University of West Bohemia in Pilsen
Faculty of Applied Sciences
Department of Cybernetics

BACHELOR THESIS

Pilsen, 2016 **Jana Rachová**

The official assignment

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 22. srpna 2016 .....

Jana Rachová

Declaration

I declare that I carried out this bachelor thesis independently, only with the cited sources and literature.

Acknowledgements

I would like to thank my supervisor, MSc. Daniel Georgiev, PhD. for his professional guidance and advice.

I also thank all members of Cell cybernetics lab, in particular to Tereza Puchrová and Anna Sosnová for their help and advice.

Abstrakt

Ve všech živých organismech se vyskytuje biologický šum projevující se náhodnými výchylkami v úrovni genové exprese. Za určitých podmínek může zvýšená variabilita genové exprese vést ke vzniku onemocnění. Mezi takovéto rizikové faktory patří ztráta jedné funkční kopie genu. Tato práce se zabývá studiem vlivu tohoto faktorů na stochastickou odezvu feromonové signální dráhy v kvasinkách. Dále je studován vliv genetické modifikace klíčového komponentu feromonové signální dráhy – nukleárního transkripčního faktoru – na stochastickou odezvu této dráhy. Je ukázáno, že přirozený transkripční faktor dokáže potlačovat náhodné změny signálu. Syntetický transkripční faktor tuto schopnost nemá.

Klíčová slova: šum v genové expresi, stochastická charakterizace, feromonová signální dráha v kvasinek

Abstract

The biological noise is present in all living organisms. It manifests as the stochasticity of gene expression. Increased gene expression noise can lead to disease onset under certain circumstances including the loss of one functional copy of the gene. This work study the effect of this factor on the stochastic response of the yeast mating pheromone signal pathway. Further, the effect of genetic modification of a key pathway component – nuclear transcriptional factor – on the stochastic response of this pathway is studied. It is shown that the wild-type transcriptional factor is able to suppress the stochastic fluctuations in signal. The synthetic transcription factor do not perform this ability.

Keywords: gene expression noise, stochastic characterization, yeast pheromone signal pathway

Contents

Introduction

Proteins are the basic regulatory elements in a cell. They participate in the function of enzymes, hormones and they occur as functional elements of cell signal pathways. The production of proteins in cells is influenced by many extrinsic an intrinsic factors. It can lead to occurance of fluctuations in protein level. These fluctuations usually stems from the variability in gene expression. This phenomenon is called gene expression noise and is usual in all living forms. However, the phenomenon of noise can be harmful in the context of the proper signal transmission through the regulatory circuits in the cell.

The presence of increased gene expression noise is sometimes associated with the harmful effects on the fitness of an organism. These aspects are strong motivation for research of gene expression noise. The first part of my thesis is trying to find a context of gene expression noise and human diseases in the professional literature.

Human diseases stem from physiological disorders. There is a simple unicellular organism which is a quite a good model of human physiological mechanisms – yeast. There are observed many signal pathways in yeast, the components of which have homologues in human. In my thesis, the pheromone mating signal pathway is studied. The aim of my thesis is to characterize the stochasticity in the response of the yeast mating pheromone signal pathway to different doses of pheromone input.

Currently, the biotechnology benefits from genetic modifications of biological systems. Therefore, it is interesting to study, how the stochastic characterization of the yeast mating pheromone pathway changes, when a crucial pathway component - the nuclear transcription factor - is genetically modified. Therefore, the characterization of the stochasticity in the response of the modified yeast mating pheromone signal pathway to different doses of pheromone input was also the aim of my work.

Chapter 1

Biological background

1.1 The phenomenon of noise

All living cells receive signals from their environment and response to them [1]. The sequence of these events is called signal pathway. Single components of signal pathway are connected into a circuit, where proteins represent nodes and protein interactions links between them [2]. Actions in the cell are regulated by enzymes either at the level of the total amount of enzyme or at the level of its activity. The process of signal transmission is influenced by many stochastic factors resulting in observed stochastic fluctuations in the level of gene expression and consequently in the concentration of protein or enzyme. This phenomenon is what we called biological noise.

The noise can be divided to an extrinsic and intrinsic noise. Extrinsic noise can be caused by environment or the global pool of housekeeping genes, which affect all parts of a cell system at once. Intrinsic noise stems from stochastic fluctuations in gene expression, when there is a component in a network which is present only in small amount [3],[4]. Further, the effect of noise on the cell can be found both advantageous (e.g. for adaptation or cell to cell communication) or deleterious [**?**],[4].

The presence of increased gene expression noise is also associated with diseases. Although, the evidence of a direct implication from noise to disease is quite hard to find. The authors of the study [5] are dealing with the idea that noise – in synergy with other factors – can be a switch-on mechanism in the onset of autosomal dominant diseases. The idea of this study became the inspiration for the formulation of the research problem solved in my thesis. Therefore in the following chapter the main idea of the study [5] will be described and examples of real autosomal dominant diseases will be given.

1.2 Autosomal dominant diseases

Autosomal dominant diseases (further ADD) are hereditary originated disorders caused by a change in gene located on an autosomal chromosome (autosome). There is either mutation or loss of function in one of two homologous gene loci. The transmission of this sign is conditioned by a dominant allele, which means the disease state occurs when mutated allele (A) is dominant over wild-type allele (a). Phenotypically it means that ADD occurs in heterozygote (Aa) (and in

homozygote (AA) of course). However, some variable characteristics in disease transmission and manifestation were observed. For example the incomplete penetrance (the state when the affected allele manifests in less than expected number of carrying individuals) or variability in time of the disease onset. Currently there are two theories trying to explain the cause of ADD outbreak.

The first one is based on the Knudson's two-hit theory of hereditary cancers and suggests ADD outbreak is a consequence of a wild-type copy damage, either by loss of heterozygosity (and getting homozygote (AA) which is more severe in the context of the disease course) or by somatic mutation in wild-type allele. However, this theory describes rather conditions under which stable disease is maintained while variable time of disease onset stays unexplained.

The second theory takes into account the phenomenon of haploinsufficiency, which can stem from the heterozygosity. Dysfunction of one allele cause reduction of wild-type gene copy number compared with the healthy individual. In the case when copy number reduction causes the loss of sufficient gene function we are talking about haploinsufficiency.

Three step mechanism of ADD onset

Haploinsufficiency often involves some additional mechanisms. Such a mechanism is a noise in gene expression. Synergy of these phenomenon has a potential to be the true cause of variable time ADD onset. When one gene copy is lost, there is stronger possibility that gene expression noise causes a fall of an essential product concentration below a critical level and it results the disease state. However, as the noise is a stochastic variability in gene expression, sooner or later such a deviation is compensated and the disease state disappears. It means the noise explains variable but temporary onset of ADD, thus it seems to be just a primary mechanism. There must be some other influence, which causes the switching and maintaining a stable disease state.

At this point, the structure of key pathway which regulates cellular physiology plays the important role. Global appearance of a network topology - loops, feedbacks, cascades - and functional qualities of single elements – activators, inhibitors – are crucial for noise propagation. The effect of noise can be either diminished or prolonged by certain structures of networks. Such a structure common in biological systems is a feedback loop. It has two functions at a time: firstly, it allows rapid switch on of a process and secondly, it behaves like a noise buffer in order to withstand long-termed noise.

Simulation of ADD onset

In silico experiments were done by authors of the study [5] in order to achieve results of the theory made above. Simulations were made in order to explore the hypothesis that three factors – haploinsufficiency of a gene, noise in its expression and the network structure which the gene is part of – together in synergy explain the onset of ADD.

A model of a key pathway structure of ADD was simplified to a three element network (examples of such networks are shown in the figure 1.1). The most upstream element X represents the signaling protein at the top of the pathway. Following downstream element Y represents an intermediate protein. Finally, the element D at the bottom of the pathway represents the disease state marker protein. Always one of elements X or Y was considered to be influenced by haploinsufficiency. Simulating haploinsufficiency, the answering gene expression level was reduced by 50% of a normal expression level. The normal (100%) level was determined to be when both gene copies express at the critical level. The critical level was determined empirically and set to be the basal expression level of a single gene copy. The presence of noise in gene expression was simulated by the Gaussian noise. The random variable chosen from Gaussian distribution with zero mean was added in each step to the basal production rate of haploinsufficient gene.

It was find out that there are structures, which performs the behavior of the switch-like mechanism and can be switched by a noise. The probability of switch-on the disease state is influenced by haploinsufficiency of appropriate gene in this structure. In the figure 1.1 are presented structures together with the results of simulations. It was find out that these sructures are involved in signal pathways of real autosomal dominant diseases. Examples of such diseases - polycystic kidney disease and the maturity onset diabetes of the young are disused above.

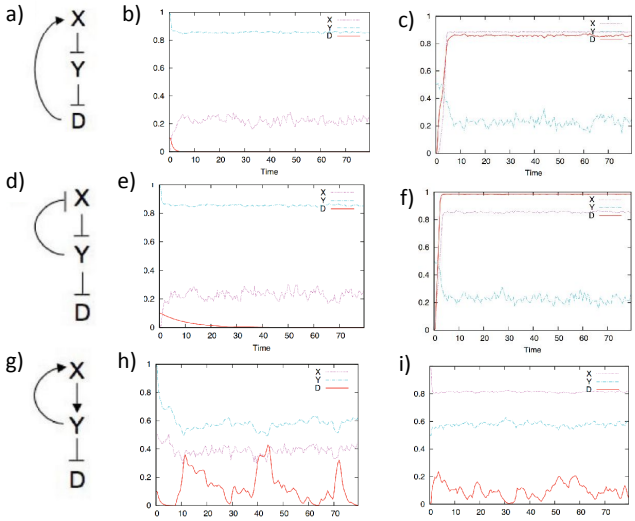

Figure 1.1: (a),(d),(g)- diagrams of simulated networks, (b),(e),(h) - crresponding simulations results for heterozygosity of X positioned gene, (c),(f),(i) - crresponding simulations results for heterozygosity of Y positioned gene. Figures inherited from [5]

Autosomal Dominant Polycystic Kidney Disease (ADPKD)

ADPKD manifests by a cystic dilation of the kidney tubules and cyst formation. Progressive enlargement of cysts leads to kidney enlargement, deformation and eventually failure, which occurs at a later age [6]. Patients have to undergo hemodialysis or organ transplantation. ADPKD is caused by mutation in gene *PKD1* or *PKD2*. *PKD1* gene is coding for PC1 (polycystine-1), the G-protein coupled receptor. *PKD2* gene is coding for PC2 (polycystine-2), the cation channel protein regulating the permeability of Ca^{2+}. [7].

According to the prevalent theory, the ADPKD onset is the result of a second–hit mutation inactivating the functional copy of mutated gene and consequent expansion of affected cells. However, referring to *in vivo* experiments on mice [8] the authors of a study [5] point out two insufficiences of this theory. Firstly, in the most of cyst lining cells both PC1 and PC2 are expressed, which means neither *PKD1* nor *PKD2* can be completely deactivated by the second-hit mutation. Secondly, in the presence of *PKD1* mutant cells, even wild-type cells can contribute to cyst growth. There is an alternative hypothesis which suggests somatic mutations are not the only mechanism of ADPKD onset. Non-genetic factors such as topology of the ADPKD key pathway or stochastic noise in gene expression or in surrounding environment may be possible alternative mechanisms. Simulations dealing with this hypothesis were made.

In the figure 1.2*a*, there is a diagram of the ADPKD key pathway. In this pathway the role of signaling protein plays the TNF-α (tumor necrosis factor α) inflammatory cytokine, the level of which can increase as a result of renal injury, infection or cystic conditions. TNF-α (through the induction of the FIP2 protein) negatively regulates the function of PC2. PC2 in turn negatively regulates the TNFR (tumor necrosis factor receptor). In simulations, PC1 and PC2 are modeled as a functional unit, haploinsufficiency of which is considered to be crucial for the disease onset. The ADPKD pathway contains two motives (in the figure 1.1*a* and *d*) both of which are able to switch on a stable disease state in the case of Y-positioned haploinsufficient element (indeed, here it is PC1-PC2 unit). Further, simulations consider two sources of a noise – firstly the stochastic fluctuations in gene expression and secondly renal injury as the stochastic influence of environment.

In simulations of heterozygous population (figure1.2*b*) the magnitude of stochastic fluctuation in PC1-PC2 level is amplified. Therefore, although in heterozygous population the PC1-PC2 level is generally sufficient to prevent cyst formation, it can easily randomly fall under the threshold so that the feedback loop switches the disease onset. Worthy to mention is that when the simulation considers TNF-α feedback loop is blocked (figure1.2*c*), the disease onset is inhibited which would not be possible considering the ADPKD onset is caused by the second-hit mutation. This is also confirmed by *in vivo* experiments on mice [8] and suggests an important consequence for treatment.

Maturity onset diabetes of the young (MODY)

MODY [9] is heterogeneous group of disorders caused by insulin deficiency predominantly arising from defect in pancreatic β-cells function. The disease onset is usual before 25 years of age. The dynamics of the onset varies as well as the severity of symptoms. In all cases, untreated elevation of blood sugars can result in a damage of many organs (e.g. neuropathy, retinopathy, renal or heart diseases). At genetic level the MODY stems from the mutation in one gene (versus diabetes of the type I and II which stems from polygenic abnormalities [9]) of the key pathway (presented in the picture 1.2d). This pathway contains motive of the type illustrated in the figure 1.1g which has not clearly switch on character of a response (shown in the figures 1.1h and i). However, the MODY pathway includes two embedded subnetworks of the same type and the extra feedback loop, and therefore the results of simulations suggest the MODY key pathway is able to switch on the disease state. Further, the discussed motive (fig. 1.1g) is more sensitive to haploinsufficiency of X-positioned element. Therefore, it can be expected, and simulations confirm it, that the deficiency in gene production higher up the cascade has more severe effect on downstream insulin production.

The noise was simulated as stochastic expression level of involved genes. The random variation in each allele is independent of the other. When a gene is unaffected by haploinsufficiency the total production rate is relatively high and random variation of each allele offsets the other. However, in haplodefficient gene random fluctuations are higher and the positive feedback loop can be easily switched on, which causes the disease onset.

The less severe form is MODY 2 caused by a loss of single allele of the gene coding for glycokinase (glycolytic enzyme). The simulation results (fig. 1.2e) corresponds to the known disease characteristics – an early onset and mild, uniform decrease of insulin level. The MODY 2 does not get worse with the time. It is suggested this is caused by the absence of glycokinase in any feedback loop.

More severe forms are MODY 1 and MODY 3 caused by mutated hepatocyte nuclear factors hnf-4α resp. hnf-1α positioned in the cascade upstream to glycokinase. Both forms have similar clinical manifestation (slightly milder symptoms at MODY 1) at most consistently to simulations (fig. 1.2f and g). The only difference observed in the age of onset (when for MODY 1 the onset is earlier and less variable compared to MODY 3) is suggested to be a consequence of answering genes placement in a pathway structure, which is complicated with double feedback loop from hnf-4α. Further, as in the case MODY1 so if MODY3 the insulin level was found generally reduced in mean but stronger in fluctuations compared to MODY 2 simulations.

The most severe and forms MODY 4 (slightly milder) and MODY 5 (very severe form) are caused by deficiencies in transcriptional factors coding genes pdx1 resp. hnf-1β. MODY 4 and MODY 5 simulations (fig. 1.2h and i) shows switch-like behaviour. There are two stable states, either no disease or very high disease level (corresponding to very low level of insulin). When the insulin level falls low enough due to random fluctuation, the stable disease state occurs.

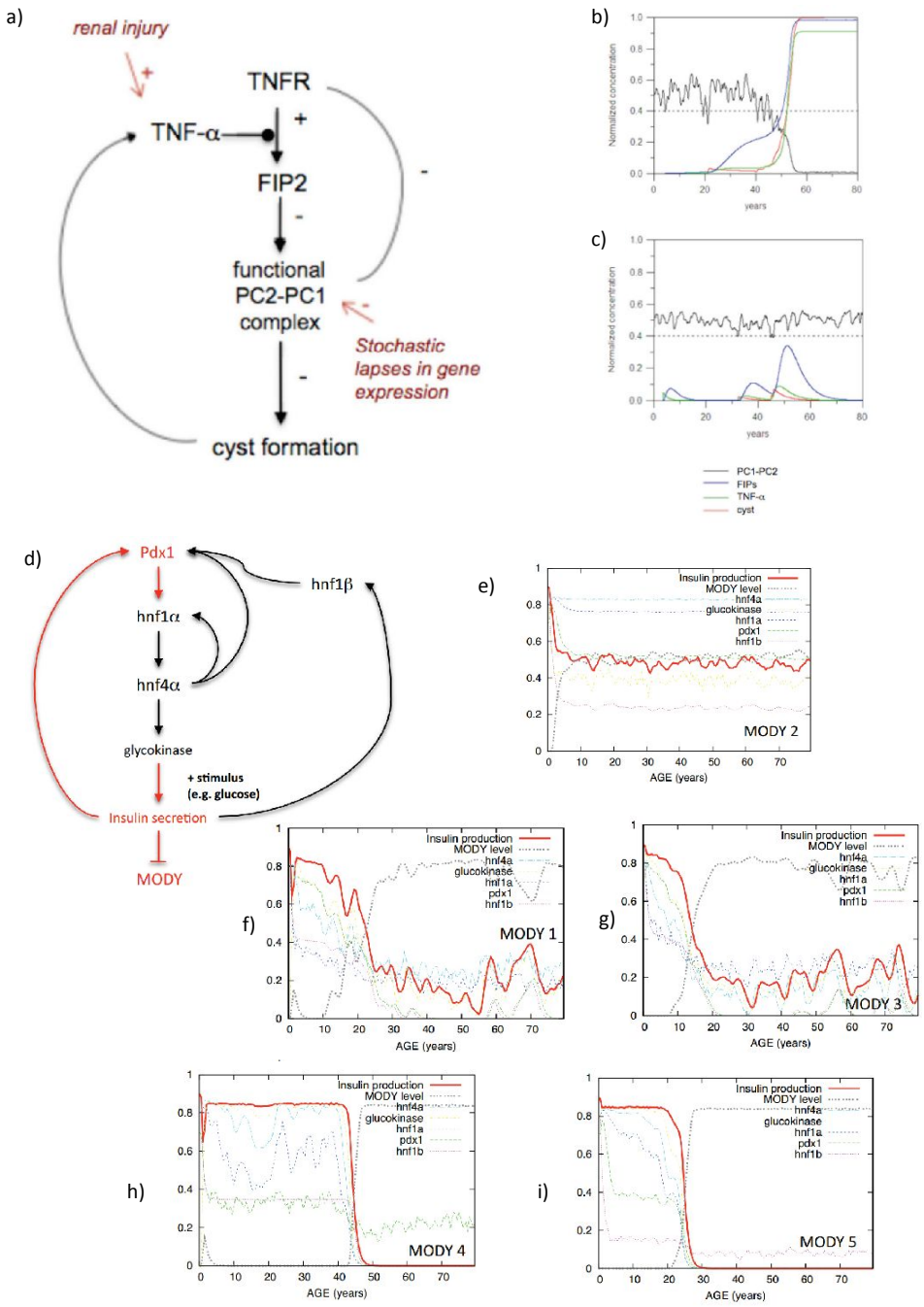

Figure 1.2: (a)- key signal pathway of ADPKD (b) - simulation of haplodeficient population (level of polycystin is lower and the fluctuations are higher) (c) - simulation of the state, when TNF-α feedback loop is blocked (d) key signal pathway of MODY (e),(f),(g),(h),(i) - simulations for haploinsufficiency of glycokinase (MODY 2), hnf-4α (MODY 1), hnf-1α (MODY 3), Pdx1 (MODY 4) and hnf-1β (MODY 5). Figures inherited from [5]

1.3 The signal pathway in yeast

1.3.1 The life cycle of yeast

Baker's yeast [10] (*Saccharomyces cervisiae*) is one of the most studied eukaryotic organisms. It was the first eukaryotic organism whose genome was completely sequenced and many other primacy were made even when examining *Saccharomyces cervisiae*. Yeast is strongly used in the research of the cell cycle, intracellular signaling and protein-protein interactions. They are appreciated as a good model organism also because they are simple unicellular organisms which can be easily manipulated. At the same time there are great similarities in genes, proteins and cellular processes in cells of yeasts and of higher organisms, consequently human. This fact gives a potential to the research of yeasts, that gained results could be generalized.

As with all eukaryotes, yeast has genetic information saved in the form of nuclear DNA in chromosomes. There is also extranuclear DNA, namely mitochondrial and plasmid which usually do not carry genes coding for essential life functions. Worth mention plasmids are very important for research, because they are used like basic vectors for genetic modifications. Yeast cells occur in two life forms – haploid and diploid [11]. Haploid cells contain in its nucleus just one set of chromosomes, while nucleus of diploid cells contains two sets of chromosomes. Both haploid and diploid cells undergo classical vegetative cycle which consists of growth period and period of asexual reproduction (mitotic division) which we call "budding" in case of yeast. However, yeast is also capable of sexual reproduction. In the period of stress diploid cells enter the process of sporulation (meiotic division) which results in formetion of spores. A spore contains four haploid cells always two and two of opposite mating types. When the period of stress is over, the spore lapses and haploid cells are released into the environment. Haploid cells of opposite mating type can undergo the process of mating, which results in fusion into one diploid cell.The life cycle of yeast is illustrated in the figure 1.3

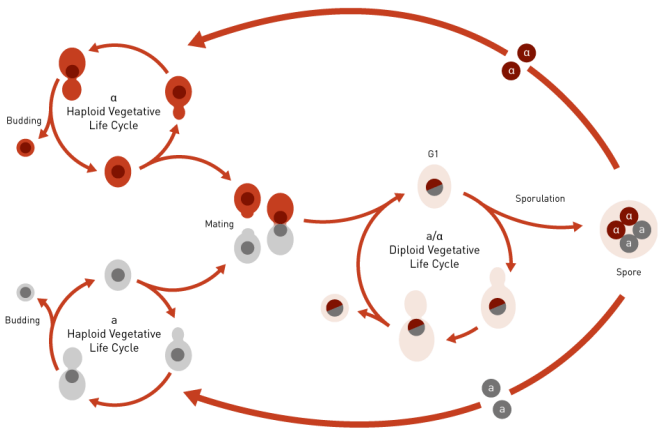

Figure 1.3: The illustration of the yeast life cycle aviable in [12]

1.3.2 The yeast mating pheromone signal pathway

The process of mating [2],[13] is quite complex and begins relatively long time before the event of two haploid opposite mating type cells fuse. All starts in the moment when one cell recognizes the presence of a mating partner in its proximity. Capturing this information starts a complex process of signal transmission and transduction at the intracellular level. At the end of this cascade, the signal gets into the cell nucleus and triggers the cell fusion. The most important of these changes is the pheromone induced expression of more than 200 genes. Other pheromone induced changes are the cell cycle arrest and the cell polarization towards its mating partner (change of the shape into so called "shmoo"). Actually, the process of fusion is the last step and result of a long cascade of biochemical processes which we call "the yeast mating pheromone signal pathway" (further YMP). The whole signal transmission system including this pathway will be described below and schema of the YMP is illustrated in the figure 1.4 At this place I would like to emphasize, that I will focus primarily on those elements of YMP which are part of a mathematical model used in next work, so that description will be quite simplified compared to natural biological systems.

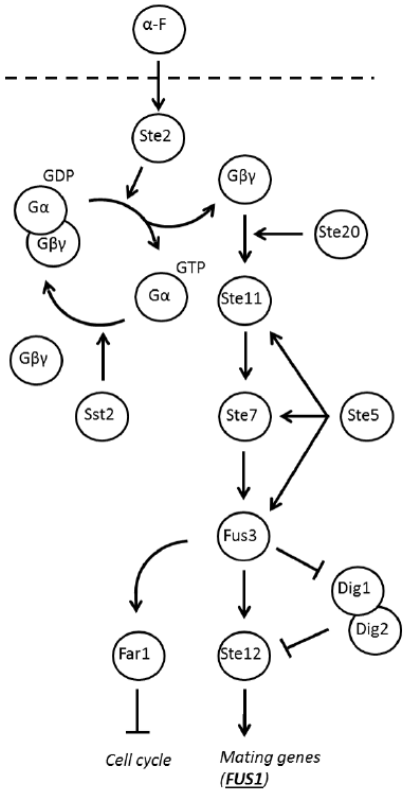

Figure 1.4: The diagram of the yeast mating pheromone signal pathway

Before I will talk about the signal transmission I would firstly introduce how the signal arises. Generally in biological systems signal molecule is secreted into environment by the cell [1]. Such a signal molecule is pheromone in case of YMP. Under normal conditions, each mating type produces and secretes into environment the specific amount of its own characteristic pheromone. Each type has its own specific receptor which is sensitive to pheromone of the opposite mating type. In the case of *Saccharomyces cervisiae* mating types are MATa and Matα, so that cells of the type MATa secrets the a-factor and its receptor (Ste2) is sensitive to α-factor. In the case of MATα the situation is opposite (and the receptor is Ste3). The receptor is the only part of the YMP and in which mating types differ; the rest of the pathway is identical.

GPCR (G-protein coupled receptors)

The first step of YMP is the binding of pheromone molecule to the receptor (GPCR) on cell surface. Function of the receptor is the regulation of intracellular protein activity so that signal is transmitted through the cell membrane. Achieving this, GPCR are cooperating with G-proteins (guanine nucleotide binding proteins). After the receptor binds a ligand, there are retrieved conformational changes of receptors itself and consequentl binding of G-protein.

In the idle state, G-protein is the heterotrimer consisting of subunits $G\alpha$ (Gpa1), $G\beta$ (Ste4), $G\gamma$ (Ste18). The $G\alpha$ is responsible for regulation of the G-protein own activity. In the idle state $G\alpha$ also binds GDP (guanosine diphosphate). After pheromone induced activation $G\alpha$ is stimulated to release the heterotrimer and to exchange of GDP for GTP (guanosine triphosphate). In the active state, $G\beta$ and $G\gamma$ subunits form a heterodimer and participate in signal transduction further along the pathway by binding to three effectors: Ste5 scaffold protein, Ste20 protein kinas and Far1 protein.

The activated state can be terminated by hydrolysis of GTP to GDP on $G\alpha$, which consequently re-associate with the $G\beta\gamma$ heterodimer (the protein Sst2 regulates the termination).

Scaffold protein

Scaffold protein is quite large and multifunctional but catalytically inactive protein. Its main function is the interaction with members of a signal pathway and their co-localization. Scaffold protein Ste5 (the first discovered signal scaffold protein) organizes and co-localizes single members of MAPK cascade (Ste11, Ste7, Fus3 and Kss1 kinases),and enables signal transmission.Ste5 forms a complex with the first member of the MAPK, Ste11.

The fact that Ste5 binds to $G\beta\gamma$ heterodimer close to the site, where Ste20 protein kinase binds and is activated (Ste20 is activated by phosphorylation after it binds to the inner side of the membrane), is not just a coincidence. Activated Ste20 is responsible for activation of Ste11. Overall it means that Ste5 plays a

role of the MAPK member's holder and Ste20 is the finger which sets in motion the first domino of the MAPK cascade.

The MAPK cascade (Mitogen activated protein kinase cascade)

The MAPK cascades occurs in all eukaryotes and they generally plays a role in regulation of hormonal activity, cell differentiation and stress responses.

MAPK present in yeast mating pathway belongs to the group of ERK (extracellular signal regulated kinases) and through the MAPK cascade the signal is transmitted from receptor to the nuclear transcription factors. The signal propagation is a gradual change in the activity of the individual components which through covalently modification – phosphorylation, when the phosphate group PO_4 is transferred from one element of the cascade to another. MAPK cascade consists of three components: the component nearest to the nucleus is MAPK (Fus3, Kss1) which is phosphorylated by upstream MAPKK (Ste7) which is in turn phosphorylated by its upstream MAPKKK (Ste11).

The Ste11 protein does not seem to have especially high affinity to its downstream substrate Ste7. There is usual a transient enzyme-substrate interaction and the key component of signal transmission from Ste11 to Ste7 is a scaffold protein Ste5, which put them stabilizes the bond between Ste11 and Ste7.

In contrast, the Ste7 protein has quite high affinity to its downstream substrates Fus3 and Kss1. The significantly stronger interaction than normal enzyme-substrate is due to the D-site (docking site) motif of Ste7 (D-site motif was firstly discovered in Ste7). D-site motif is the mediator of the bond between Ste7 and its substrates. Moreover Ste7 and its substrates bind to specific regions of Ste5 scaffold protein. These mechanisms – docking and scaffolding - are mutually reinforcing. Their functional overlap serves as a safety factor in order to accomplish effective signal transmission.

Nuclear transcription factors

Through the MAPK cascade the signal gets into the nucleus. Substrates of MAPKs Fus3 and Kss1 are nuclear transcription factor complex Ste12/ Dig1/ Dig2, the Far1 protein (its function lies primarily in the stimulation of the cell polarized growth and the mediation of the pheromone induced arrest of a cell cycle in G1 phase) and other substrates.

Ste12 is a DNA binding transcriptional factor, which binds with its DNA-binding site to the "pheromone response element" (the DNA motif in A/TGAAACA) in the promoter of answering genes. The pheromone stimulation induces transcription of genes. The strains without Ste12 are defective in this pheromone induced gene expression.

Dig1 and Dig2 are protein repressors that in the idle state bind and repress Ste12 transcription factor. Genes, which have pheromone induced expression are

upregulated in strains without Dig1, Dig2.

Inactive Kss1 regulates the Ste12 by specific mechanism - repression of transcription by inactivated MAPK. When unphosphorylated, Kss1 binds directly to Ste12 and represses its transcription. (In contrast, the second MAPK - Fus3 is a weaker repressor of Ste12.) When Kss1 is phosphorylated, it releases the bond to Ste12. Further, active Kss1 and Fus3 directly phosphorylate the Ste12/Dig1/Dig2 complex. Consequently, Dig1 and Dig2 repression of Ste12 is inhibited and the pheromone induced gene expression is activated.

Between genes whose transcription is activated by Ste12, there are components of mating pathway (*STE2, FUS3, FAR1*), negative feedback regulators of the pathway (*SST2, MSG5, GPA1*), genes associated with the fusion process (e.g. *FUS1*). Ste12 also binds its own promoter and it stimulates its own expression (positive autoregulation).

1.4 Pheromone activated factor

Numerous hybrid proteins were created in the study [14] by fusion of different regions of transcription factors Ste12 and Gal4.

The Ste12 transcription factor consists of 688 amino acids. At the N-terminus Ste12 has a DNA-binding domain (residues 1 - 215). At the C-terminal end, there is a transcription activation domain (residues 384 - 688) which is necessary for activation of the basal transcription in the absence of pheromone as well as pheromone induced transcription in the presence of pheromone. The region between these two domains (residues 216 – 383) forms a pheromone induction domain it is capable for response to pheromone stimulation.

In the study [14] was defined the region of pheromone induction domain sufficient to confer pheromone induction to hybrid protein. It was found, this minimal induction domain is a region bound by residues 301 and 335. Further foundation was that pheromone induction domain alone activates the transcriptional activity weakly in the presence of pheromone. To achieve a significant pheromone induced transcriptional activity, the pheromone induction domain cooperates with the adjacent transcriptional activation domain.

The fusion protein H30 from the study [14] was constructed and internally named PAF (pheromone activated factor) by my colleague Anna Sosnová [15]. This fusion protein contains the minimal induction domain and activation domain of the Ste12 transcription factor (residues 301-688) and the DNA-binding domain (GBD) of the Gal4 transcription factor. Gal4 is a DNA-binding transcription factor which serves as a positive regulator for the gene expression of galactose induced genes (e.g.*GAL1, GAL2*) by binding to the upstream activating sequence of these genes [16].

Taken together, Ste12 and PAF differ primarily in their DNA-binding domain. Ste12 binds to pheromone response element of pheromone induced mating genes

(e.g. *FUS1, FAR1*). Among them there are genes encoding proteins placed upstream of Ste12 in the yeast pheromone mating pathway. Hence, the feedback mechanism is mediated by Ste12.

PAF binds to the upstream activating sequence of these galactose induced genes (e.g. *GAL1, GAL2*). By this hybrid transcription factor the pheromone induced transcription of any gene, having the upstream activating sequence of galactose induced genes, can be achieved. In contrast to Ste12, PAF does not activate the pheromone induced transcription of genes coding for proteins placed upstream to Ste12 in the yeast pheromone mating pathway. It means using PAF, the feedback mechanism can be inactivated.

Chapter 2

Research

The interaction of such effects as a topology of a signal pathway, a deficiency of gene copy number in some key component of a pathway and a presence of stochasticity (noise) in gene expression which was examined with respect to the autosomal dominant disease onset in the study [5], raises the question how some other signal pathway would behave under similar conditions. In the case of my research problem it is the yeast pheromone mating pathway. Many components of YMP have human homologues. Therefore the study of this pathway is interesting from the perspective of an examinination of the function of human physiology.

2.1 Formulation of the problem

The key component of YMP is considered to be the nuclear transcription factor. As mentioned in chapter 1.4, it is possible to replace the wild-type Ste12 transcfription factor with the hybrid PAF, so that the synthetic signal pathway with a different topology is created. Both these pathways are examined in following research. In the figures 2.1a resp. 2.1b there are simplified block diagrams of a wild-type resp. synthetic pathway, where the pheromone is an input and green fluorescent protein (GFP) is an output.

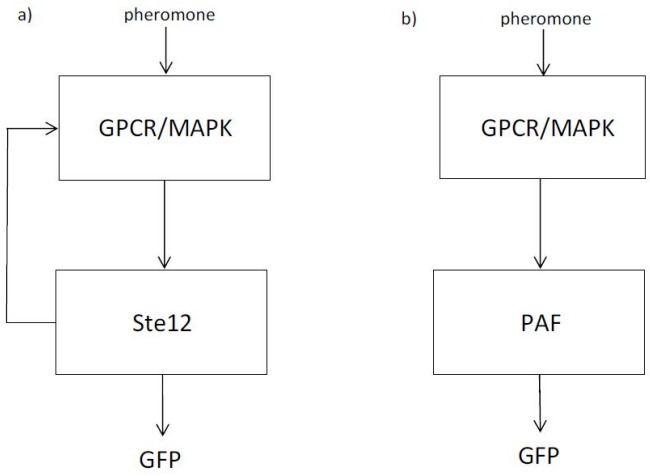

Figure 2.1: Block diagrams of a wild-type (a) and synthetic (b) pathway

The wild-type pathway contains a feedback loop, as the Ste12 transcriptional factor binds to promoters of genes placed upstream in the YMP. Induction of gene expression of mating genes is not considered, because there was used sterile strain in expreriments in vivo. Autoregulatory feedback loop is not considered, because for experimental purpose, constitutive promoters were placed upstream to STE12 gene instead of its own promoter. It means this pathway is not truly wild-type at all, however it is used as a model of wild-type pathway and it is termed by this way in the following text. As the reporter it is used GFP. The synthetic pathway does not contain the feedback loop and as the reporter gene it is used GFP again.

The aim of my work is to analyze the stochastic characterization of wild-type resp. synthetic signal pheromone pathway response to different doses of pheromone input and to compare the responses of both pathways. Another aim is to analyze how the response of both pathways changes, when gene coding for the key pathway component - transcriptional factor Ste12 resp. PAF - is present in the cell in full(2) resp. half(1)copy number resp. What is the difference between responses of pathway with haplodeficient gene and the pathway with gene present in full number of copies?

In order to achieve these aims, in silico and in vivo experiments were performed.

2.2 In silico performance

The Yeast Pheromone Signalling Model [17] was used for the purpose of the experiment in silico. This model is written in the BioNetgen rule-based modeling language (BNGL). Modifications of the former model are made using RuleBender [18],[19], free tool for work with rule-based models in the BioNetGen Language, which enables simple definition of intermolecular reactions. RuleBender is able to communicate with MATLAB software, which is useful when there is a need to run simulations repeatedly and to change their parameters continuously. Stochastic simulations are performed using the NFSim algorithm.[20].

2.2.1 Computational model

To model pathways defined above it is necessary to modify some single parts of the complex Yeast Pheromone Signaling Model. The original Yeast Pheromone Signaling Model contains the rule representing the process of Ste12 binding to its own promoter. In my work, this process of the Ste12 autoregulation is not considered. Therefore, corresponding interaction iss deleted. In the model of a synthetic pathway all interactions of transcriptional factor PAF with promoters of upstream pathway genes are deleted. Last modification is adding of rules representing the interaction of particular transcription factor with the particular promoter of reporter gene GFP and its expression. Equations of chemical reactions responding to this process are given bellow, together with their RuleBender

notations and contact maps acquired from the RuleBender tool. In figure 2.2)is shown a contact map of the wild-type pathway and in the figure 2.3)is shown a contact map of the synthetic pathway.

In order to simulate haplodeficiency state of a gene coding for tha particular transcription factor, corresponding initial conditions of the Yeast Pheromone Signaling Model are modified. To simulate the full(2) resp. half(1) gene copy number, the initial level of particular transcription factor is set to the value 3000 molecules (30nM) resp. 1500 molecules (15nM). The model considers the interaction of transcription factor (either Ste12 or PAF) with the Dig1 and Dig2 protein repressors. The most important: the complex $Ste12 - Dig1^*$ is present from the start of simulation. The total amount of Dig1 in the original model is set to 4799 molecules. As Ste12 and Dig1 form a complex, the amount of free Dig1 (and of course also Ste12) molecules decreases. Therefore, the initial conditions has to be modified for three components of the model - $Ste12 - Dig1^*$ complex, $freeDig1$ and $freeSte12$. Under the above conditions, the amount of the transcription factor is always lower than the amount of Dig1 - in both cases simulating full or half gene copy number. Consequently, all molecules of the transcription factor are consumed to form a complex with Dig1 and there are initially no free transcription factor molecules. Specific values of initial conditions are summarized in the table 2.1. Simulations of Ste12 resp. PAF considering the full resp. half gene copy number are termed $Ste12^{(2)}$, $Ste12^{(1)}$, $PAF^{(2)}$, $PAF^{(1)}$.

Simulations are performed for varying pheromone input(from 0.1nM to 10 nM of pheromone). The time of simulations is 2000s. It is considered that the steady state is reached after 2000s. In order to study the stochasticity of the simulated process, the sufficient number of simulations has to be performed. The Monte Carlo method is used for determination of apropprate number of simulation (subsection 2.2). The concrete calculated numbers of simulations for $Ste12^{(2)}$, $Ste12^{(1)}$, $PAF^{(2)}$, $PAF^{(1)}$ and for all pheromone inputs are summarized in the table 2.2.

Table 2.1: The table of simulations sets and corresponding initial conditions (values are given in absolute number of molecules).

Simulation	Initial conditions		
$Ste12^{(2)}$	`Ste12_Dig1^*` = 3000	`Dig1_free` = 1799	`Ste12_free` = 0
$Ste12^{(1)}$	`Ste12_Dig1^*` = 1500	`Dig1_free` = 3299	`Ste12_free` = 0
$PAF^{(2)}$	`PAF_Dig1^*` = 3000	`Dig1_free` = 1799	`PAF_free` = 0
$PAF^{(1)}$	`PAF_Dig1^*` = 1500	`Dig1_free` = 3299	`PAF_free` = 0

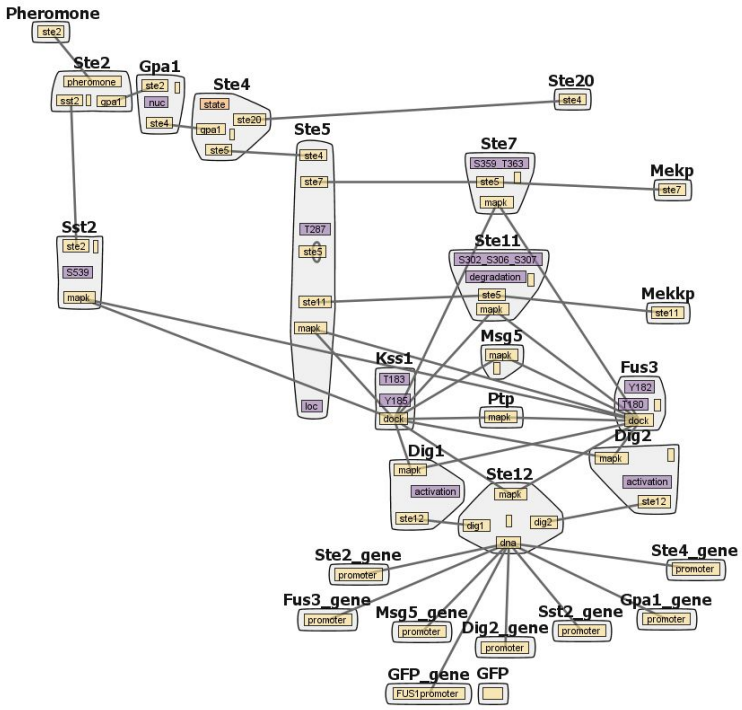

Figure 2.2: The contact map of wild-type pathway model from RuleBender

Chemical equations of reporter gene expression

$$Ste12 + FUS1 \underset{k_B}{\overset{k_F}{\rightleftharpoons}} Ste12{-}FUS1^* \tag{2.1}$$

$$Ste12{-}FUS1^* \xrightarrow{kT} Ste12{-}FUS1^* + \text{Gfp} \tag{2.2}$$

$$\text{Gfp} \xrightarrow{kD} \emptyset \tag{2.3}$$

RuleBender notation

```
Ste12(dna) + GFP_gene(FUS1_promoter) ->
-> Ste12(dna!1).GFP_gene(FUS1_promoter!1) 2.145e-05

Ste12(dna!1).GFP_gene(FUS1_promoter!1) ->
->Ste12(dna) + GFP_gene(FUS1_promoter) 0.03

Ste12(dig1,dig2,mapk,dna!1).GFP_gene(FUS1_promoter!1) ->
-> Ste12(dig1,dig2,mapk,dna!1).GFP_gene(FUS1_promoter!1) + GFP() 15

GFP() -> Trash() 6.9e-5 DeleteMolecules
```

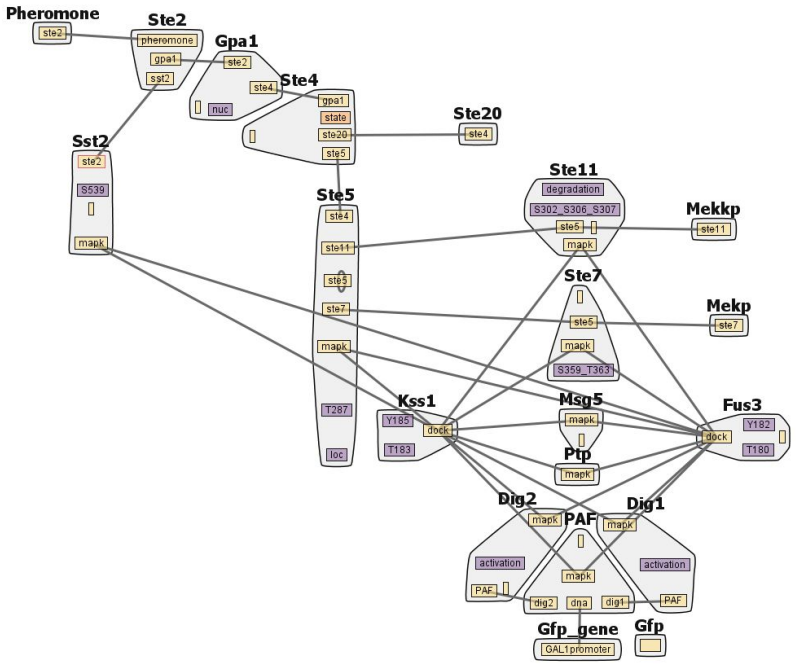

Figure 2.3: The contact map of synthetic pathway model from RuleBender

Chemical equations of reporter gene expression

$$Ste12 + GAL1 \underset{k_B}{\overset{k_F}{\rightleftharpoons}} Ste12{-}GAL1^* \tag{2.4}$$

$$Ste12{-}GAL1^* \xrightarrow{kT} Ste12{-}GAL1^* + \text{Gfp} \tag{2.5}$$

$$\text{Gfp} \xrightarrow{kD} \emptyset \tag{2.6}$$

RuleBender notation

```
PAF(dna) + GFP_gene(GAL1_promoter) ->
-> PAF(dna!1).GFP_gene(GAL1_promoter!1) 2.145e-05

PAF(dna!1).GFP_gene(GAL1_promoter!1) ->
-> PAF(dna) + GFP_gene(GAL1_promoter) 0.03

    PAF(dig1,dig2,mapk,dna!1).GFP_gene(GAL1_promoter!1) ->
-> PAF(dig1,dig2,mapk,dna!1).GFP_gene(GAL1_promoter!1) + GFP() 15

GFP() -> Trash() 6.9e-5 DeleteMolecules
```

2.2.2 Monte Carlo convergence

The Monte Carlo methods [21] are numerical methods of solving mathematical problems, which uses modeling of random variables and statistical estimation of their characteristics. The most frequent approach of the Monte Carlo method is to model such a random variable X that mean value $E(X)$ equals searched value a. It means, in order to calculate the value a, firstly, random variable X is searched, which $E(X) = a$. As independent realization $X_1, X_2, ..., X_N$ are found, the value a can be calculate as 2.7. There is an infinite number of random variables X which are subject to $E(X) = a$.

$$a = \frac{1}{N}(X_1 + X_2 + \cdots + X_N) \tag{2.7}$$

The Monte Carlo method is also associated with the problem of estimation of parameters of normal distribution. The estimation of a mean value $E(X)$ gives the searched value a and the estimation of variance $D(X)$ gives the estimation of error of Monte Carlo method. The error of Monte Carlo method can be estimated according to the following perscription[22].

Let the random variable X represents the number of GFP molecules synthesized as a result of gene expression at time t. The stochasticity in gene expression (noise) can be defined by the random variable Y (expression 2.8 responds to one realization of random variable Y). Let say in following expressions X_i resp. Y_i represent realization of random variables X resp. Y. $E_N(X)$ resp. $E_N(Y)$ represent the estimated mean value of random variable X resp. Y calculated from N realizations.

$$Y_i = |X_i - E_N(X)| \tag{2.8}$$

Let the error of the estiomation of the $E(Y)$ value is given by residue 2.10. The value $E(Y)$ is unknown and the value of $E_N(Y)$ is estimated using the known formula for sample standard deviation 2.10.

$$r = E(Y) - E_N(Y) \tag{2.9}$$

$$E_N(Y) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - E_N(X_i)))} \tag{2.10}$$

The required error is wanted to satisfy the expression 2.11, where I set the option of accuracy $\varepsilon = 100$ molecules and the level of significance $\alpha = 0.05$.

$$P(|r| < \varepsilon) = 1 - \alpha \tag{2.11}$$

According to the central limit theorem [21],[22] for the sequence $A_1, A_2, ..., A_N$ of independent random variables with the same distribution F, mean $E(A_i) = \mu$ and non-zero variation $D(A_i)$, the probability given by left side of the expression 2.12 goes with $N \to \infty$ to the normal distribution N(0,1). (The value $\sigma(A)$ is the standard deviation given by $\sigma(A) = \sqrt{D(A)}$.)

$$P\left(\frac{E(A) - E_N(A))}{\sigma(A)/\sqrt{N}} < \varepsilon\right) \approx N(0,1) \tag{2.12}$$

Applied on the above problem, with $N \to \infty$ it is:

$$P\left(|E(Y) - E_N(Y)| < \varepsilon\right) \approx N\left(0, \frac{\sigma(Y)}{\sqrt{N}}\right) \tag{2.13}$$

Expressions 2.11 and 2.13 together suggest, the value ε responds to the quantile of normal distribution and its numerical value can be calculated using the expression 2.14, where $z_{1-\frac{\alpha}{2}}$ is the $100(1-\frac{\alpha}{2})\%$ quantile of normal distribution.

$$\varepsilon = z_{1-\frac{\alpha}{2}} \frac{\sigma(Y)}{\sqrt{N}} \tag{2.14}$$

The expression 2.14 was used for calculation of appropriate number of simulation (expression 2.15) in order to satisfy the condition given by expression 2.11.

$$N = \frac{z^2_{1-\frac{\alpha}{2}} \sigma^2(Y)}{\varepsilon^2} \tag{2.15}$$

For the calculation of the number of simulations (2.11), the value of standard deviation $\sigma(Y)$ is required. As the real value of this characteristic is unknown, it can be estimated with the help of the $\sigma_N(Y)$ value, calculated by formula 2.16. This formula represents the deviation of the deviation. The value of $\sigma_N(Y)$ is calculated from results of N pilot simulations. I set the $N = 5$ because of length of simulations course (one simulation run takes about 20 minutes).

$$\sigma_N(Y) = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} (X_i - E_N(X_i))} \tag{2.16}$$

Table 2.2: The simulations numbers satisfying the condition 2.11

Pheromone dose [nM]	0.1	1	10	20	100	1000
$Ste12^{(2)}$	0	0	1	202	219	229
$Ste12^{(1)}$	0	0	0	23	31	51
$PAF^{(2)}$	0	0	0	157	97	411
$PAF^{(1)}$	0	0	0	21	27	34

2.3 In vivo performance

Experiments performed *in silico* using computer simulations were also transferred to the laboratory and performed *in vivo*. For this experimental purpose was used the sterile yeast strain, which is specifically defective in pheromone response. Cells of this sterile strain have the ability to activate transcription in response to pheromone stimulation. However, they do not arrest the cell cycle and do not change the shape towards the potential mating partner after pheromone induction which took advantage of.

2.3.1 Design of experiments

As an initial strain was used MLY215ΔSTE12 Mat$a$ (yeast strain of a mating type a with knocked-out gene STE12). The purpose of experiments was to examine both wild-type and synthetic yeast mating pathway. The only component, which these pathways differ in, is the transcription factor. A plasmid carrying the STE12 gene was transformed into strains used for examining the wild type pathway. A plasmid carrying the coding sequence of the PAF gene construct transformed into strains used for examining the synthetic pathway.

Further, to simulate the conditions of the full (2) and half (1) gene copy number (normal state and haplodeficient state) of the gene coding for the particular transcription factor, their coding sequences were placed downstream of two different promoters: pTET20 and pLAC13. The pLAC13 promoter is approximately half as strong as pTET20. Therefore pLAC13 was used to simulate the state, when downstream gene is haplodeficient. The pTET20 promoter was used to simulate the full gene copy number of a downstream gene.

A plasmid carrying a reporter GFP gene placed downstream of either the pFUS1 (in strains containing STE12 gene) or pGAL1 (in strains containing PAF gene construct) promoter was transformed into each strain.

Using the initial strain, four other yeast strains were created. In the following text these strains and cultures of these strains will be referred under names $Ste12^{(2)}$, $Ste12^{(1)}$, $PAF^{(2)}$, $PAF^{(1)}$. The table 2.3 summarizes which DNA coding sequences were added in particular strains by transformation.

Table 2.3: The table of strains and corresponding plasmid inserts

The strain name	Plasmid inserts	
$Ste12^{(2)}$	pTET20-STE12	pFUS1-tGFP
$Ste12^{(1)}$	pLAC13-STE12	pFUS1-tGFP
$PAF^{(2)}$	pTET20-PAF	pGAL1-tGFP
$PAF^{(1)}$	pLAC13-PAF	pGAL1-tGFP

In cultures of each of strains $Ste12^{(2)}$, $Ste12^{(1)}$, $PAF^{(2)}$, $PAF^{(1)}$, the signal pathway was induced by different dose of pheromone. The pheromone induction was performed by adding variously diluted cultures of α-cells. A flow cytometer was used for fluorescence assays. More detail description of used strains, plasmids and laboratory protocols is specified in appendix A.

2.3.2 Cytometric data processing

Figures 2.4-2.7 show data obtained by the cytometric fluorescence assay of the $Ste12^{(2)}$, $Ste12^{(1)}$, $PAF^{(2)}$, $PAF^{(1)}$ cultures. In each of figures are six graphs corresponding to six samples induced by different pheromone dose. Graphs are sorted by increasing dose of pheromone (by increasing concentration of added α-cell culture respectively). In most of samples, we can see two clusters. The upper cluster represents cells with higher fluorescence (as the x axis represents the level of fluorescence) and therefore I assume this cluster correspond to pheromone induced a-cells which I am interested in. The lower cluster probably correspond to a-cells, which are not induced, and α-cells (even these cells have a certain level of basal fluorescence). As dose of pheromone increases, the upper cluster is more distinctly separated from the lower one. It has to be noted that as the dose of the pheromone increases the number of points in the upper cluster decreases. It is caused by the fact that the sample induced by higher amount of pheromone contains naturally higher number of α-cells. Consequently, the sample has to contain lower number of a-cells because all assayed samples had the same volume (as mentioned in appendix A).

In order to perform statistics, it is necessary to make pre-processing of raw cytometric data. This pre-processing comprises separation of the points corresponding to cells of interest from the other points. This process is called gating; for this purpose simple empirical linear method was implemented in the software MATLAB. Results of gating are also shown in figures 2.4-2.7. A diagonal line separates the upper cluster from the lower one. The left vertical line separates points corresponding to the cells, which are too small (as the y axis represents the size of cells). The right vertical line separates the small and poorly recognizable cluster, which probably corresponds to double events (when cytometer accidentally takes two cells at once). Infigures 2.4-2.7, red points represents the cells of interest, however, the number of points changes significantly between samples, which can be misleading for consequent statistic calculations. Therefore, from each sample, the same number of values has been chosen (these are represented by black points). This data set was used for further calculations given in following sections.

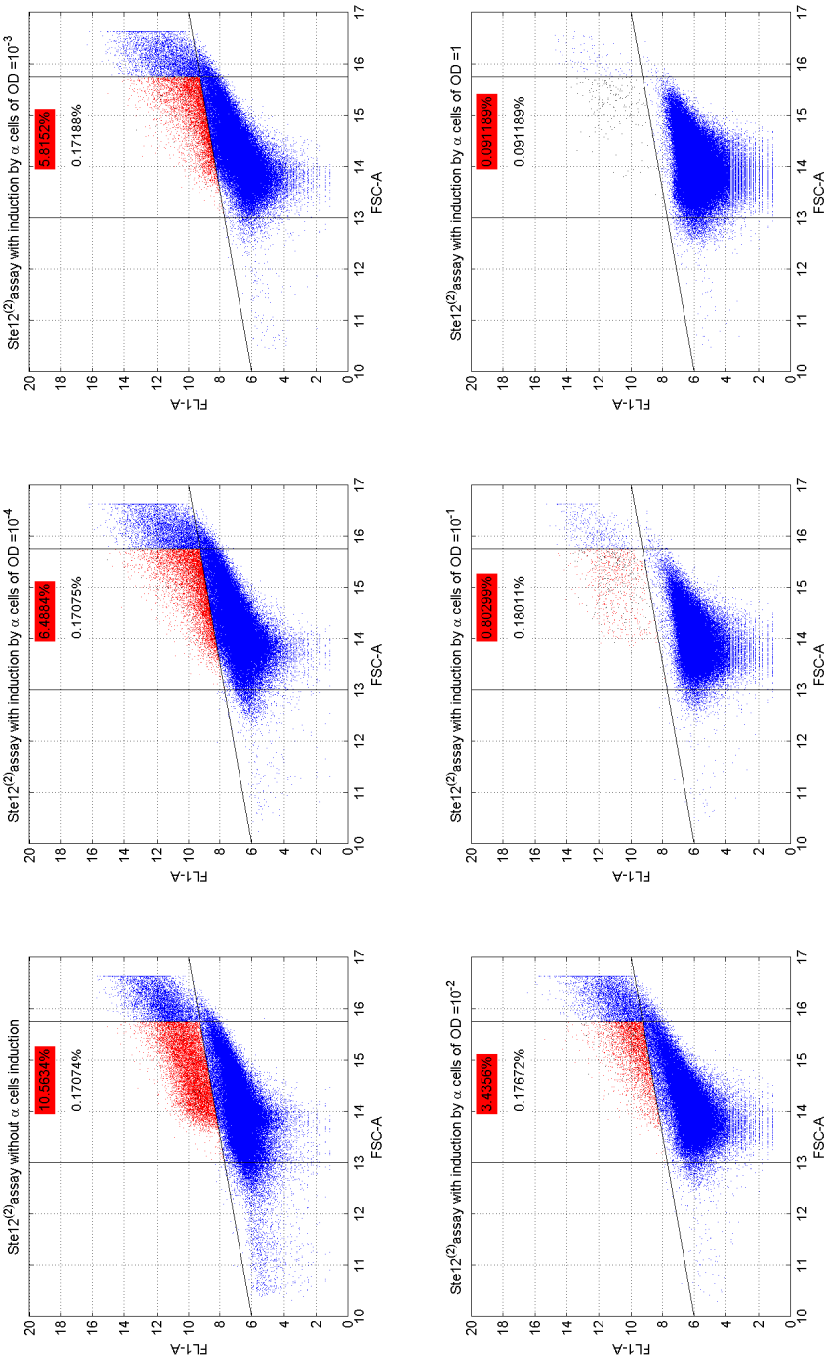

Figure 2.4: The cytometric data of $Ste12^{(2)}$ assay. Red points - fluorescent a-cells of interest, black points - cells used for statistics, red/black number - the proportion of corresponding cells to all.

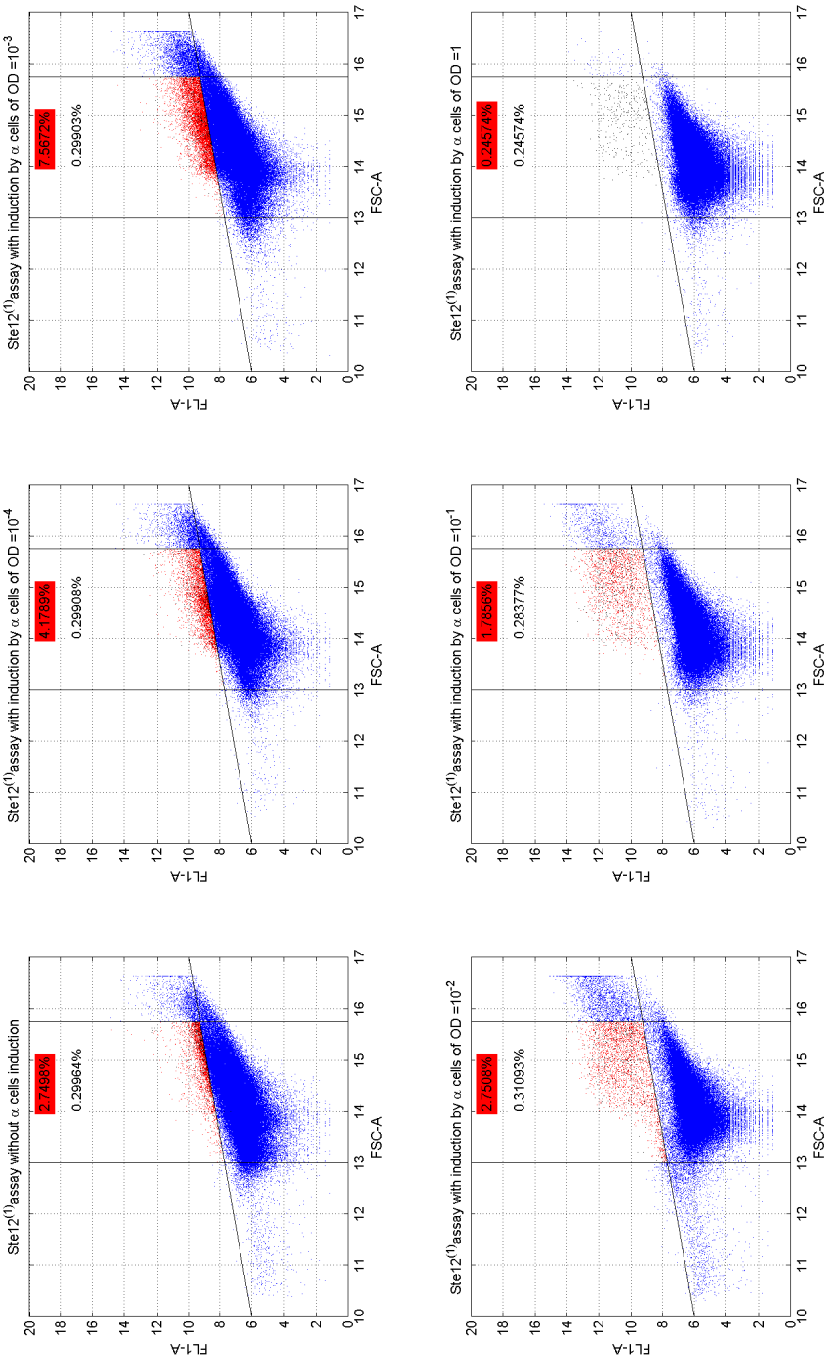

Figure 2.5: The cytometric data of $Ste12^{(1)}$ assay. Red points - fluorescent a-cells of interest, black points - cells used for statistics, red/black number - the proportion of corresponding cells to all.

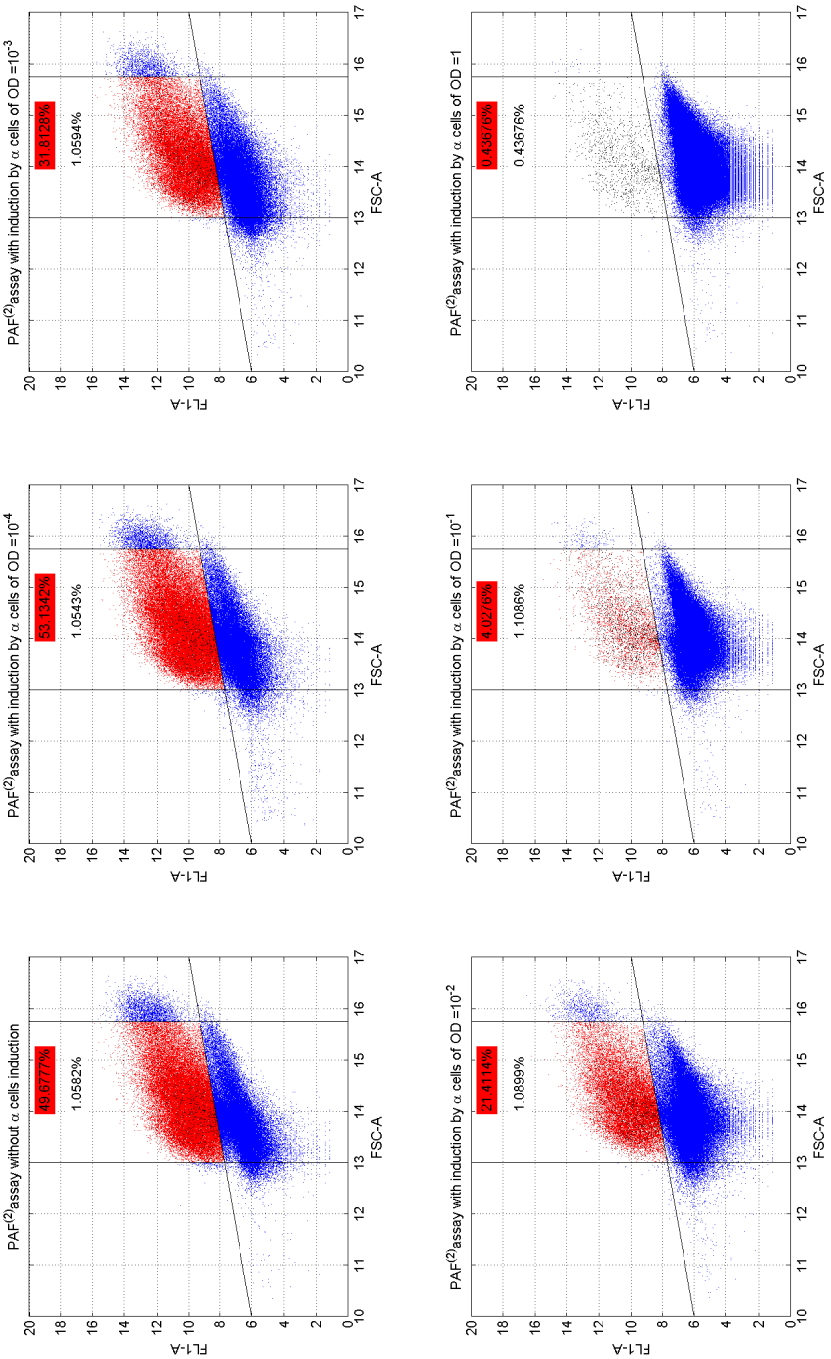

Figure 2.6: The cytometric data of $PAF^{(2)}$ assay. Red points - fluorescent a-cells of interest, black points - cells used for statistics, red/black number - the proportion of corresponding cells to all.

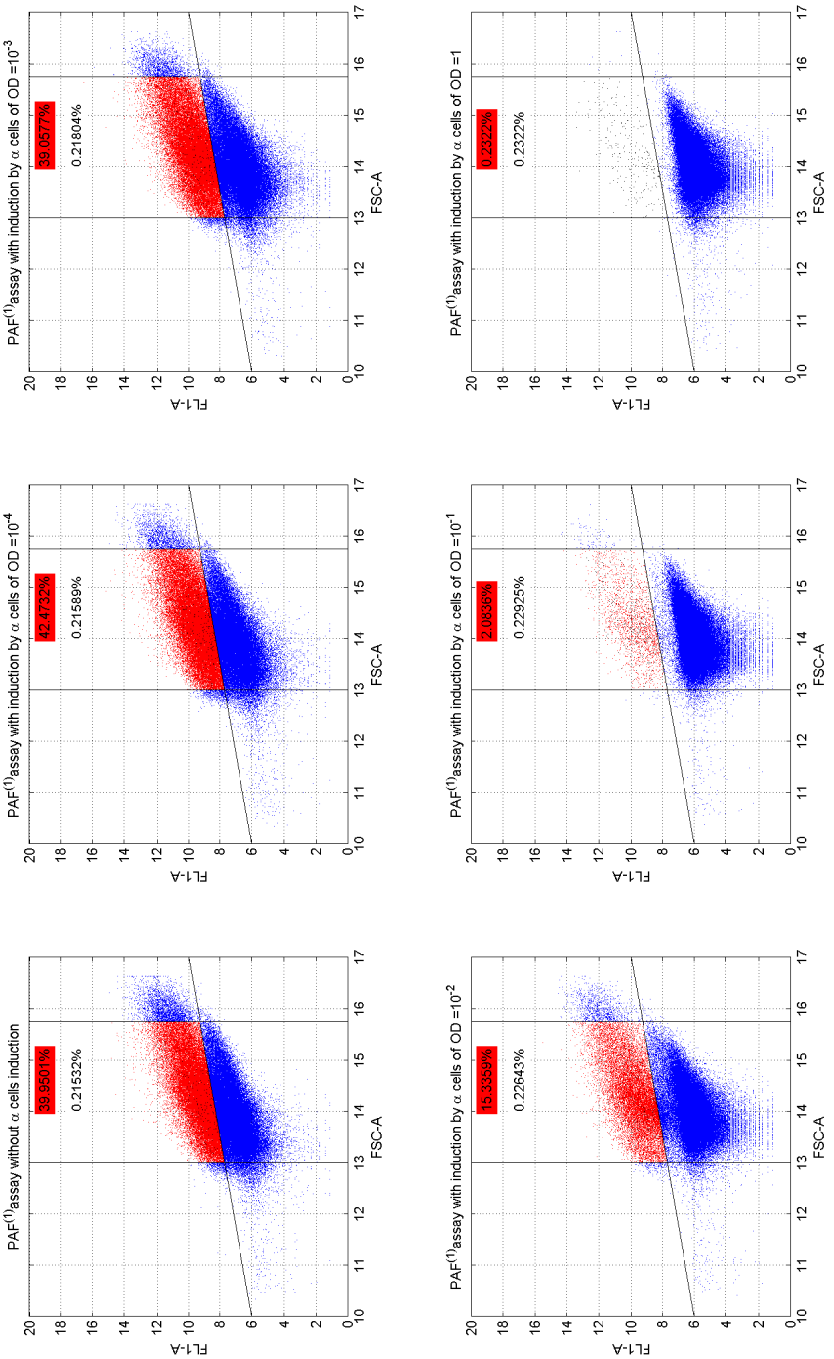

Figure 2.7: The cytometric data of $PAF^{(1)}$ assay. Red points - fluorescent a-cells of interest, black points - cells used for statistics, red/black number - the proportion of corresponding cells to all.

Chapter 3

Data analysis and results

3.1 Explorative analysis

The purpose of explorative data analysis is to reveal their features and to validate assumptions for subsequent statistical processing[23]. Data sets gained from both in silico and in vivo experiments represent one-dimensional selections from a certain distribution. To be able to correctly calculate proper characteristics of tendency and variability, it is necessary to determine what is the distribution, which the data comes from.

In silico experimental data

In the figure 3.1 there are shown histograms of the GFP molecules frequency. These histograms corresponding to simulations of $Ste12^{(2)}$ (details about this simulations in chapter 2.2). The pathway is induced by varying dose of pheromone, therefore, there are six histograms and each of them corresponds with different pheromone dose (as labeled). Simulations stay deterministically at zero level for the first two histograms, because used model does not assume the activation of a pathway by too weak concentrations of pheromone. Other histograms are considered to show normally distributed data sets. The data from the rest of simulations ($Ste12^{(1)}$,$PAF^{(2)}$ and $PAF^{(1)}$) has the similar character.

The set of suitable statistics is calculated from the normal distributed data sets. These statistics are arithmetic mean 3.1, variation 3.2 and the confidence interval for mean value 3.3 (where t_α is a quantile of Student's distribution).

$$E(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i) \tag{3.1}$$

$$D(X) = \sqrt{D(X)} \tag{3.2}$$

$$\left(E(X) - t_{1-\frac{\alpha}{2}}(N-1)\sqrt{\frac{D(X)}{N}}; E(X) + t_{1-\frac{\alpha}{2}}(N-1)\sqrt{\frac{D(X)}{N}}\right) \tag{3.3}$$

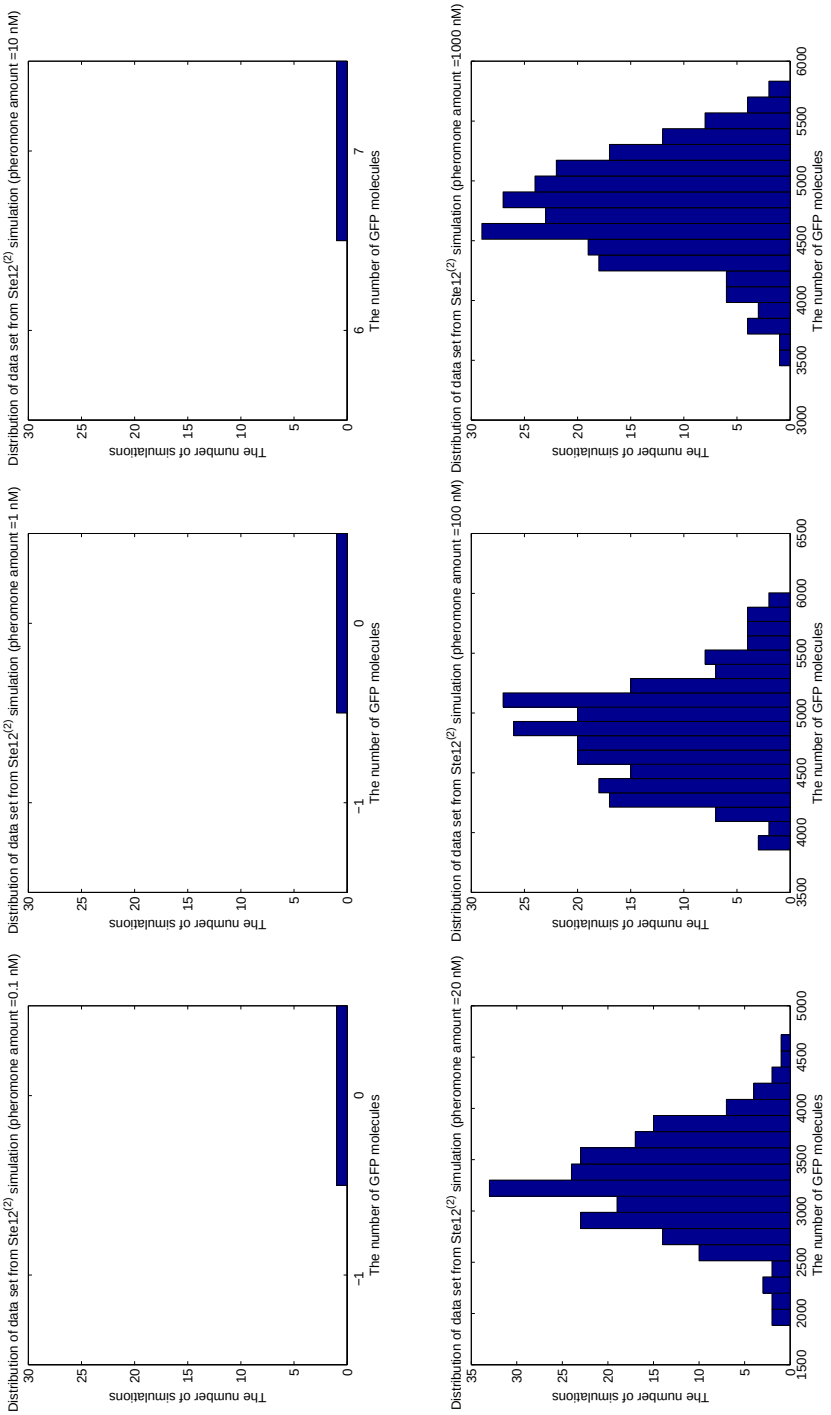

Figure 3.1: The histogram of the GFP molecules for $Ste12^{(2)}$ simulation.

In vivo experimental data

Histograms of GFP fluorescence values frequency, getting from $Ste12^{(2)}$ assay are represented in the figure 3.2. The rest of data has the similar character. Again, each of six histograms correspond to induction by different pheromone dose(as labeled). Obviously, the data do not come from a normal distribution. For that reason it is necessary to make a suitable data transformation, in order to convert them to normal distribution.

The Box-Cox transformation 3.4is suitable for the approximation to the normal distribution in the view of the skewness and kurtosis. [23]. For the zero value of the parameter λ the Box-Cox transformation responds to logarithmic transformation.

$$Y = g(X) = \begin{cases} \dfrac{X^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ ln(X) & \lambda = 0 \end{cases} \tag{3.4}$$

The diagnostic tool for estimation of optimal parameter λ is Hines-Hines selection graph [23]. It is based on the requirement of a symetrically distributed quantiles around median. This requirement is given by expression 3.5, where x_{P_i} is $P_i\%$ quantile of the data set's empirical distribution.

$$\left(\frac{x_{P_i}}{x_{0.5}}\right)^{\lambda} + \left(\frac{x_{0.5}}{x_{1-P_i}}\right)^{-\lambda} = 2 \tag{3.5}$$

In order to compare the trend of experimental points $[(x_{0.5}/x_{1-P_i}); (x_{P_i}/x_{0.5})]$ with theoretical trend corresponding to particular value of λ parameter, the teoretical curves are ploted. These theoretical curves represent the solutions of the equation 3.6.

$$y^{\lambda} + x^{-\lambda} = 2 \qquad 0 \leq x \leq 1 \wedge 0 \leq y \leq 1 \tag{3.6}$$

The suilable λ parameter can be estimated by location of experimantal points on theoretical curves of Hines-Hines graph. The set of six Hines-Hines graphs ploted for data sets from $Ste12^{(2)}$ simulations is presented in the figure 3.3. The result of data transformation is presented in the figure 3.4. The correction of the data distribution is clearly visible. The data from the rest of experiments ($Ste12^{(1)}$,$PAF^{(2)}$ and $PAF^{(1)}$) has similar character and were processed analogically.

By using transformed data, statistics 3.1-3.2 are calculated. However, in order to get the correct values of statistics of original data sets, the reverse transformation has to be done. The reverse transformation process stems from Taylor series expansion of a function $Y = g(X)$ around the mean value $E(Y)$. Approximate formulas for calculation of retransformed statistics can be derived [23] and are given by expressions 3.7 (retransformed arithmetic mean), 3.8(retransformed variation).

$$E_R(X) = g^{-1}\left[E(Y) - \frac{1}{2}\frac{d^2 g(X)}{dX^2}\left(\frac{dg(X)}{dX}\right)^2 D(Y)\right] \tag{3.7}$$

$$D_R(X) = \left(\frac{dg(X)}{dX}\right)^2 D(Y) \tag{3.8}$$

The retransformed confidence interval can be calculated according to the formula mentione d bellow, where I_D resp. I_H is the lower resp. upper limit of the confidence interval and t_α is a quantile of Student's distribution).

$$
\begin{aligned}
I_D &= E_R(X) - g^{-1}\left(E(Y) + -\frac{1}{2}\frac{d^2g(X)}{dX^2}\left(\frac{dg(X)}{dX}\right)^2 D(Y) - t_{1-\frac{\alpha}{2}}(N-1)\sqrt{\frac{D(Y)}{N}}\right) \\
I_H &= E_R(X) + g^{-1}\left(E(Y) + -\frac{1}{2}\frac{d^2g(X)}{dX^2}\left(\frac{dg(X)}{dX}\right)^2 D(Y) + t_{1-\frac{\alpha}{2}}(N-1)\sqrt{\frac{D(Y)}{N}}\right)
\end{aligned}
$$

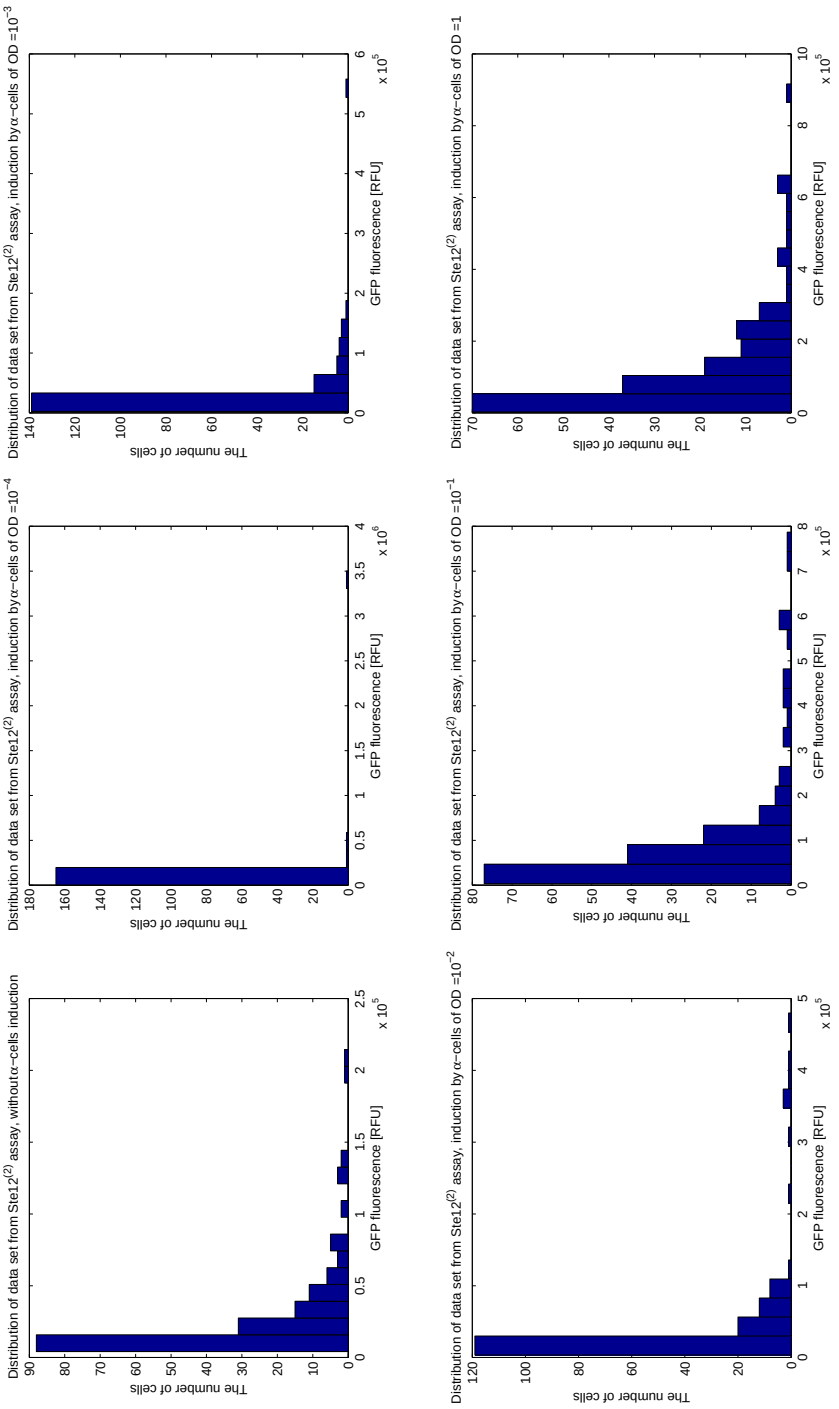

Figure 3.2: Histograms of the GFP fluorescence values frequency for $Ste12^{(2)}$ assays

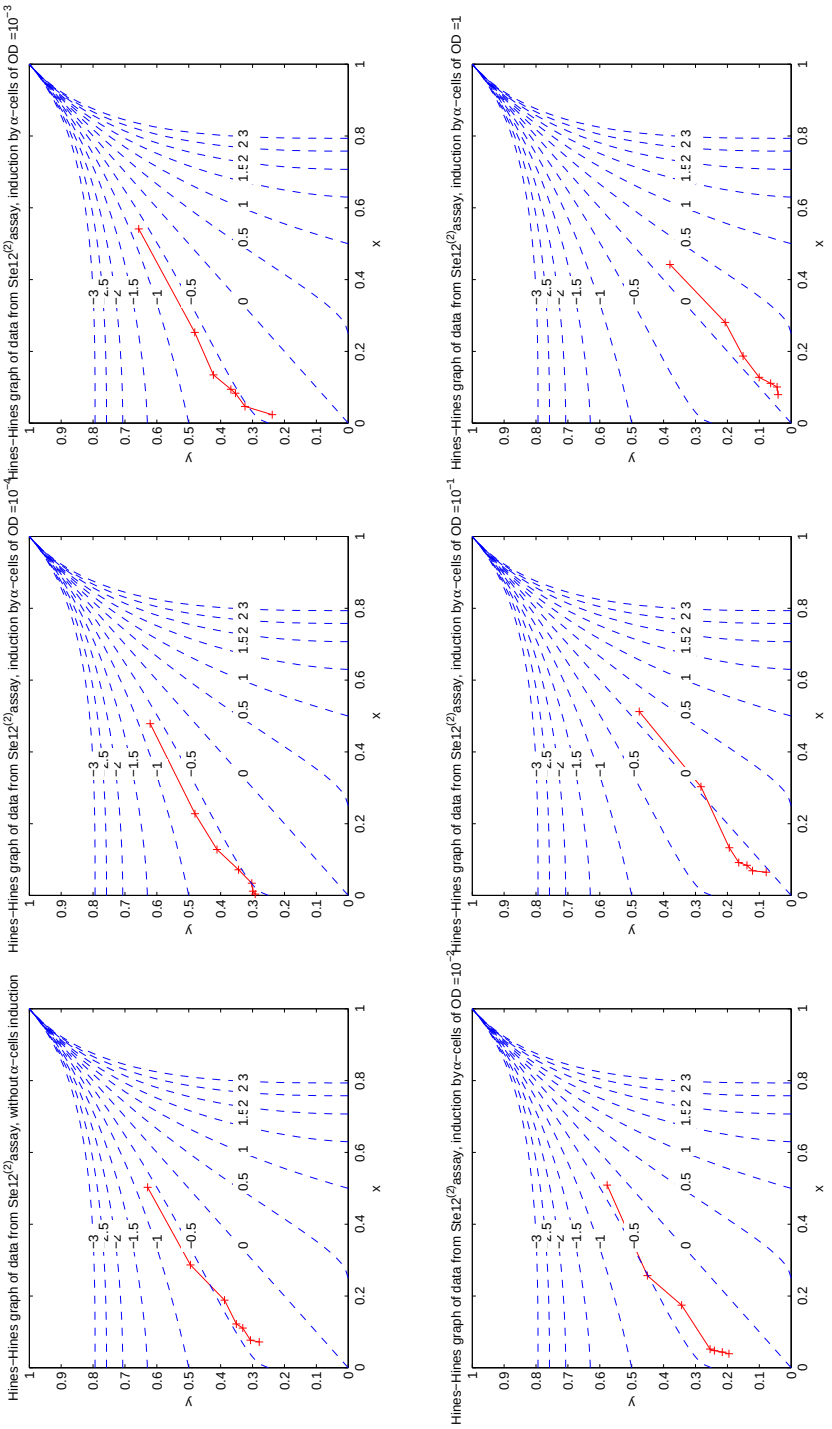

Figure 3.3: Hines-Hines selection plot for $Ste12^{(2)}$ assays.

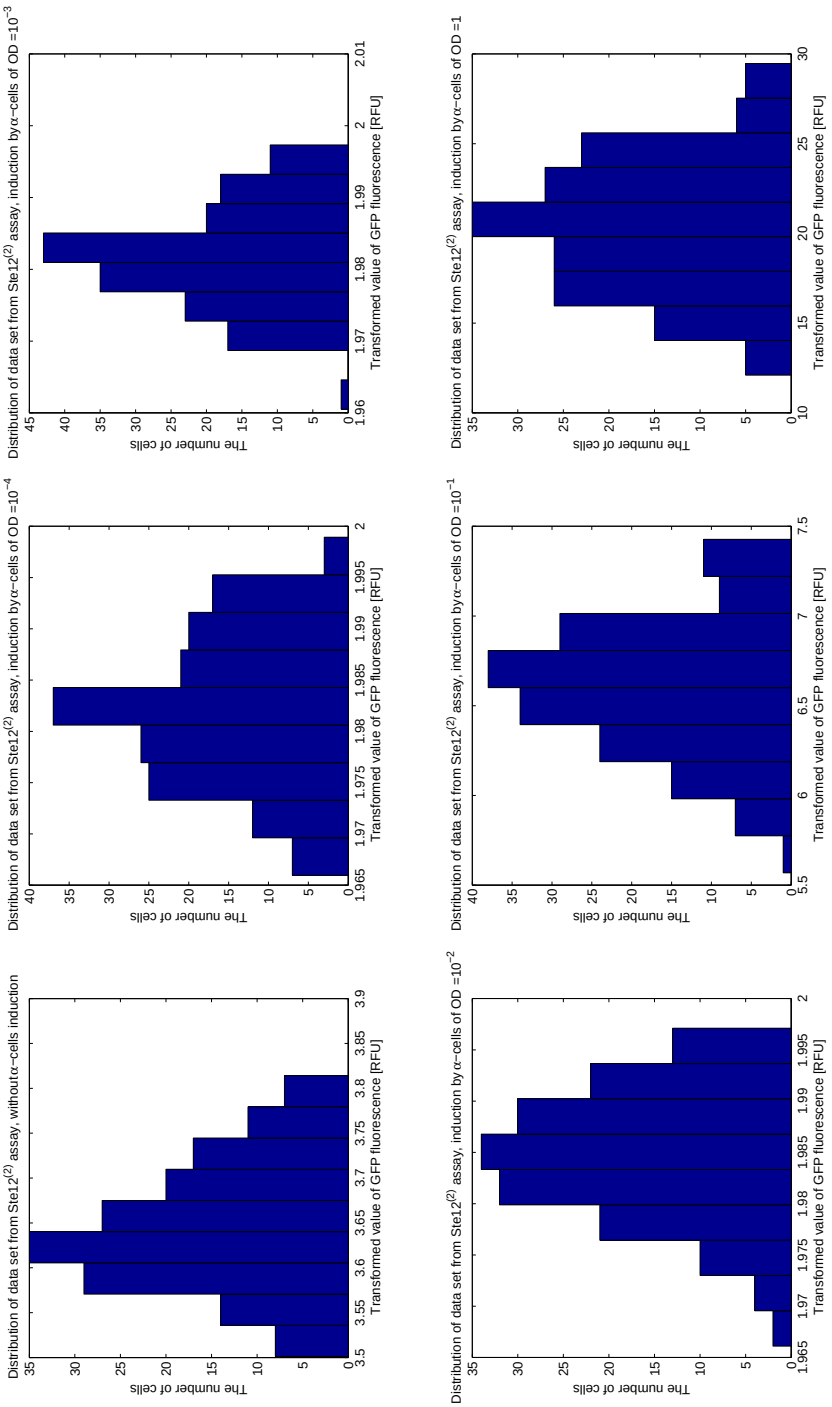

Figure 3.4: Histograms of the GFP fluorescence values frequency for $Ste12^{(2)}$ assays after Box-Cox transformation

3.2 Statistic analysis

Dose response curve is used to plot the change in the effect on an organism caused by varying doses of a chemical after a certain exposure time.[24]. The horizontal axis represents the input - concentration of a stressor (e.g. drug, hormone). Usually, a logarithmic scale is used. The vertical axis represents a response of an organism. It can be represented by any biological function (e.g. enzyme activity, production of a protein). The dose response curve represents the input-output characteristics or static characteristic (when the exposure time is long enough to system reaches a steady state).

Coefficient of variation is represented by ratio of a standard deviation and mean value 3.9. It gives a relative measure of variability. It is a realistic characteristic of variability when comparing data sets with significantly different mean values [25]. Its value is dimensionless and it can be given in percentage.

$$CV(X) = \frac{\sqrt{D(X)}}{E(X)} \tag{3.9}$$

Results of experiments performed *in silico*

Simulations of the wild-type ($Ste12^{(2)}$, $Ste12^{(1)}$) and synthetic ($PAF^{(2)}$, $PAF^{(1)}$) pathways were performed for pheromone doses varying from 10 to 100000 molecules (resp. 0.1nM - 10^3nM). The output was considered to be the mean level of the GFP production. Simulation time was 2000s.

The dose response curves obtained from simulations are shown in the figure 3.5.The activation of all examined systems starts significantly as the level of pheromone stimulation reaches 10nM. The figure 3.5 shows that the half gene copy number (simulations $Ste12^{(1)}$ resp. $PAF^{(1)}$) causes approximately a half of the amplification compared to simulations considering the full gene copy number (simulations $Ste12^{(2)}$ resp. $PAF^{(2)}$). The amplification of a response of the synthetic pathway ($PAF^{(2)}$,$PAF^{(1)}$) is generally lower compared to the amplification of a response of the swild-type pathway ($Ste12^{(2)}$, $Ste12^{(1)}$). Comparing $Ste12^{(2)}$ and $Ste12^{(1)}$, the reduction of amplification of the output resulting from decrease of the gene copy number is less than 50% of $Ste12^{(2)}$ response. In the case of PAF, the reduction of amplification resulting from the gene copy number decrease is more than 50%. of $PAF^{(2)}$.

Error bars in the figure 3.5 represent the 95% confidence intervals for the mean value of the GFP level. Decrease of a gene copy number ($Ste12^{(1)}$, $PAF^{(1)}$) results in a wider range of confidence intervals. Consistently, the coefficient of variation of both Ste12 and PAF responses increases as the gene copy number decreases (figure 3.6). It suggests the output of haplodeficient system shows greater variability compared to the output of system with full gene copy number. The figure 3.6 shows the higher variability in the response of a synthetic pathway compared to wild-type. The $PAF^{(2)}$ (although having the full gene copy number) response reaches the same variability as haplodeficient wild-type ($Ste12^{(1)}$).

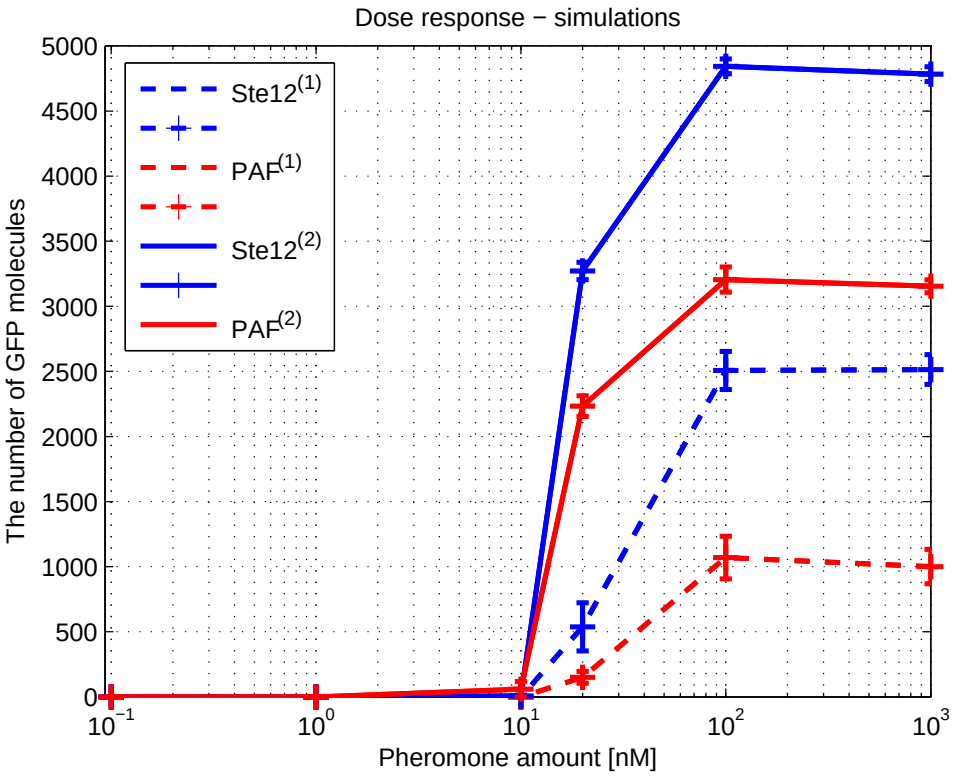

Figure 3.5: Dose response curves obtained from experiments *in silico*

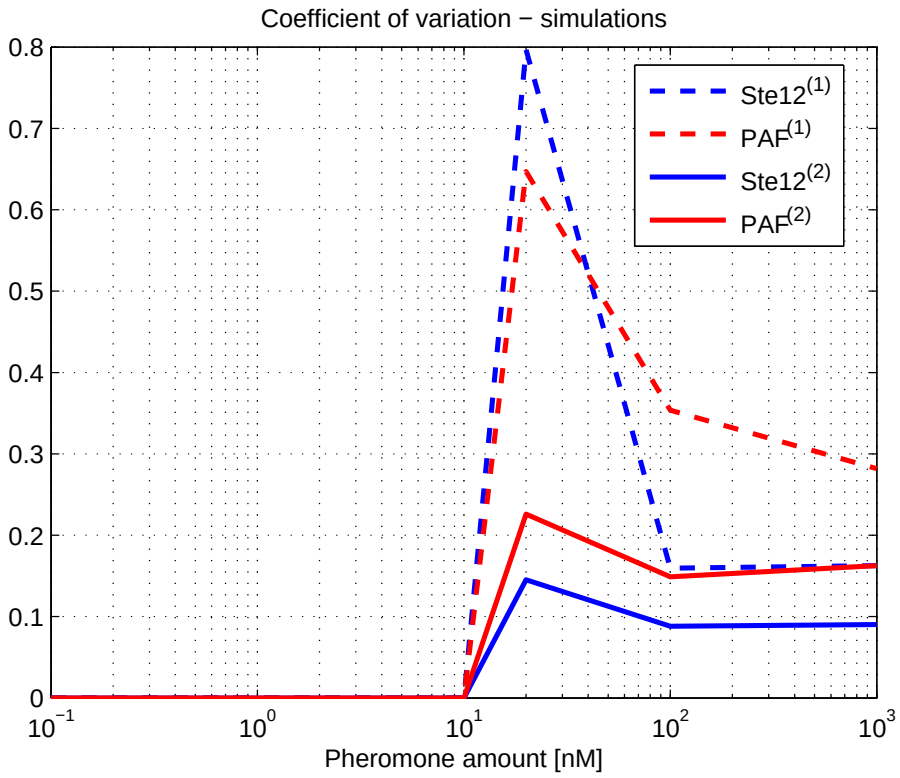

Figure 3.6: Trend of the coefficient of variation obtained from *in silico*

Results of experiments performed *in vivo*

Cultures $Ste12^{(2)}$, $Ste12^{(1)}$, $PAF^{(2)}$, $PAF^{(1)}$ (described in chapter 2.3) were assayed after two hours of induction by α-cells of the OD varying from 10^{-4} to 1. A negative control without α cells induction was also assayed. In the figure 3.7, there are dose response curves obtained from assays. The vertical axis represents the mean level of GFP fluorescence (given in relative fluorescent units).

Dose response curves of $Ste12^{(2)}$ and $Ste12^{(1)}$ have nearly the same trend and amplification, except of the difference in the response to the highest dose of α-cells. However, it can be caused by a measurement error and do not have to be significant. Therefore, it is possible to say that the decrease of *STE12* gene copy number does not affect quantitatively the ability of the transcription factor Ste12 to activate the transcription of pheromone induced genes. This result does not correspond to simulations.

Dose response curves of $PAF^{(2)}$ and $PAF^{(1)}$ obtained from assays have significantly lower amplification compared to $Ste12^{(2)}$ and $Ste12^{(1)}$. Amplification of $PAF^{(2)}$ is higher compared to amplification of $PAF^{(1)}$. The decrease in the gene copy number of PAF gene leads to lower ability to activate the pheromone induced transcription of reporter gene. In this aspect, *in vivo* experiments correspond to the experiments performed *in silico*. However, there is one difference between *in vivo* and *in silico* experiments. Concerning *in vivo* experiments, the leaky expression of $PAF^{(2)}$ (the ability to induce gene expression without the pheromone stimulation) is non-zero, significantly higher than the leaky expression of $PAF^{(1)}$ (which is lower, but non-zero as well). The leaky expression is not considered by a computational model and it is the reason why results of experiments *in vivo* and *in silico* differ. The higher leaky expression may indicate that the transcription factor PAF is not repressed by Dig1, Dig2 as much as *Ste*12. Therefore, the loss of repression due to pheromone induction does not have as strong effect as in the case of Ste12 and the amplification of dose response is lower compared to Ste12.

Significantly wide confidence intervals (which are represented by error bars in the figure 3.7) are observed in all assays $Ste12^{(2)}$, $Ste12^{(1)}$, $PAF^{(2)}$, $PAF^{(1)}$. Because error bars overlap, single dose responses are plotted separately and presented in figures 3.9-3.12. The widest confidence intervals are observed in $PAF^{(2)}$ assay. Also the coefficient of variation (figure 3.8) reaches highest value for $PAF^{(2)}$ assay. $PAF^{(1)}$ confidence intervals are much narrow compared to $PAF^{(2)}$ as well as coefficient of variation $PAF^{(1)}$ is lower compared to $PAF^{(2)}$. It suggests, as the copy number of PAF gene decrease, the variability of the response of the synthetic pathway also decreases. This finding does not correspond to simulations.

In the case of wild-type pathway, the variability in the response does not depend on the STE12 gene copy number. As the pheromone dose increases, the variability of both $Ste12^{(2)}$, $Ste12^{(1)}$ decreases and tends to the same level. The coefficient of variation of $Ste12^{(2)}$, $Ste12^{(1)}$ at the level activated by pheromone reaches the significantly lower value compared to $PAF^{(2)}$.

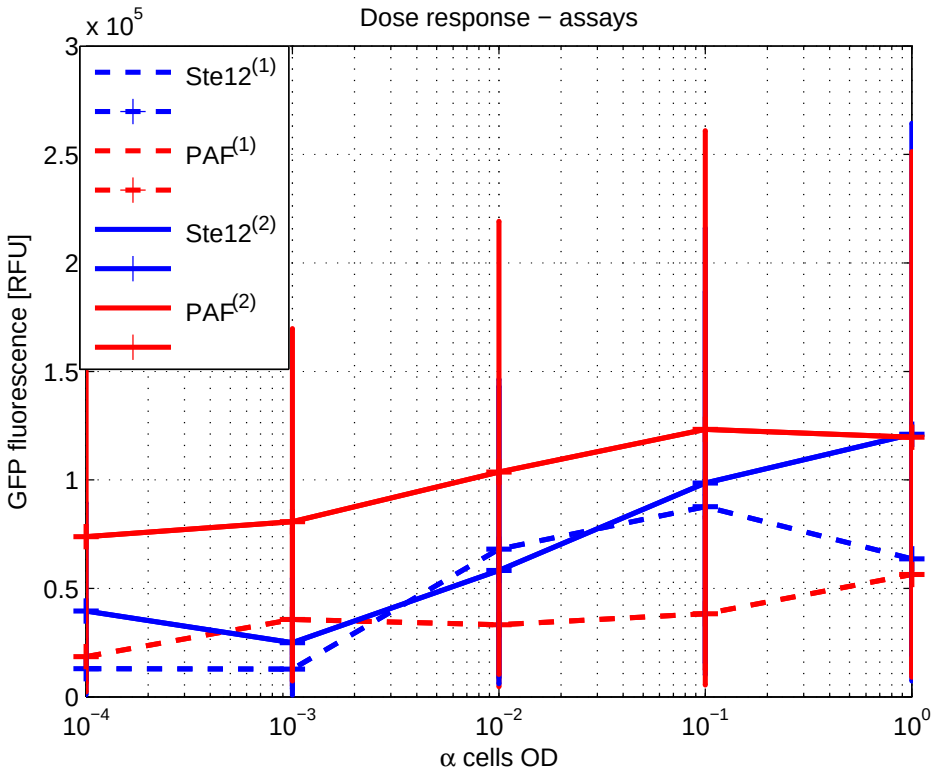

Figure 3.7: Dose response curves obtained from experiments *in vivo*

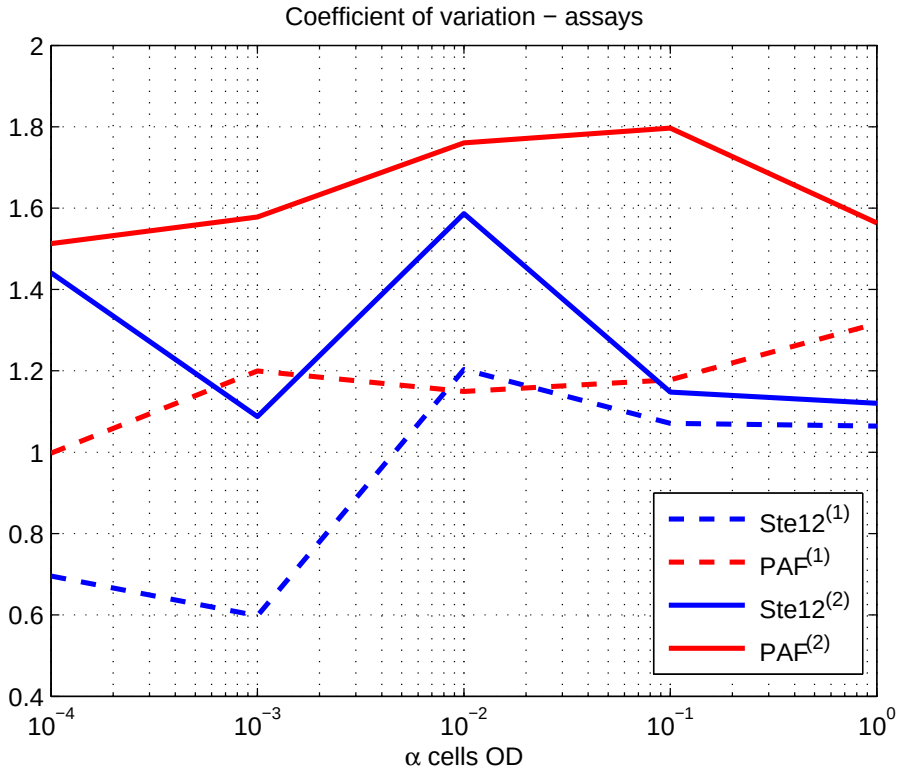

Figure 3.8: Trend of the coefficient of variation obtained from experiments *in vivo*

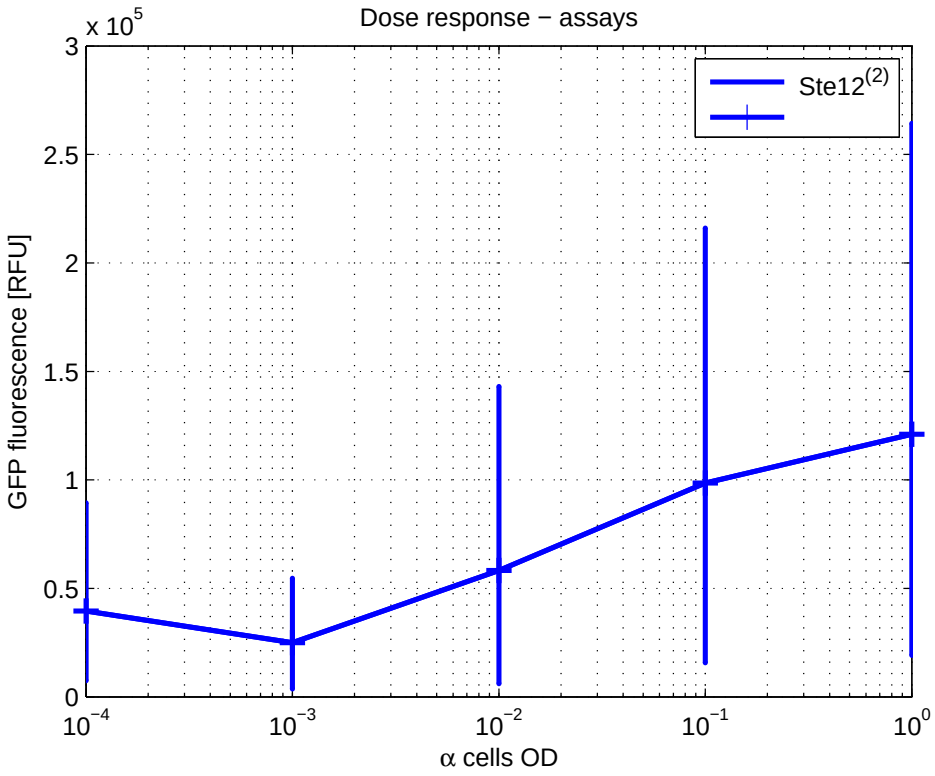

Figure 3.9: Dose respose for $Ste12^{(2)}$ obtained from experiments *in vivo*

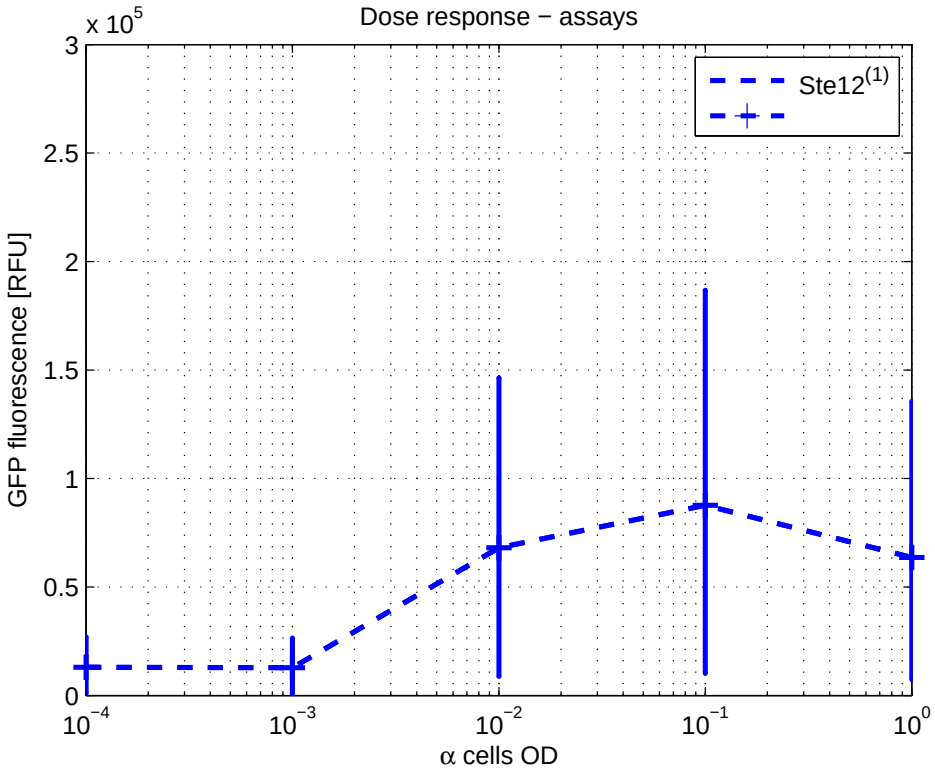

Figure 3.10: Dose respose for $Ste12^{(1)}$ obtained from experiments *in vivo*

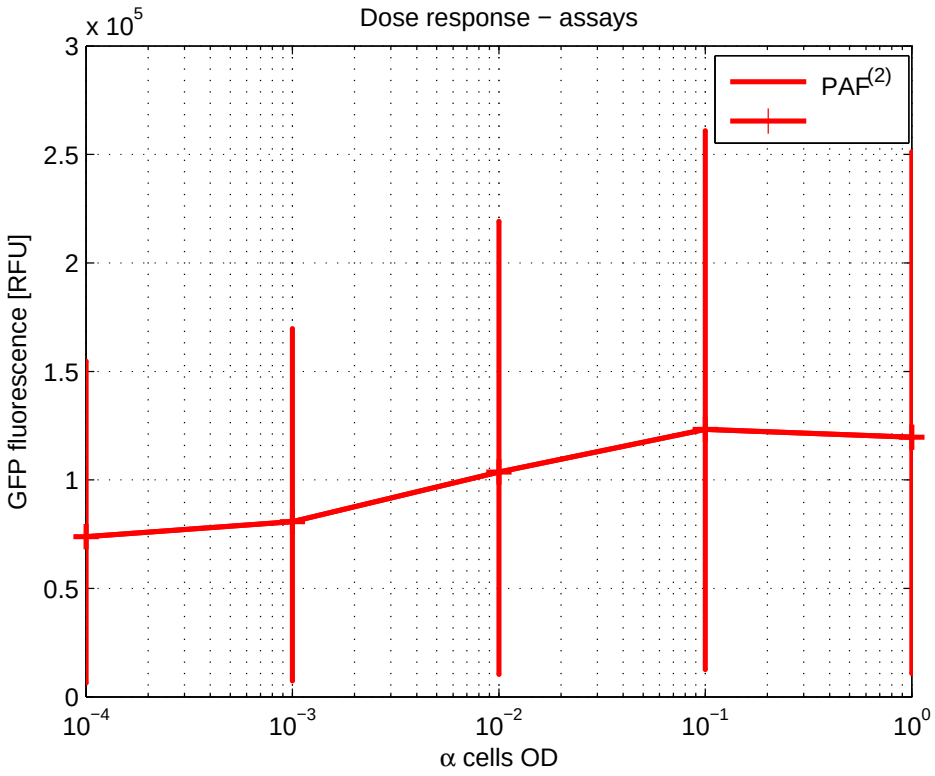

Figure 3.11: Dose respose for $PAF^{(2)}$ obtained from experiments *in vivo*

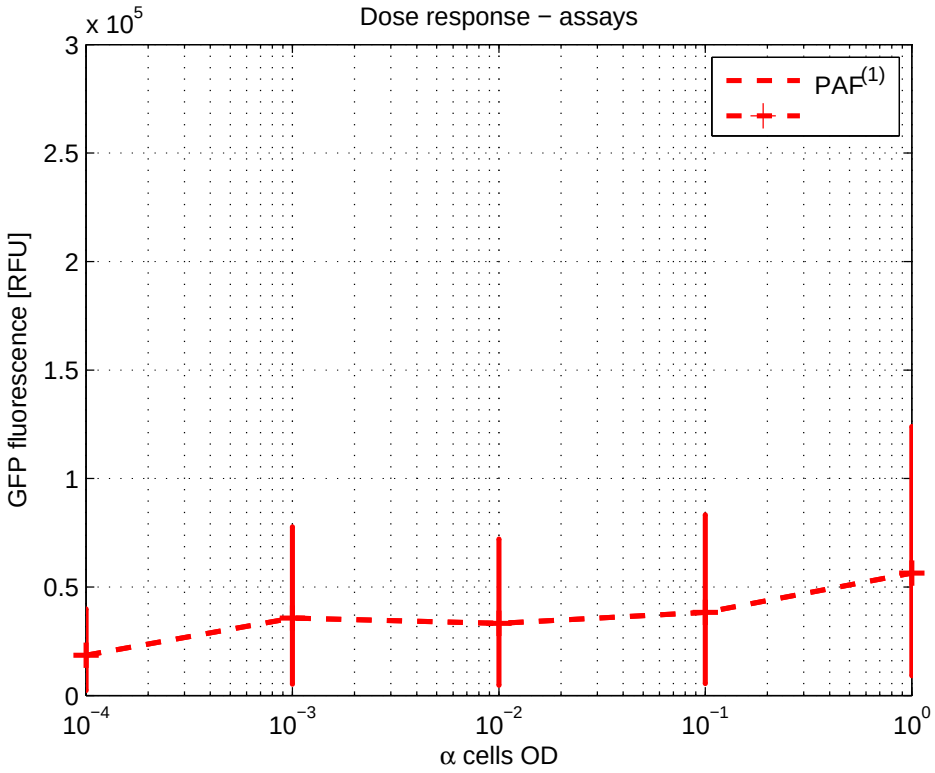

Figure 3.12: Dose respose for $PAF^{(1)}$ obtained from experiments *in vivo*

3.3 Probabilistic analysis

The probabilistic analysis was performed studying how the decrease in gene copy number affects the system. This probability is described by the expression 3.10. It refers to the probability of the state, when the output of the system with half gene copy number (designation $X^{(1)}$ refers to the output of $Ste12^{(1)}$ resp. $PAF^{(1)}$) decreases below the mean value of the output of the system with full gene copy number (designation $E(X^{(2)})$ refers to the the mean value of the output of $Ste12^{(2)}$ resp. $PAF^{(2)}$).

$$P\left(X^{(1)} < E\left(X^{(2)}\right)/2\right) \tag{3.10}$$

Results of experiments performed *in silico*

It can be seen in the figure 3.13 that after pheromone activation, the probability 3.10 is stably high for the synthetic pathway (labeled PAF) and does not decrease under the value 0.9. There is significantly high probability, that haplodeficiency in PAF causes the system output decreases under the half of the mean level of the output when full gene copy number of PAF is present in the system. It is not surprising, as the comparison of $PAF^{(2)}$ and $PAF^{(1)}$ dose responses (figure 3.5) suggests, the amplification of $PAF^{(1)}$ is more than half lower than amplification $PAF^{(1)}$. The probability 3.10 for the wild-type pathway (label Ste12) is generally lower compared to synthetic pathway. Again, it corresponds with the amplification of the $Ste12^{(1)}$ response to the pheromone is less than half lower than the amplification of the $Ste12^{(2)}$ response(figure 3.5).

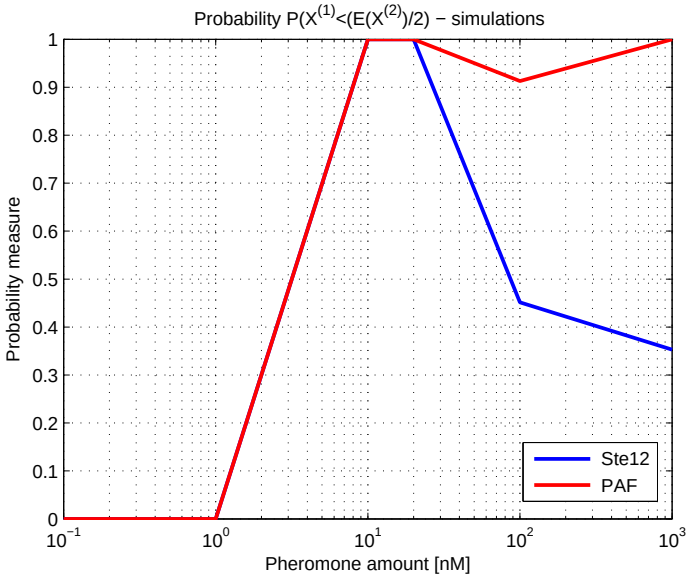

Figure 3.13: Results of probabilistic analysis obtained from experiments *in silico*

Results of experiments performed *in vivo*

Results of the probabilistic analysis (figure 3.14) suggest that the probability 3.10 is significantly lower in the case of Ste12 compared to PAF. The probability 3.10 is stably above 75% in the case of PAF while in the case of Ste12 it decreases even below 50%. It suggests the ability of Ste12 to suppress the fluctuations in signal and to be resistant towards a decrease in gene copy number.

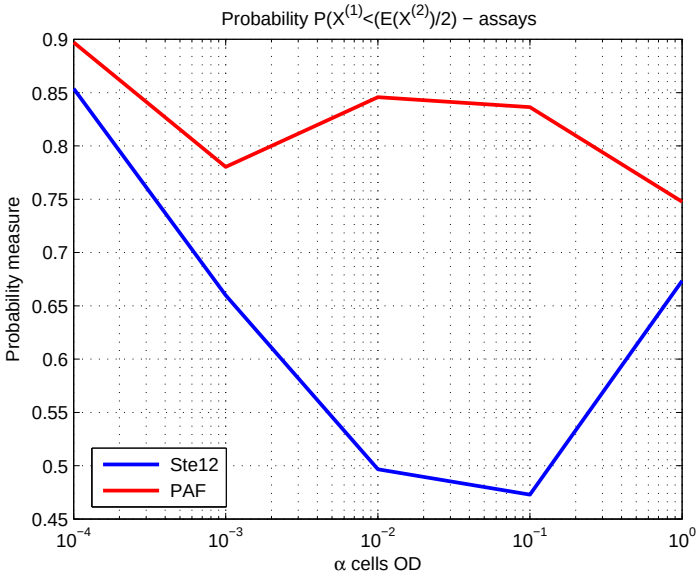

Figure 3.14: Results of probabilistic analysis obtained from experiments *in vivo*

3.4 Summary

Taken together, the decrease of the copy number of a gene encoding PAF affects the synthetic pathway by decrease of both mean value and variability of the pathway response. Full gene copy number of PAF results in higher variability of a pathway response. There is no regulation mechanism in PAF which is able to suppress the stochastic fluctuations of gene expression. Exactly opposite, it seems, PAF has an ability to amplify these fluctuations.

In the contrary, the wild-type pathway response is not significantly affected by the change of a $STE12$ gene copy number. The mean value and the variability of a pathway response are similar in both cases $Ste12^{(2)}$ and $Ste12^{(1)}$. Moreover, the variability of a pathway response decreases at the level of high pheromone induction. It is in direct contradiction with results of simulations. It suggests the wild-type Ste12 transcription factor has some regulation mechanism which suppresses the effect of stochastic fluctuations in the signal.

This mechanism is not considered by the model. The model considers the Ste12 binds to the promoters of upstream genes in the yeast pheromone pathway - the positive feedback from Ste12. However, the negative feedback must be carried out by Ste12 as well. It would not be possible to correct the fluctuations of

a system without the negative feedback [26]. It seems, the Ste12 transcription factor behaves as the filter of stochastic fluctuations in the signal, especially at the higher level of pheromone induction.

Results of data analysis confirms the thesis that crucial components of biological circuits are noise resistant. The synthetic transcription factor PAF does not have this noise-resistant mechanism. It suggests that any intervention in the natural cell environment can damage some important function of complex biological system.

Conclusion

The aim of my thesis was to characterize the stochasticity of the wild-type and synthetic signal pheromone pathway response to different doses of pheromone input. The difference of the wild-type and synthetic signal pheromone pathway was in crucial component of the pathway – the nuclear transcription factor. The wild-type signal pheromone pathway contained the wild-type transcription factor Ste12 while the synthetic signal pheromone pathway contained the hybrid transcription factor PAF. My aim was to analyze how the change of the transcription factor affects the stochasticity of the pathway response to varying pheromone input.

In the theoretical part of my thesis, I was trying to find a context of gene expression noise and human diseases. The direct impication from noise to disease state is quite hard to find. However, the possible role of noise in the process of disease onset was found to be discussed in professional literature. This theory suggests, that nois can play the role of switch-on mechanism of the disease state under the specific conditions. One of these conditions is a loss of one functional copy of the specific gene. This finding suggested one more aim of my thesis, namely, to analyze how the response of the wild-type respective synthetic pathway changes, when the gene encoding the responding transcriptional factor is present in full (2) respective half (1) gene copy number.

Consequently, in order to achieve the aims of my thesis, in silico and in vivo experiments were performed. There was found a significant difference between results of in silico and in vivo experiments. It was found that the real wild-type Ste12 transcription factor has the ability to suppress the effect of stochastic fluctuations in the signal, especially at the higher level of pheromone induction. It suggests the Ste12 transcriptional factor behaves as a filter of biological noise. Moreover, this ability of the Ste12 transcriptional factor is independent on the $STE12$ gene copy number. It means the Ste12 performs the robust behavior towards to change of parameters of the pathway. This robust behavior is probably due to negative feedback mediated by Ste12 itself. The computational model did not consider the regulation mechanism involved in real wild-type pathway. The synthetic transcription factor PAF does not have such a mechanism and therefore it is unable to suppress the effect of stochastic fluctuations in the signal. Exactly opposite, it seems, PAF has an ability to amplify these fluctuations when full PAF gene copy number is present.

Results of my thesis support the assumption that crucial components of natural biological circuits carry out the robust and noise resistant behaviour. It

may be advantageous for people to perform genetic modifications of natural systems. However, it should be remembered that any intervention in the natural cell environment can damage some important function of complex biological system.

Appendix A

Materials and methods

Yeast strains and plasmids Yeast strains used for experiments in vivo were: 6194 *(FY 23::ura3-52leu2Δ1 his3Δ200 MATα)* in case of α-cells used for induction. MLY215 Δ *pde2 :: G418Δste12 :: leu2 :: hisG*Δ *leu2 :: hisG ura3-52 MATa)* in case of a-cells used for all Ste12 and PAF assays. Plasmids with inserts carrying the gene encoding the particular transcription factor were: pRS416-pLAC13-STE12, pRS416-pTET20-STE12 for Ste12 assays, pRS416-pLAC13-PAF, pRS416-pTET20-PAF for PAF assays. Plasmids with inserts carrying the reporter gene tGFP were: pRS416-pFUS1-tGFP for Ste12 assays pRS416 - pGAL1-tGFP for PAF assays. Transformations were made using the High efficient yeast transformation protocol. Colonies after transformation were streaked on selection plates with appropriate amino acids.

Cytometric assay Cultures of a-cells and α-cells in SD —ura —leu dropout medium were prepared and grown for 24 hours. Cultures of a-cells resp. α-cells were diluted to OD 0.2 resp. 0.4 and were grown for 4 hours before the induction. After 2 hours of induction, samples were diluted to OD 0.1 and the fluorescence assay was performed. A negative control, sample without α-cells induction was also assayed. The flow cytometer BD Accuri C6 was used for the fluorescence assay. The fluorescence of samples was assayed at wavelength 530 nm (FL1-A data set obtained from CFlow software) corresponding to the wavelength of the radiation of the tGFP protein used as a reporter.

Bibliography

[1] Helena Handrková. O signálních drahách obecně. [online], 2005, [cited 2016-08-10] Availble at http://cellula.wz.cz/sigob.html,.

[2] Lee Bardwell. A walk-through of the yeast mating pheromone response pathway. *Peptides*, June 2006.

[3] Juan M. Pedraza and Alexander van Oudenaarden. Noise propagation in gene networks. *Science*, 5:1965–1969, March 2005.

[4] Zhi Wang and Jianzhi Zhang. Regression models and life-tables (with Discussion). *PNAS*, April 2011.

[5] William J Bosl and Rong Li. The role of noise and positive feedback in the onset of autosomal dominant diseases. *BioMed Central*, June 2010.

[6] Polycystická choroba ledvin. [online], [cited 2016-08-10] Availble at http://www.ledviny.cz/nemoc-polycysticka-choroba-ledvin.

[7] Miroslav Merta, Jana Reiterová, Jitka Štekrová. Polycystická choroba ledvin. [online], 2007, [cited 2016-08-10] Availble at http://www.internimedicina.cz/pdfs/int/2007/06/08.pdf.

[8] Xia S Johnson T Wallace DP Calvet JP Li R Li X, Magenheimer BS. A tumor necrosis factor-alpha-mediated pathway promoting autosomal dominant polycystic kidney disease. *Nat Med*, pages 863–868., June 2008.

[9] Tomáš Edelsberger. MODY [online], 2014, [cited 2016-08-10] Availble at http://www.lecbacukrovky.cz/slovnicek/mody,.

[10] Fred Sherman. Getting Started with Yeast. [online] 2002, cited[2016-08-10], Availble at https://instruct.uwo.ca/biology/3596a/startedyeast.pdf.

[11] Vladimír Zicháček Jan Jelínek. *Biologie pro gymnázia*. Vydání třetí. Olomouc, Praha, 2003, isbn 978-80-246-2173-9.

[12] The yeast life cycle. [online], 2016, cited[2016-08-10], Availble at http://www.singerinstruments.com/resource/what-is-yeast.

[13] Tereza Puchrová. Modelling and experimental validation of signalling pathways with relevance to homologous mammalian systems. Master Thesis. University of West Bohemia, September 2015.

[14] Cheng-Ting Chien Haiwei Pi and Stanley Fields. Transcriptional activation upon pheromone stimulation mediated by a small domain of *Saccharomyces cerevisiae* Ste12p. *Molecular and Cellular Biology*, page 6410–6418, 1997.

[15] Anna Sosnová. Identification and modulation of the pheromone response pathway step response. Bachelor Thesis. University of West Bohemia, May 2016.

[16] Gal4. [online] 2016, cited[2016-08-10], Availble at http://www.uniprot.org/uniprot/p04386.

[17] Ryan Suderman and Eric J. Deeds. Machines vs. ensembles: Efective mapk signaling through heterogeneous sets of protein complexes. *PLoS Computational Biology*, 2013.

[18] Rulebender-1.1.415-tutorial. [online], 2015, cited[2016-08-10], Availble at http://visualizlab.org/rulebender/tutorial.html.

[19] J.J. Tapia J.R.Farder G.E.Marai J. Wenskovitch, L.A.Harris. MOSBIE:A Tool for Comparision and Analysis of Rule-Based Biochemical Models. *BMC Bioinformatics Journal*, 15:1–22, 2014.

[20] James R Faeder Michael W Sneddon and Thierry Emonet. Efficient modeling, simulation and coarse-graining of biological complexity with nfsim. *Nature Methods*, page 177–185, February 2011.

[21] Jiří Dřímal, David Trunec, Antonín Brablec. ÚVOD DO METODY MONTE CARLO.[online] duben 2006, cited[2016-08-10], Availble at http://www.physics.muni.cz/ trunec/mc.pdf.

[22] Daniel Georgiev. Přednášky z předmětu Modelování a simulace 1.[online] 2014, cited[2016-08-10], Availble at http://ccy.zcu.cz/files/MS2_2014/Lectures/Lecture8_2014.pdf,.

[23] Jiří Militký Milan Meloun. *Interaktivní statistická analýza dat*. Vydání třetí. Karolinum, Praha, 2012, isbn 978-80-246-2173-9.

[24] Dose-response -relationship. [online], 2016, [cited 2016-08-10] Availble athttp:https://en.wikipedia.org/wiki/dose–response_relationship.

[25] Míry variability. [online] 2014, cited[2016-08-21], Availble at http://www.wikiskripta.eu/index.php/míry_variability.

[26] Jiří melichar. Lineární systémy 1, skriptum str.77 [online] 2010, cited[2016-08-10], Availble at http://www.kky.zcu.cz/uploads/courses/ls1/ls1-ucebni-texty-2010.pdf.