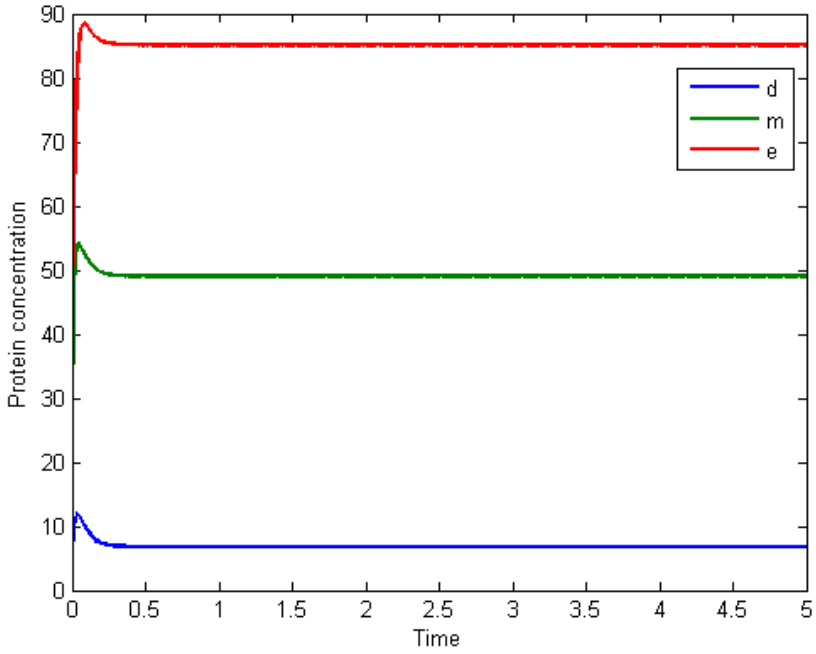

Figure 4.1: The time response of the simple Min model showing the species' concentrations settling to the stable equilibrium.

It was clear that the four basic reactions do not describe the system completely. Hence, it was necessary to include some extra reactions that may take place in the cell and give rise to the oscillations. The model was further modified with respect to real Min system's behaviour.

4.1.1 Modified biochemical network

First, the fact of cooperative binding was included. It means an increased affinity of MinD molecules to bind to membrane when some MinDs are already bound there. This effect of cooperative binding is mentioned in literature [9]. The principal is seen in Figure 4.2 A.

$$d_p + m + 3d_m \xrightarrow{k_m} 4d_m$$

Second, the possibility of MinE proteins "hopping" from a state of MinDE complex right to another membrane-bound MinD which releases MinD-ADP

and unblocks a place on membrane (Figure 4.2 B) was considered. It might happen during the movement of MinE ring towards the pole, as illustrates Figure 2.3, and is also discussed in recent in vitro studies of the Min system [10].

$$d_e + d_m \xrightarrow{k_r} d_e + d + m$$

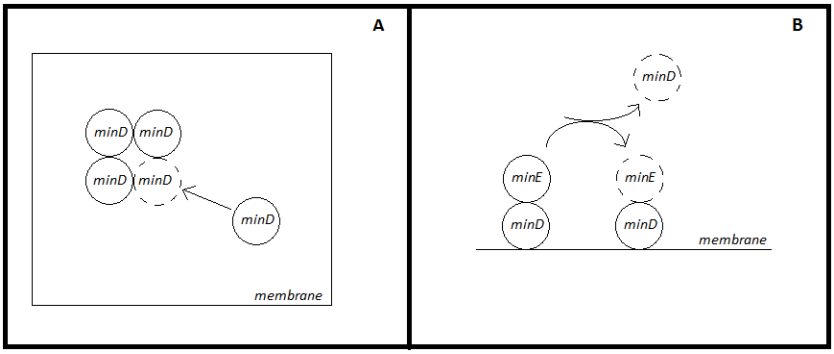

Figure 4.2: Schematic of two added reactions.

4.1.2 Corresponding mass action model

No extra species were added, so the number of variables in the system persists. The new differential equations describing the evolution of d and m are as follows:

$$\begin{aligned} \dot{d} &= k_r d_e - k_p d + k_r d_e d_m \\ \dot{m} &= k_r d_e - k_m d_p m(1 + d_m^3) + k_r d_e d_m \\ \dot{e} &= k_r d_e - k_e d_m e \end{aligned}$$

All the following procedures build on this model.

4.2 Parameter classification

A challenging task was to generate a sufficiently extensive data set of admissible parameter values. In evaluating each parameter value, for a randomly given p, the first step was to compute the system's equilibria.

4.2.1 Equilibria

Procedure of evaluating the system's equilibria needs to be as quick as possible due to the large amount of parameter values that need to be examined (10^8). The Symbolic Toolbox was used to transform the three algebraic equations into a single polynomial equation for e. The roots of the resulting 7th order polynomial could be easily and efficiently computed and transformed using MATLAB.

Real, negative, and complex roots were commonly encountered. Since the examined equations represent physical system, only nonnegative equilibria are permissible and nonnegativity is required not only for the state vector $(d,\ m,\ e)$ but also for the other species $d_p,\ d_m$ and d_e. All these conditions reduced the 7 equilibria to at most one real positive equilibrium for every random parameter vector. In some cases, no real positive equilibrium was found. While this is in general permissible, after further investigation, the absence of a real positive equilibrium was attributed to numerical error where repeated positive roots included a small imaginary part ($\ll 1$). In all such cases, these equilibria were found to be stable and, hence, the associated parameter values were deemed inadmissible.

4.2.2 Eigenvalues

The case of multiple real positive equilibria did not occur so the existence of stable limit cycle for a specific parameter value depended on the stability of the only real positive equilibrium found. Using the linearisation method described in 3.2.4 showed that 2% of the randomly generated parameter values correspond to a stable limit cycle. These can be further separated in two classes based on the Jacobian eigenvalues:

- Class 1 (5%): 2 real positive + 1 real negative eigenvalue
- Class 2 (95%): 2 complex with positive real parts + 1 real negative eigenvalue

Dependence between the eigenvalues and the oscillation characteristics was explored. Intuition suggests that eigenvalues with a greater real part λ_{max} correspond to a more unstable system leading to greater oscillation magnitudes. After analysing several sample values, however, the data showed

no such dependence. In addition to λ_{max}, other eigenvalue functions were explored:

- $$v_D = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix},$$

  where λ_1 and λ_2 are real positive eigenvalues and v_1 and v_2 are corresponding eigenvectors.

- $$v_{Dmax} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T v_{max}\lambda_{max},$$

  where λ_{max} is the maximum eigenvalue and v_{max} is corresponding eigenvector.

No dependence was found between the oscillation period or magnitude and any of the above eigenvalue functions.

One thing that this analysis did reveal was that the system with only real eigenvalues shows significantly grater oscillations than the system with complex eigenvalues.

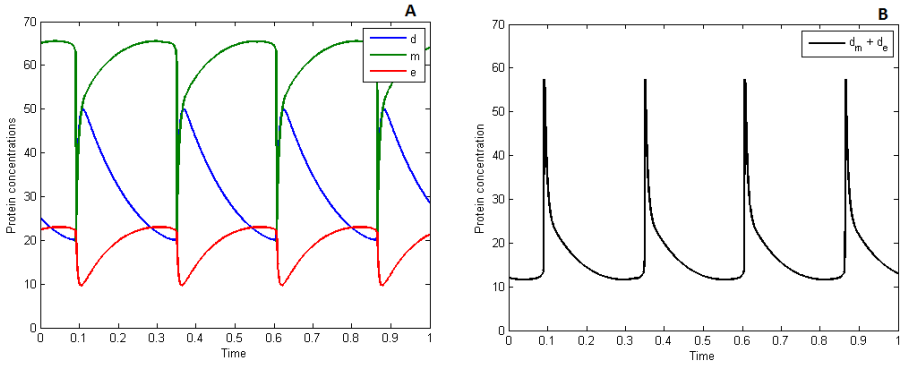

Figure 4.3: Parameters that correspond with only real eigenvalues show significantly grater oscillations (**A**). Figure **B** shows oscillation of filled membrane.

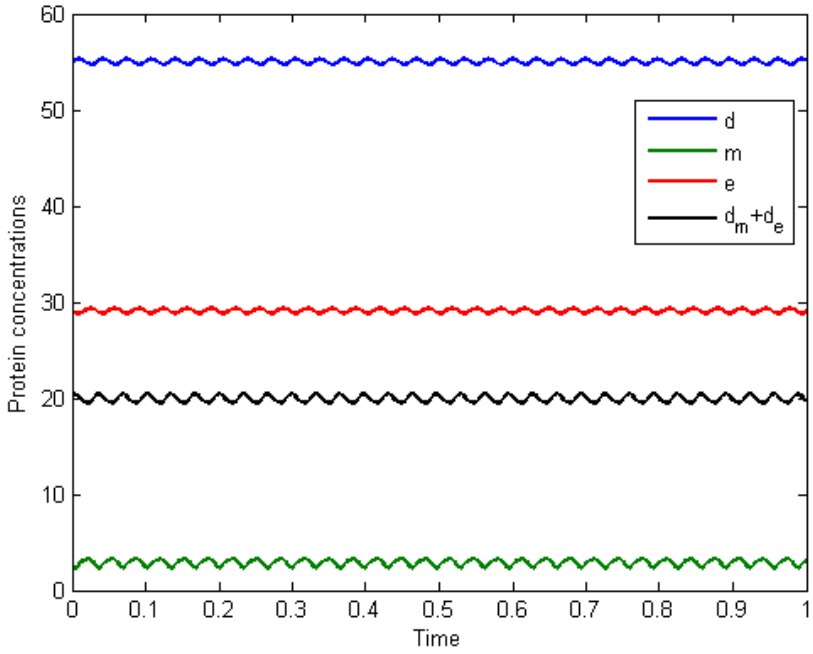

Figure 4.4: A typical example of oscillations for system with only complex eigenvalues with a positive real part. Magnitudes are minimal.

According to these results it was decided to work further only with parameters that represent systems with only real eigenvalues to ensure greater average oscillation amplitudes similar to those seen in the real system. This will guarantee at least greater oscillations, on average.

4.3 Cluster analysis

For the purpose of defining clusters, a density-based method DBSCAN was implemented in MATLAB.

4.3.1 The DBSCAN method

DBSCAN defines separate areas with higher density than the rest of the data set. The less dense points are considered as noise.

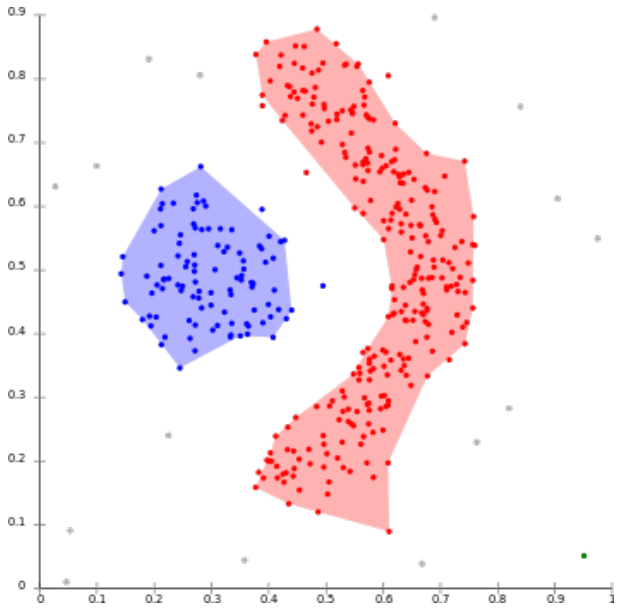

Figure 4.5: DBSCAN can find non-linearly separable clusters. Source: http://en.wikipedia.org/wiki/File:DBSCAN-density-data.svg

The method has two parameters. First, the minimal distance between two points at which these points are considered neighbours (*eps*). Second, the minimal number of points ($MinPts$) that is necessary for starting a cluster formation. It means each point in a cluster (except the border points) has at least $MinPts$ points in a radius of *eps*. The final number and size of clusters depends on both parameters.

Pseudocode

(Source: http://en.wikipedia.org/wiki/DBSCAN)

```
DBSCAN(D, eps, MinPts)
   C = 0
   for each unvisited point P in dataset D
      mark P as visited
      NeighborPts = regionQuery(P, eps)
      if sizeof(NeighborPts) < MinPts
         mark P as NOISE
      else
         C = next cluster
         expandCluster(P, NeighborPts, C, eps, MinPts)

expandCluster(P, NeighborPts, C, eps, MinPts)
   add P to cluster C
   for each point P' in NeighborPts
      if P' is not visited
         mark P' as visited
         NeighborPts' = regionQuery(P', eps)
         if sizeof(NeighborPts') >= MinPts
            NeighborPts = NeighborPts joined with NeighborPts'
      if P' is not yet member of any cluster
         add P' to cluster C

regionQuery(P, eps)
   return all points within P's eps-neighborhood
```

The algorithm of clustering is a significantly time and memory demanding procedure. DBSCAN visits all points in dataset multiple times. In addition, it needs to keep the set of neighbours, which is a dynamic array.

The data set had approximately 120,000 points, each a vector in $\mathbb{R}^7$. Distances between two points p and q are given by the Euclidean distance

$$d(p,q) = \sqrt{\sum_{i=1}^{7}(q_i - p_i)^2}.$$

The algorithm is written to work with a distance matrix whose a_{ij} entry represents the distance between p_i and p_j. This requires memory $O(n^2)$, which

is excessive given $n = 120,000$. As a result, distances had to be recomputed for each point while visiting it, slowing down the algorithm significantly.

On the other hand, constant lengthening of neighbours array was avoided. The algorithm was built on $n \times 3$ array where each row stands for one point of the dataset. The first column states whether this point was visited or not. The second column states the cluster the point belongs to. The third column states whether the point is in the set of neighbours right now. To each point thus belongs a vector of information:

$$v = (v_1, v_2, v_3), \text{ where}$$

$$v_1 = \begin{cases} 1 & \text{visited} \\ 0 & \neg\text{visited} \end{cases}$$

$$v_2 = c,\ c = 1, ...r, \text{ where } c \text{ is cluster ID}$$

$$v_3 = \begin{cases} 1 & \text{P} \in \text{NeighborPts} \\ 0 & \text{P} \notin \text{NeighborPts} \end{cases}$$

A simple testing data set in $\mathbb{R}^2$ (shown in Figure 4.6) was created for debugging the DBSCAN algorithm.

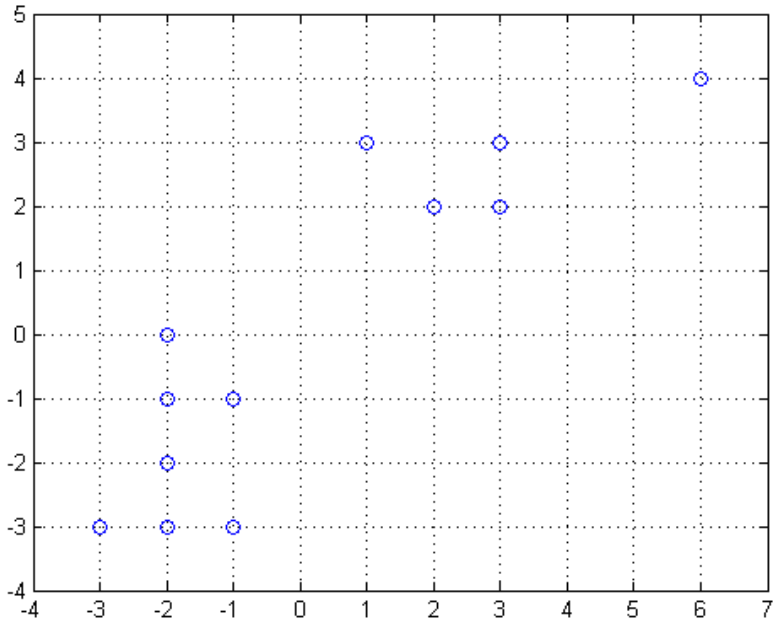

Figure 4.6: Visualization of the testing data set including 12 points which form two clearly separable clusters.

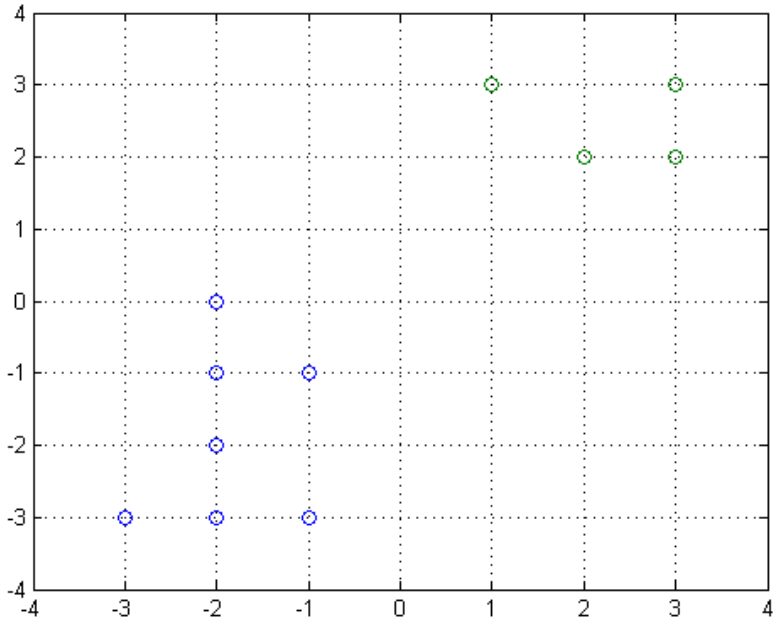

Figure 4.7: The result of DBSCAN algorithm ran with data shown in Figure 4.6. Two clusters shown in different colours are clearly separable and the point $[6, 4]$ is considered as noise. The corresponding parameters for the algorithm are $eps = \sqrt{2}$ and $MinPts = 3$

4.3.2 Cluster visualization

The method of trial and error was used for finding the suitable values for *eps* and $MinPts$. Some balance between these two parameters had to be found so that the result is one significantly localized cluster and maybe few smaller ones.

The parameters for DBSCAN that were considered most suitable are

$$\begin{aligned} eps &= 3 \\ MinPts &= 15 \end{aligned}$$

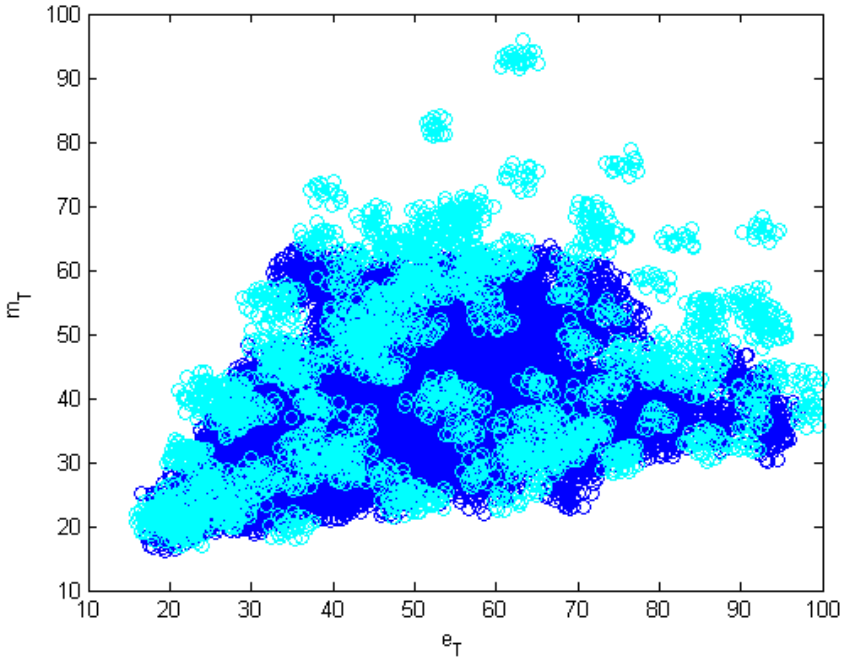

Figure 4.8: Number and size of detected clusters can be quickly identified, but some visualisation is rather difficult. Clusters in $\mathbb{R}^7$ are impossible to imagine or visualize completely. Only $\mathbb{R}^2$ projections can be visualised. This is a projection of identified clusters in e_T and m_T. Corresponding DBSCAN parameters are $eps = 3$ and $MinPts = 15$. The largest cluster contains 20,389 points from data set of 120,256.

The largest cluster is taken to be the set of potential parameter values. This cluster is further examined in the following section.

4.3.3 Cluster labelling

To better understand cluster compositions, the data points were labelled according to corresponding system properties, e.g., the functions defined in Section 4.2.2 ($\lambda_{max}, v_D, v_{Dmax}$).

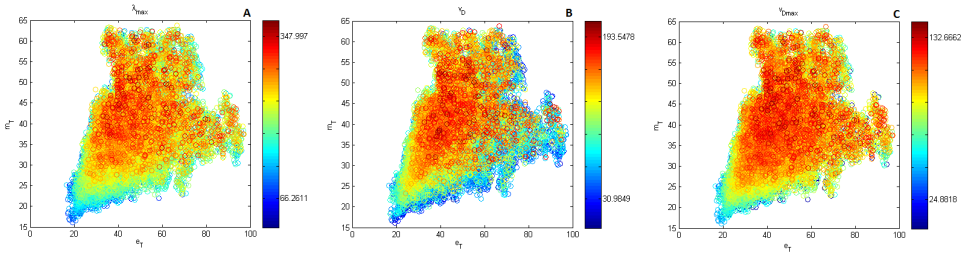

Figure 4.9: Values corresponding to eigenvalue functions were assigned to the points in the major cluster. In all three pictures, an area of most unstable parameters is defined. The most convincing is probably picture C where the points, labelled according to v_{Dmax}, are the most dense.

The cluster points already satisfy the desired system properties. In order to better understand the changes in behaviour caused by changes in parameter values, the resulting oscillations were further characterized for the points in the cluster. For the purpose of obtaining representative amount of data, brute force was used to further characterize clusters. A representation of points was selected at random and simulated (approx. 10% of the cluster points). The relative magnitude and period of oscillations in the time response was subsequently numerically calculated. Monitoring the amount of actually filled membrane is desired. It means, maxima and minima of $d_m + d_e$ were detected and relative magnitude thus refers to the value:

$$m_r = \frac{m_{max} - m_{min}}{\frac{1}{2}(m_{max} + m_{min})},$$

where m_{max} and m_{min} stands for maximum and minimum of $d_m + d_e$.

Results show that period T and magnitude m_r of points in the cluster belong to intervals:

$$\begin{aligned} T &= [0.036, 0.85] \\ m_r &= [0.2, 1.6] \end{aligned}$$

Labelling the whole cluster with either period or magnitude, however, showed no interpretable result. The dependence on period was found only after more specified investigation.

Dependence of period and magnitude on relation between total concentrations d_T and e_T, and d_T and m_T was also investigated. A point from the

centre of the major cluster was selected. For variable d_T and e_T, respective d_T and m_T, the other five parameters were fixed and new data set was generated and further clustered. For points of the major cluster found, periods and magnitudes were again calculated.

The result of period dependence is shown in Figure 4.10 and 4.11 but no dependence of magnitude was revealed. Magnitude of points in cluster for varying d_T and e_T is rather constant and is 0.5, for varying d_T and m_T it is in range from 0.04 to 0.6.

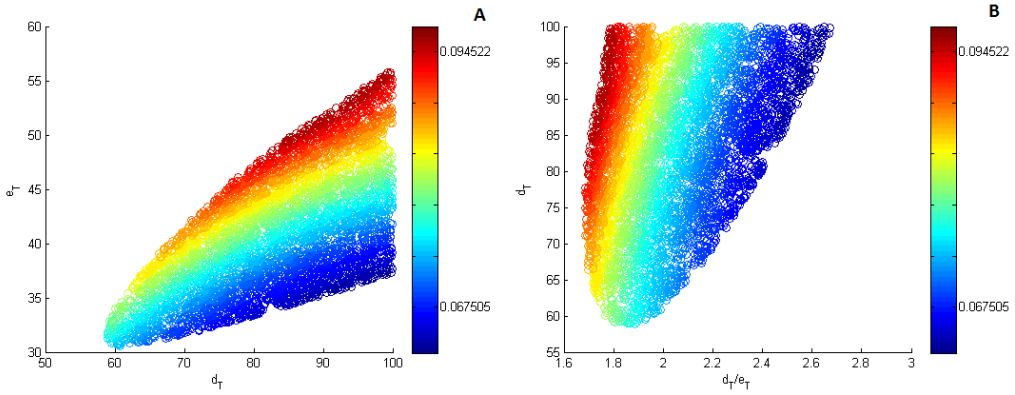

Figure 4.10: Visualisation of a cluster found for fixed parameters k_p, k_m, k_e, k_r, m_T. From **A** is quite clear that period is proportional to MinE and inversely proportional to MinD concentration. **B** just better demonstrates that considering fixed MinD to MinE ratio, period increases with increasing MinD concentration.

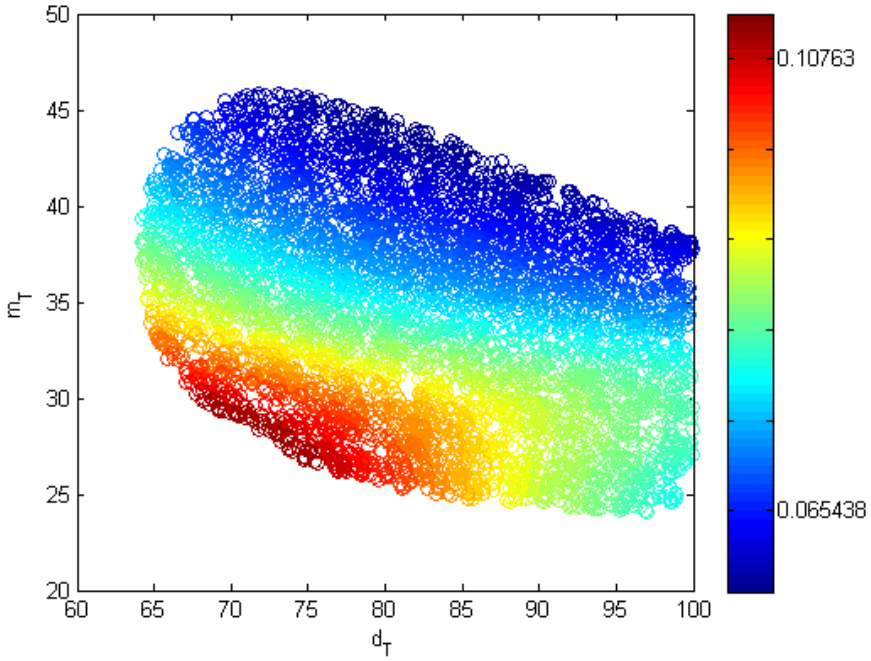

Figure 4.11: Visualisation of a cluster found for fixed parameters k_p, k_m, k_e, k_r, e_T. It is important to understand, that decreasing concentration of membrane places corresponds to increasing cell volume. Thus, the figure shows that period is slower in larger cells, according to this figure.

4.4 Experimental validation

Approximately 30 cells were analysed for each IPTG induction. Following table shows obtained results:

IPTG concentration [uM]	Period [s]	Magnitude [-]
50	29.9±3.6	1.2±0.16
100	29.8±3.3	1.1±0.13
150	29.9±4.3	1.2±0.11
200	26.6±3.6	1.3±0.14

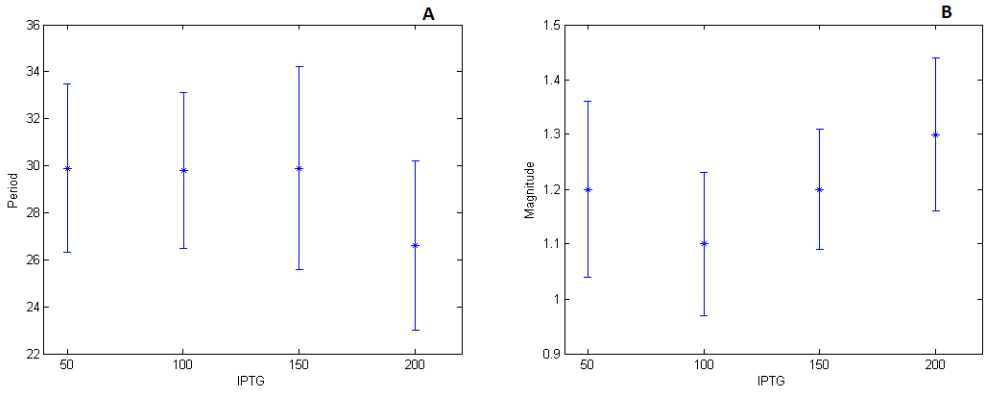

Figure 4.12: **A** shows measured values of oscillation period, **B** magnitudes. Both dependences, thanks to quite big variance, could be interpretable either as linear or inversely linear.

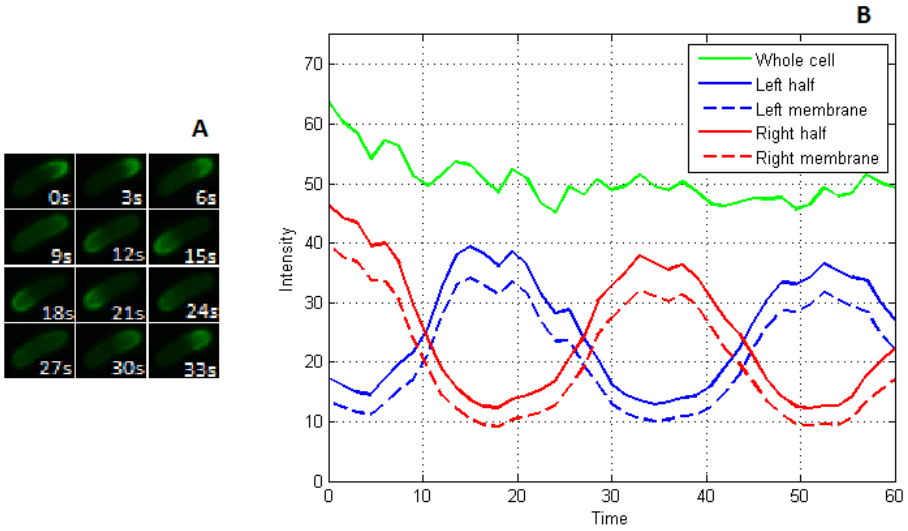

Figure 4.13: **A** Sequence of fluorescence images of a cell induced with 100uM of IPTG. This sequence corresponds to **B**. Left/right membrane stands for membrane range of left/right half of the cell. Effect of photobleaching is obvious as mean value of fluorescence intensity decreases.

The model shows proportional relation between the oscillation **period** and MinD concentration, assuming the fixed MinD to MinE ratio (see Figure 4.10). Experimental results, however, are not conclusive in this respect (see Figure 4.12). Other two experimentally investigated parameters of oscillation, magnitude and shape, both rather accord with the model. While monitoring the oscillation of membrane bound MinDs, the relative **magnitude** extracted from microscopy images ($m_r \doteq 1.2$) belongs to the interval $[0.2, 1.6]$. Considering the cluster identified in Section 4.3.3 as a set of potential parameter values for the system, this interval defines ranges in which the magnitude fluctuates. Also the experimentally measured **shape** of oscillation indicates, that first MinD proteins relatively quickly fill the membrane and then more slowly dissociate, which corresponds to behaviour of the model (see Figure 4.3). When increasing, the wave is more steep.

4.5 Remaining issues

The model predicts the oscillation period varies inversely with the ratio of MinD to MinE. The literature isn't clear on how the ratio of MinD to MinE concentrations influences the oscillation period. Some findings correspond to the theoretical result of this work [3], more, however, claim the oscillation period is proportional to MinD and inversely proportional to MinE concentrations [6], [9]. None of these opinions could be here either confirmed or disproved by experiments. The problem of the dependence of oscillation period on the ratio of MinD to MinE thus remains unsolved. Bacterial strain used for the experiments can be further modified to allow independent variation of MinD and MinE to further validate the presented model.

5 Discussion

Knowledge obtained by investigating the Min system behaviour can be further applied to design cell division control mechanisms. By changing only the protein concentrations, the oscillation period be varied during specific phases of the cell cycle (see 2.2.3). This might, for example, influence time to next division. As a result of high concentrations, the cell might be unable to start the division process. This could work to delay or synchronize cell division. Such a tool could be useful in studying cell cycle properties or developing antibacterial therapies.

Also the the fact that Min proteins itself distribute between the daughter cells equally could be used. Min proteins could serve as carriers for another proteins whose expression is monitored. Their equal division would eliminate noise that naturally arises from the division.

Bibliography

[1] Ganhui Lan, Charles W Wolgemuth, and Sean X Sun. Z-ring force and cell shape during division in rod-like bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 104(41):16110–5, October 2007.

[2] D M Raskin and P a de Boer. Rapid pole-to-pole oscillation of a protein required for directing division to the middle of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 96(9):4971–6, April 1999.

[3] Joe Lutkenhaus. Min oscillation in bacteria. *Advances in experimental medicine and biology*, 641:49–61, January 2008.

[4] Jennifer R Juarez and William Margolin. Changes in the Min oscillation pattern before and after cell birth. *Journal of bacteriology*, 192(16):4134–42, August 2010.

[5] Barbara Di Ventura and Victor Sourjik. Self-organized partitioning of dynamically localized proteins in bacterial cell division. *Molecular systems biology*, 7(457):457, January 2011.

[6] Kerwyn Casey Huang, Yigal Meir, and Ned S Wingreen. Dynamic structures in Escherichia coli: spontaneous formation of MinE rings and MinD polar zones. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22):12724–8, October 2003.

[7] Hassan K. Khalil. *Nonlinear Systems*. Prentice Hall, third edit edition, 2002.

[8] V Sundaram. C OMPARATIVE S TUDY OF D ATA M INING A LGORITHMS FOR. 4(2):173–178, 2012.

[9] Filipe Tostevin and Martin Howard. A stochastic model of Min oscillations in Escherichia coli and Min protein segregation during cell division. *Physical biology*, 3(1):1–12, March 2006.

[10] Martin Loose, Karsten Kruse, and Petra Schwille. Protein self-organization: lessons from the min system. *Annual review of biophysics*, 40:315–36, January 2011.