

MASARYK UNIVERSITY
FACULTY OF INFORMATICS



Modelling of cell signalling pathways by using boolean networks

BACHELOR'S THESIS

Kateřina Hemalov

Brno, 2015

Declaration

Hereby I declare, that this paper is my original authorial work, which I have worked out by my own. All sources, references and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

Kateřina Hemalov

Advisor: RNDr. David řafrnek, Ph.D.

Acknowledgement

I would like to thank my advisor RNDr. David Šafránek, Ph.D. for guidance and many consultations. I would also like to thank Mgr. Pavel Krejčí, Ph.D. for providing me with experimental data and Vojtěch Havel for critical reading of this thesis.

Abstract

In this thesis, we study the characteristics of boolean networks which mimic signalling pathways. We use the FGF (Fibroblast Growth Factor) pathway as a case study. Exact molecular interactions and feedbacks in the FGF pathway are unclear. Therefore, the boolean modelling is suitable formalism for its analysis. Based on literature we have built boolean models of the FGF pathway.

The activity of the FGF pathway can have either transient or sustained profile. However, the molecular mechanism responsible for differences in the pathway activity is unknown. We have expressed this two qualities formally in LTL. Using the NuSMV model checker we have analysed our models for given qualities. Based on our analysis we suggest that the negative feedbacks in the FGF pathway regulate the pathway activity.

To comprehensively visualize the model dynamics we have implemented a DFS algorithm which depicts in gnuplot all paths in the state transition graph reachable from an initial node.

Keywords

Cell Signalling Pathways, Boolean Networks, FGF Pathway, Transient Activity, Sustained Activity

Contents

1	Introduction	1
2	FGF2 signalling pathway	3
2.1	<i>Biological background of FGF2 signalling pathway</i>	3
2.2	<i>Behaviour of FGF2 signalling</i>	7
3	Methods	9
3.1	<i>Boolean models</i>	9
3.1.1	<i>Boolean models definition</i>	9
3.1.2	<i>Boolean models example</i>	10
3.1.3	<i>Boolean models semantics</i>	11
3.2	<i>Linear temporal logic</i>	12
3.3	<i>GINSim</i>	13
3.4	<i>NuSMV</i>	14
3.5	<i>Implementation of DFS algorithm</i>	14
3.6	<i>Experimental data</i>	16
4	Results	20
4.1	<i>Analysed properties</i>	20
4.2	<i>Models</i>	20
4.3	<i>Model \mathcal{M}_1</i>	21
4.3.1	<i>\mathcal{M}_1 behaviour</i>	21
4.4	<i>Model \mathcal{M}_2</i>	23
4.4.1	<i>\mathcal{M}_2 behaviour</i>	24
4.5	<i>Model \mathcal{M}_3</i>	24
4.5.1	<i>\mathcal{M}_3 behaviour</i>	24
4.6	<i>Model \mathcal{M}_4</i>	24
4.7	<i>Model \mathcal{M}_5</i>	25
4.8	<i>Summary of analysis</i>	27
5	Discussion	29
6	Conclusion	31
A	Attached files	36
B	Signalling pathways	37

1 Introduction

Surviving of living cells depends on their ability to perceive and correctly response to their environment. Cells accept chemical signals by receptors on their plasma membrane. The receptors trigger chain of biochemical reactions which result in complex cellular response. The process of spreading the signal on particular molecules is called signalling pathway.

Signalling pathway can be viewed as a network of mutual interactions among signalling molecules. It is a complex system which is not intuitive due to various feedbacks and crosstalks. Recently, the computational modelling is more often used to analyse the signalling pathways structure and behaviour. We can use the computational models to make hypotheses about the role of signalling molecules in diseases and to make predictions which can be experimentally tested.

Continuous models using ODE (Ordinary Differential Equations) provide accurate system analysis. However, their accuracy strongly depends on the exact molecular concentrations and reaction rates which are often unknown due to the limitations of today's experimental techniques. Therefore, in practice most of the model parameters are based on estimation and inference. In many cases, we can clarify the basic system mechanisms without the need of complicated continuous model, by using logical models which discretize the unknown continuous parameters.

In this thesis, we study the characteristics of boolean networks which mimic signalling pathway. We use the FGF (Fibroblast Growth Factor) pathway as a case study. Exact molecular interactions and feedbacks in the FGF pathway are unclear and thus the boolean modelling is suitable formalism for its analysis.

Cells respond differentially depending on how long the FGF pathway is active [16]. The pathway activity is either transient (approximately 2 hours) or sustained (more than 12 hours). The molecular mechanism responsible for different duration of FGF pathway activity is not known.

To address this issue, we have built several models of FGF pathway based on the literature and unpublished experimental data pro-

duced by Pavel Krejčí's group from the Department of Experimental Biology MU. The models explain the behaviour of the pathway on the logical level.

2 FGF2 signalling pathway

FGF signalling pathway is one of the most important pathway in human physiology and development. During embryonic stage, it is involved in cell proliferation, differentiation and migration. In adults, it plays an important role in nervous system, tissue repair and tumor angiogenesis[10].

Considering the number of biological processes that are regulated by FGF pathway, it is clear that deregulation of the pathway has severe impact on the organism. Mutations in FGF receptors (FGFRs) were identified in a variety of human cancers as well as variety of human skeletal dysplasias[1, 11].

The FGF protein family has 22 members and is recognized by 4 receptors (FGFRs). In this thesis, we focus on FGF2 and its receptor FGFR3.

2.1 Biological background of FGF2 signalling pathway

The following section introduces main proteins which are involved in FGF2 signalling. They are summarised in the figure 2.1.

FGF2

FGF2 is external signal which binds to specific membrane receptors and triggers the signal transduction in the cell. FGF2 activates a spectrum of signal transduction pathways. Among those, MAPK pathway which is necessary for the growth arrest [16].

FGFR3

FGF receptor 3 (FGFR3) is a cell surface receptor. It has extracellular domain with high affinity towards FGF2 and intracellular domain with tyrosine kinase activity which in active state gains the ability to phosphorylate other proteins on tyrosine residues.

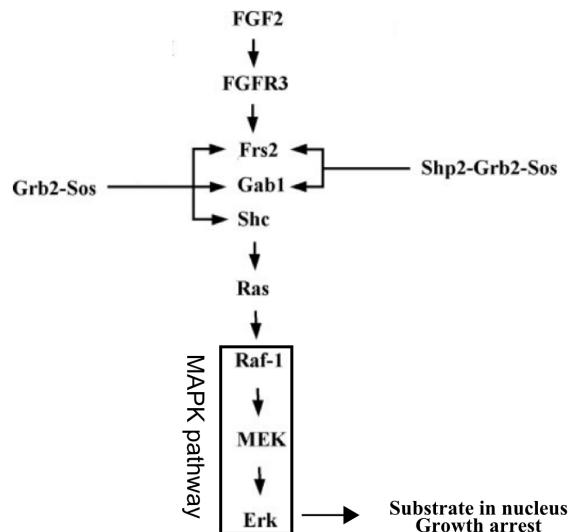


Figure 2.1: Schema of FGF2 pathway[17]. FGF2 activates Erk via the FGFR3/Ras/MAPK pathway. FGFR3 utilizes at least three adaptors (Frs2, Gab1, and Shc) to recruit Sos in order to activate Ras. Sos is complexed with Grb2 and also might be complexed with Shp2. Both Frs2 and Gab1 recruit Shp2-Grb2-Sos and Grb2-Sos complexes, whereas Shc recruits mostly Grb2-Sos.

FGF2 binds to FGFR3 and promotes the formation of FGFR3 dimer [14]. The FGFR3 changes its conformation which triggers its tyrosine kinase activity [20]. It leads to transphosphorylation of key tyrosine residues on receptor intracellular part. The activated receptor then phosphorylates multiple intracellular proteins, including adaptor proteins Frs2, Shc and Gab1.

Signal is propagated downstream by the recruitment of signalling complexes Grb2-Sos or Shp2-Grb2-Sos which consequently activate MAPK pathway[17]. The complexes bind to Frs2, Shc and Gab1 or less frequently they bind directly to FGFR3.

Frs2

Fibroblast Growth Factor Receptor Substrate 2 (Frs2) functions as a major mediator of signalling via FGFRs [18]. The Figure 2.2 depicts

the structure of Frs2 protein. It has multiple tyrosine residues which are phosphorylated by FGFR3. These residues serve as binding sites for downstream molecules Grb2 and Shp2. Shp2-Frs2 binding leads to MAPK pathway activation. The Grb2-Frs2 binding mostly leads to the activation of a different pathway (PI-3 kinase) and plays only a secondary role in activation of the MAPK pathway.

Negative feedback on Frs2

Apart from tyrosine residues, phosphorylation on Frs2 also occurs on several threonine residues. This phosphorylation is mediated by the downstream kinase ERK and has an inhibitory effect on Frs2. Frs2 threonine phosphorylation results in reduced tyrosine phosphorylation and decreased recruitment of downstream molecules and consequent attenuation of the MAPK pathway [18].

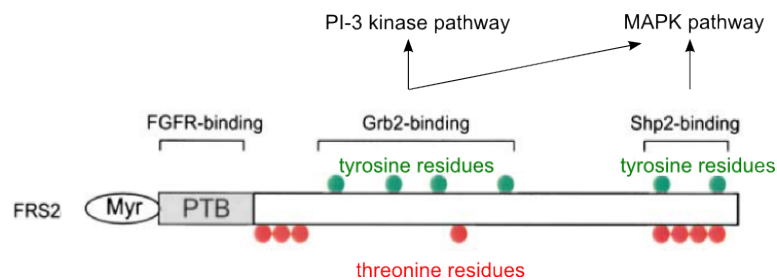


Figure 2.2: Structure of Frs2 docking protein [18]. Myristyl group (Myr) anchors the Frs2 in the plasma membrane. PTB (phosphotyrosine binding domain) binds to FGFR3. Several tyrosine residues function as binding sites for downstream molecules Grb2 and Shp2. Several threonine residues are crucial for negative feedback loop.

Shc

Shc is an adaptor protein that binds through SH2 domain to FGFR3 and forms complex with Grb2-Sos under FGF2 stimulation [17].

Gab1

GRB2-associated-binding protein 1 (Gab1) is adaptor protein that binds Shp2-Grb2-Sos complex under the FGF2 stimulation[17].

Grb2

Growth factor receptor-bound protein 2 (Grb2) is an adaptor protein composed of one SH2 domain and 2 SH3 domains [19]. The SH2 domain binds to phosphotyrosine on receptors or adaptor proteins, e.g. Frs2 and Shc. The SH3 domains bind to several proline-rich peptides, manely Sos.

Sos

Son of sevenless (Sos) is guanine nucleotide exchange factor that activates the Ras protein by catalyzing formation of Ras-GTP complex[9]. Sos is constitutively bound to Grb2 in a complex Grb2-Sos.

Negative feedback on Sos

Several Sos serin residues are phosphorylated by ERK[8]. The phosphorylation results in disruption of the Grb2-Sos complex and inhibition of Grb2-Sos binding to adaptors.

Shp2

Shp2 is involved in many signalling pathways, including FGF pathway, and interacts with various proteins [18]. It contains 2 SH2 domains which bind Frs2. FGF stimulation leads to tyrosine phosphorylation of Shp2 and association with the Grb2-Sos complex [2].

Ras

Ras is a GTP binding protein which is connected to plasma membrane. It is active if it binds GTP and inactive if it binds GDP. Active

Ras is recognized by Raf protein [5]. They formate a complex which leads to activation of MAPK pathway.

MAPK pathway

The MAPK pathway represents a major pathway utilized by FGF2 signalling [11]. It consists of serin/threonin kinases RAF, MEK, ERK which transmit signal downstream in cascade-like manner. RAF phosphorylates MEK and afterwards phosphorylated MEK phosphorylates ERK. Phoshorylation of ERK results in transcription of genes involved in cell growth and division.

P-ERK provides several negative feedbacks that attenuate FGF signalling and thereby protects the cell from excessive activation of FGFR[8, 18].

2.2 Behaviour of FGF2 signalling

The overtime behaviour of FGF2 pathway can be measured as a level of phosphorylated downstream molecules, e.g. phosphorylated ERK (P-ERK) [16]. The cell response depends on how long the P-ERK level is increased.

In most of the cell types, the pathway activity is transient (approximately 2 hours). In this time range the level of P-ERK rapidly increases and then decreases to minimum (Figure 2.4a).

In chondrocytes, the cell type responsible for bone growth, the pathway activity is sustained (Figure 2.4b). The P-ERK level increases and then stays increased (with slight fluctuations) in the range of 12 hours (Figure 2.3). The sustained increased P-ERK level leads to growth arrest. The molecular mechanism responsible for differences in the FGF2 pathway activity is not clear.

2. FGF2 SIGNALLING PATHWAY

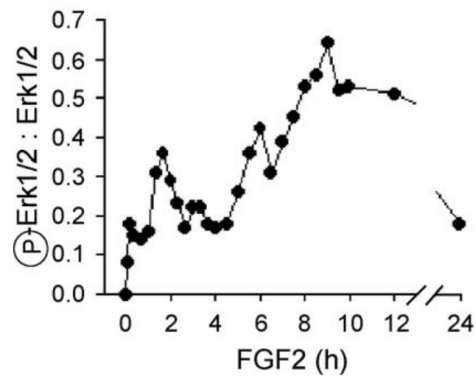


Figure 2.3: Sustained behaviour of the FGF pathway in chondrocyte-like cells [16]. The level of P-ERK was measured in the 24 hours long range of incubation with FGF2. The data are expressed as ratio between the optical density of P-Erk1/2 and total Erk1/2.

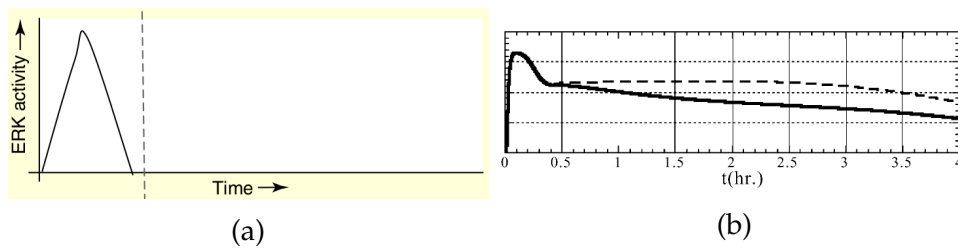


Figure 2.4: Schematic time course of ERK activity. (a) Transient ERK activity [23]. (b) Sustained ERK activity [27].

3 Methods

3.1 Boolean models

In this theses, we use the boolean model definition as it is described in the paper by Barnat *et al.* [4]. The definition comes from Thomas' extension of boolean networks which is implemented in GINsim tool [24, 22]. The framework was designed to model gene regulatory networks (GRN) but it is nevertheless suitable for modelling of signalling pathways as well since the GRN and signalling pathways share basic principles. The approach has been applied to modelling signalling pathways in the studies [25, 13].

GRN is a collection of genes which interact with each other by gene products (RNA or proteins). Thus it can be presented as a network where nodes are genes and edges are regulatory interactions. Each node can gain binary value. It is either *on* (the gene is active, transcribed) or *off* (the gene is inactive, nottranscribed).

In case of signalling pathway, the nodes represent signalling proteins and the edges represent regulations, mostly phosphorylation or dephosphorylation. The proteins are usually in *on* state when they are phosphorylated and in *off* state when they are dephosphorylated, but it can also be vice versa.

3.1.1 Boolean models definition

Here we present the boolean model definition according to Barnat *et al.* [4]. The boolean model is formally described as a tuple $\mathcal{B} = \langle G, \sigma, \theta, \rho, L \rangle$ where

- $G = (V, E)$ is a directed regulatory graph with vertices $V = \{g_1, \dots, g_n\}$ denoting *proteins of signalling pathway* and set of edges $E \subseteq V \times V$ denoting *regulations*.
- $\sigma(e) \in \{+, -\}$ denotes the type of regulation $e \in E$: positive (+) or negative (-),
- $\theta(e) \in \mathbb{N}_{\geq 1}$ denotes the activation *threshold* of $e \in E$,

- $\rho(g_i) \in \mathbb{N}_{\geq 1}$ denotes the *maximum activity level* of $g_i \in V$ determining the activity domain $\{0, \dots, \rho(g_i)\}$,
- L is *regulatory logic* defined as $L = \{K_{i,R} \mid 1 \leq i \leq n, R \subseteq \{v \in V \mid \langle v, g_i \rangle \in E\}\}$ where $K_{i,R}$ denotes *target activity level* of g_i when regulated by all signalling proteins in R , $0 \leq K_{i,R} \leq \rho(g_i)$.

3.1.2 Boolean models example

The most suitable for modelling signalling pathways is a special type of boolean models, the purely boolean models. It has the maximum activity level of all vertices set to 1 [4]. Figure 3.1a depicts an example of simple purely boolean model $\mathcal{B}' = \langle G, \sigma, \theta, \rho, L \rangle$ where

- $G = (V, E)$ is the graph with the vertices $V = \{A, B\}$ and the edges $E = \{\langle A, A \rangle, \langle A, B \rangle, \langle B, A \rangle\}$.
- $\sigma(\langle A, A \rangle) = +$
 $\sigma(\langle A, B \rangle) = +$
 $\sigma(\langle B, A \rangle) = -$
- $\theta(\langle A, A \rangle) = 1$
 $\theta(\langle A, B \rangle) = 1$
 $\theta(\langle B, A \rangle) = 1$
- $\rho(A) = 1$
 $\rho(B) = 1$
- L is regulatory logic defined as the following set of rules
 $K_A, \emptyset = 1$
 $K_A, \{A\} = 1$
 $K_A, \{A, B\} = 0$
 $K_A, \{B\} = 0$
 $K_B, \emptyset = 0$
 $K_B, \{A\} = 1$

The rule $K_B, \{A\} = 1$ declares that when activity level of A is higher than threshold then it activates B . The rule $K_A, \{A\} = 1$ declares autoactivation of A . The rule $K_A, \{A, B\} = 0$ defines the activity A when interactions with both A and B are active. And the rules

$K_A, \emptyset = 1$ and $K_B, \emptyset = 0$ define the basal activity level of A and B which corresponds to the situation when all incoming interactions are inactive.

The regulatory logic we choose here is one of 2^6 possible parametrizations since each regulatory rule can be set to 0 or 1.

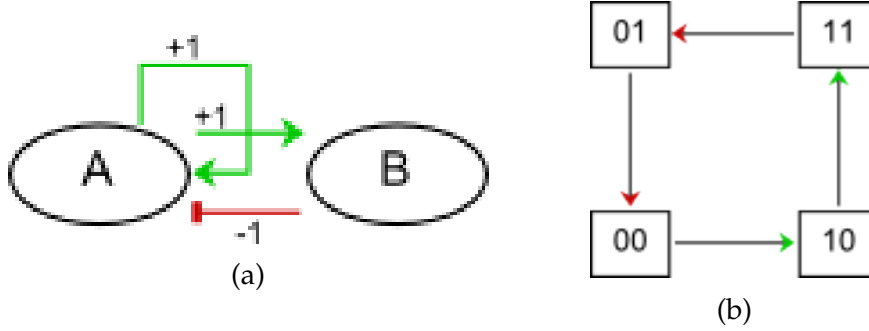


Figure 3.1: (a) Simple boolean model example. (b) Dynamics of the model depicted by the transition graph.

3.1.3 Boolean models semantics

The directed regulatory graph of a boolean model \mathcal{B} consists of nodes g_0, \dots, g_n . The node g_i can gain one of several different values $0, \dots, \rho(g_i)$. A state of the model \mathcal{B} is n-tuple of the values of all nodes \mathcal{B} [4].

For example the model \mathcal{B}' from 3.1.2 consists of two nodes A and B which both can gain the value of either 0 or 1. Therefore, the model \mathcal{B}' can occur in four distinct states 00, 10, 11, 01. The semantics of the model \mathcal{B}' can be depicted by transition graph where the nodes denote the states and the edges denote the transitions between them (figure 3.1b). The construction of transitions is defined later in the text.

Formally the semantics of a model \mathcal{B} is defined as the tuple $BTS(\mathcal{B}) = \langle S, T, S_0 \rangle$ where

- $S = \prod_{i=1}^n \{0, \dots, \rho(g_i)\}$ is the state space.
- $S_0 \subseteq S$ is the set of specified initial states.
- $T \subseteq S \times S$ is the transition relation which is defined later.

The state space is the set of all distinct states of the model. The transition relation determines the transitions between the states in the transition graph. Several strategies how to construct the transition relation have been developed [12]. In this thesis, we use the nondeterministic asynchronous approach which allows a change of the activity level of only one signalling protein in a single transition [26, 15].

The transition relation is defined as follows [4]. First we denote the level of g_i in the state $s \in S$ by $l_i(s)$. Assume there is a regulation $e = \langle g_i, g_j \rangle \in E$ between genes g_i, g_j . We say that g_i is a *resource* for g_j in s if $\sigma(e) = +$ and $l_i(s) \geq \theta(e)$, or $\sigma(e) = -$ and $l_i(s) < \theta(e)$. Let $Re(s, g_i)$ denote the set of all resources of g_i in s . There is transition $s \rightarrow s'$ according to the following rules:

- If there exists u such that $K_{u, Re(s, g_u)} > l_u(s)$ then $l_u(s') = l_u(s) + 1$.
- If there exists u such that $K_{u, Re(s, g_u)} < l_u(s)$ then $l_u(s') = l_u(s) - 1$.

3.2 Linear temporal logic

In this section we define the Linear Temporal Logic (LTL) which we use to formally describe characteristics of boolean model behaviour.

First we denote $P(\mathcal{B})$ the set of all runs π in the transition graph $BTS(\mathcal{B})$. Each infinite run $\pi = s_0, s_1, \dots \in BTS(\mathcal{B})$ belongs to $P(\mathcal{B})$. For each finite run $s_0, \dots, s_n \in BTS(\mathcal{B})$ such that there is no outgoing edge from s_n , there is an infinite run $\pi = s_0, \dots, s_n, s_n, s_n, \dots \in P(\mathcal{B})$.

Let π^i for $\pi = s_0, s_1, \dots$ denote the run s_i, s_{i+1}, \dots where first i states are removed and π_0 denote the first state of the run π .

Syntax of LTL is defined by the following equation in BNF:

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg \varphi \mid \mathcal{X}\varphi \mid \varphi \mathcal{U} \varphi,$$

where p belongs to the set of atomic propositions AP [7]. Atomic propositions are statements in form $g_i \circ t$ where $\circ \in \{=, >, <, \leq, \geq\}$ and g_i belongs to set of the vertices of the regulatory graph of model \mathcal{B} and t belongs to the activity domain $\{0, \dots, \rho(g_i)\}$.

Semantics of LTL is defined over the set of all runs $P(\mathcal{B})$ as follows [7]. We say that a run $\pi = s_0, s_1, \dots$ satisfies an LTL formula, written $\pi \models \varphi$, if and only if:

- $\pi \models p$ iff $p \in L(\pi_0)$ for $p \in \mathcal{AP}$ where $L(s)$ is a set of all atomic propositions $g_i \circ t$ such that the constrain $g_i \circ t$ holds in the state s .
- $\pi \models \varphi \wedge \psi$ iff $\pi \models \varphi$ and $\pi \models \psi$
- $\pi \models \mathcal{X}\varphi$ iff $\pi^1 \models \varphi$
- $\pi \models \varphi \mathcal{U} \psi$ iff $\exists k \geq 0$ such as $\pi^k \models \psi$ and $\forall i \in \{0, \dots, k-1\}$ holds $\pi^i \models \varphi$
- $\pi \models \neg\varphi$ iff not $\pi \models \varphi$

In addition, LTL can be extended with the temporal operators $\mathcal{F}\varphi \equiv \text{true} \mathcal{U} \varphi$ and $\mathcal{G}\varphi \equiv \neg \mathcal{F} \neg \varphi$ and with the boolean connectives \vee, \Rightarrow where $\varphi \vee \psi \equiv \neg(\neg\varphi \wedge \neg\psi)$ and $\varphi \Rightarrow \psi \equiv \neg\varphi \vee \psi$.

LTL can be directly interpreted over $P(\mathcal{B})$. Given a dynamic system $BTS(\mathcal{B})$ with a particular set of initial states S_0 we can say that $BTS(\mathcal{B})$ satisfies a formula φ , written $BTS(\mathcal{B}) \models \varphi$, only if for each $s \in S_0$ all runs $\pi \in P(\mathcal{B})$ such that $\pi_0 = s$ satisfy φ [3].

Automated process of verifying that a given model satisfies a given logical formula is called model checking[7].

3.3 GINSim

GINsim (Gene Interaction Network simulation) is a computer tool for the modelling and simulation of GRN which implements the Thomas' extension of boolean networks [22, 24]. GINSim leans on two main types of graphs: logical regulatory graphs, which model regulatory networks, and state transition graphs, which represent their dynamical behaviour.

GINsim offers several tools to analyse the state transition graph. The tool "search path" allows to find a path between two different states of the state transition graph. The result of "search path" is the first encountered shortest path. However, for the purposes of our

analysis we needed to extract all the paths reachable from initial states and visualize them in an automated manner. GINsim does not support this functionality therefore we implemented our algorithm to search paths by DFS. We present it later in the section 3.5.

3.4 NuSMV

NuSMV is a symbolic model checker [6]. It is used to verify that a given boolean model satisfies a given LTL formula. If the model does not satisfy the formula then the model checker gives one counterexample of a path which does not satisfy the formula.

Models defined in GINsim can be directly exported into a NuSMV file [21]. Before the export the initial states can be set which reduce the transition graph only to paths reachable from initial states. If the initial states are not set then a given LTL formula is interpreted over the full transition graph where all states are initial.

3.5 Implementation of DFS algorithm

For the purposes of our analysis we needed to extract all the paths in the state transition graph reachable from initial states and visualize them in an automated manner. We implemented algorithm to search paths by DFS (Depth First Search, see Algorithm 1). The algorithm takes as the input the state transition graph G and one arbitrary initial state $s_0 \in G$. The algorithm finds all paths s_0, \dots, s_t where there is no outgoing edge from s_t . And it also finds all pairs $p(\text{prefix}, \text{cycle}) \in G$ where:

- $\text{prefix} = s_0, \dots, s_n$ is a path connected to the cycle by the edge $s_n \rightarrow s'_0$.
- $\text{cycle} = s'_0, \dots, s'_n$ is a cycle such that $s'_n = s'_0$ and there is no state twice in the cycle.

Correctness of the Algorithm 1 can be proved by the fact that each path occurs on the stack S exactly once. The while cycle has at most $n!$ iterations because there is at most $n!$ paths in the graph with n vertices. Since there is at most $n!$ iterations of while cycle and single

iteration is linear in graph size, the complexity of the algorithm is $n!$. The algorithm is not scalable and it is intended for small graphs. To analyse large graphs, we use different approach based on model checking.

Algorithm 1 $DFS(G, initialNode)$

```

1:  $S \leftarrow$  empty stack
2:  $S.push((initialNode, successors(initialNode)))$ 
3: while  $S$  not empty do
4:   if  $second(S.top()) = empty$  then1
5:      $S.pop()$ ; continue
6:    $next \leftarrow$  any element from  $second(S.top())$ 
7:   remove  $next$  from  $second(S.top())$ 
8:
9:   if  $\exists n \in S$  such that  $first(n) = next$  then2
10:    report cycle
11:   else
12:      $S.push((next, successors(next)))$ 
13:     if  $successors(next) = empty$  then
14:       report path

```

Implementation details

The implementation uses the state transition graph exported from GINsim in ".layout" file where each line of the file defines two nodes connected to each other by directed edge. The program parses the ".layout" file then it processes the DFS algorithm and visualizes each path by the multiplot of the proteins' activity time courses.

The program is written in c++11. It uses Gnuplot-iostream interface which involves sending commands to Gnuplot. The Gnuplot 4.6. version is required. The source code was written in Visual Studio which is required for the compiling. The program can be run by the command ".\NPF.exe [fileName].layout" from the command

-
1. $second(s)$ returns second element of tuple s
 2. $first(s)$ returns first element of tuple s

line. The user is asked to choose one of the nodes of the state transition graph as the initial node. For each processed path the program creates a file "path.data" with the plot coordinates. For each path is generated a graph saved as "plot.png".

Figure 3.2 is example of program output. Program takes as input the state transition graph (Figure 3.2a) and initial node 1000 and returns all paths reachable from 1000. Only two of them are shown here.

3.6 Experimental data

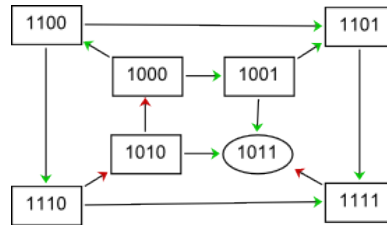
Western blot

Western blot is one of the most commonly used method in the studies of signalling pathways. The proteins extracted from the cell lysate are loaded on electrophoretic gel where they are separated by 3D structure or by length. Then they are transferred to a nitrocellulose or PVDF membrane where they keep the exact same position as they had on electrophoretic gel. Afterwards they are bound with stained antibodies specific to the target protein. The protein levels are evaluated by densitometry.

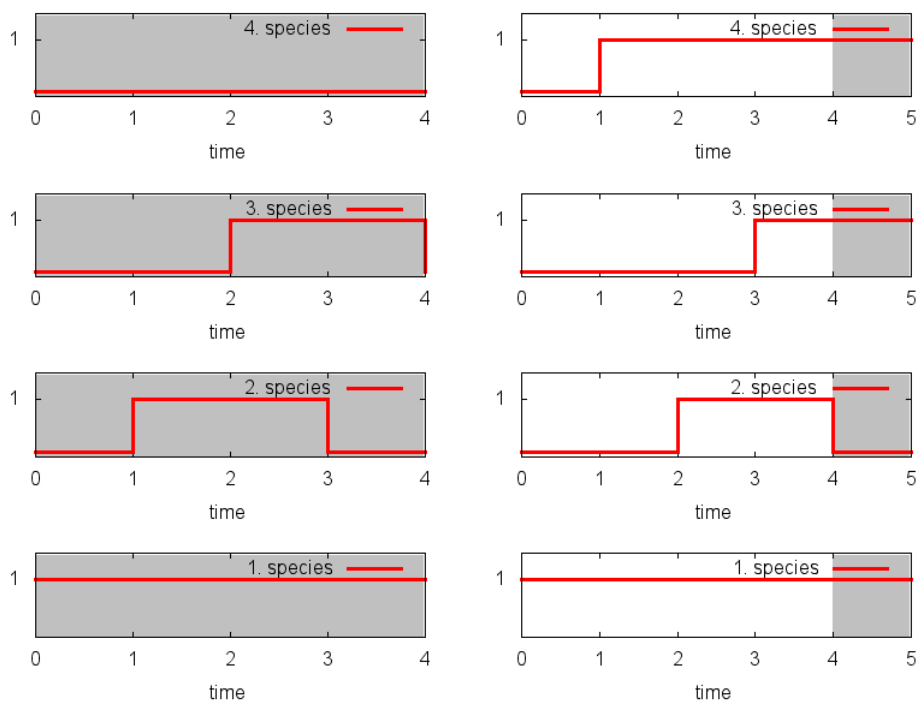
The final image consists of several bands, each band corresponds to single protein (Figure 3.4). Western blot is used as a qualitative indicator if protein activity is up or downregulated based on how intense the band is. The band intensity depends on several factors which are stochastic. Therefore, the intensity cannot be reproduced in two different measurements on different gels. Western blot generates qualitative data. It allows to find out general trends in data rather than absolute quantification of proteins.

Experimental data

We have been provided with western blot data by Pavel Krejčí's group from the Department of Experimental Biology MU. They measured the overtime behaviour of the FGF pathway as the level of phosphorylated downstream molecules, phosphorylated ERK (P-ERK) and phosphorylated FRS2 (P-FRS2).



(a) The state transition graph which the program uses as an input.



(b) Path (1000,1100,1110,1010)

(c) Path (1000,1001,1101,1111,1011)

Figure 3.2: Demonstrative run of the DFS implementation. The program takes as the input the state transition graph (a) and the initial node 1000 and returns all paths reachable from the initial node. Only two paths are shown here (b and c). The grey background denotes a cycle. The path (b) consists of a single cycle. The path (c) ends with terminal state 1011 which is shown as a single state cycle.

Western blot images (Figure 3.4) were evaluated by software to quantify the relative amount of proteins by both the area and the intensity of bands in terms of optical density. The Figure 3.3 depicts the summary of 12 independent measurements of P-ERK. The data are variable and does not allow absolute quantification of P-ERK. We can only extract general trend in data which corresponds to sustained P-ERK activity (section 2.2). The level of P-ERK rapidly increases in 2 hours and slowly decreases in the range of 12 hours. The same trend can be found in P-FRS2 measurements (data not shown).

Regarding the qualitative character of data it is eligible to abstract them as boolean values.

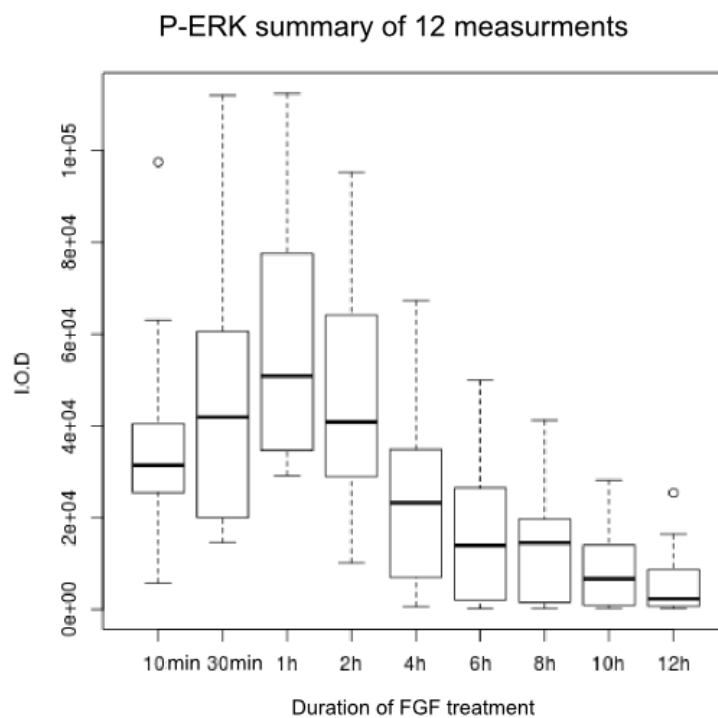


Figure 3.3: Boxplot of 12 independent measurements of ERK activity in hESC and RCS cells.

3. METHODS

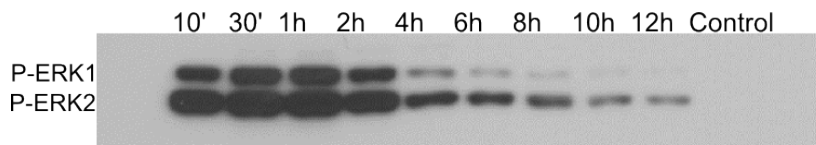


Figure 3.4: Example of western blot image. RCS cells were treated for indicated time with FGF. Control comes from untreated cells.

4 Results

Based on the biological knowledge we have built several models of the FGF signalling pathway. We have implemented the models in GINsim and we have analysed properties of their dynamics with our implementation of DFS and with the model checker NuSMV.

4.1 Analysed properties

We use the activity of ERK as a main indicator of the FGF pathway behaviour. We focus on two properties of the FGF pathway, transient and sustained ERK activity. Additionally, we analyse transient and sustained FRS2 activity. We formally define the properties by the following LTL formulae:

- $\mathcal{F}(ERK > 0) \wedge \mathcal{GF}(ERK < 1)$ expresses transient ERK activity.
- $\mathcal{F}(FRS2 > 0) \wedge \mathcal{GF}(FRS2 < 1)$ expresses transient FRS2 activity.
- $\mathcal{FG}(ERK > 0)$ expresses sustained ERK activity.
- $\mathcal{FG}(FRS2 > 0)$ expresses sustained FRS2 activity.
- $\mathcal{FG}(ERK > 0) \wedge \mathcal{FG}(FRS2 > 0)$ expresses sustained ERK and FRS2 activity.

We say that a given model \mathcal{B} allows transient/sustained behaviour if there is at least one run in $BTS(\mathcal{B})$ which satisfies the LTL formula describing transient/sustained ERK activity.

4.2 Models

We present the models in the form of schema of the regulatory graph $G(V, E)$ where positive interactions are marked by symbol \rightarrow and negative interactions by symbol \neg . The following two constraints hold for all models: $\forall g \in V : \rho(g) = 1$ and $\forall e \in E : \theta(e) = 1$. Therefore, all models are simply boolean and each node can gain binary value, either 0 or 1.

For each model we set the topology, the regulatory logic and the initial states for construction of the state transition graph. We explain the model settings based on the biology of the FGF pathway.

As initial, we choose all states where FGF is 1 because we want to mimic the dynamics of the system in FGF stimulated cells.

4.3 Model $\mathcal{M}1$

The Figure 4.1a depicts the topology of the model $\mathcal{M}1$. It is a basic model which consists of three proteins (FGF, FRS2, ERK).

The edge $\langle \text{FGF}, \text{FGF} \rangle$ simulates the constitutive stimulation of cells by FGF. The edge $\langle \text{FGF}, \text{FRS2} \rangle$ simulates the signal transduction through receptor and consequent FRS2 activation. The edge $\langle \text{FRS2}, \text{ERK} \rangle$ is an abstraction of signal transduction by the MAPK phosphorylation cascade. The edge $\langle \text{ERK}, \text{FRS2} \rangle$ mimics that ERK phosphorylates FRS2 and consequently inhibits the FRS2 activity[18].

We set the basal activity of the proteins to zero by the rules $K_{g_i}, \emptyset = 0$ because we want to fit the models to the control western blot measurements of unstimulated cells (Chapter 3.6) which show that the FRS2 and ERK is inactive when there is no FGF in the environment. Due to this settings, the activation of FRS2 results from the stimulation by FGF. Similarly, ERK is not activated spontaneously but it is activated by FRS2. We set it by the rules $K_{FRS2}, \{\text{FGF}\} = 1$, $K_{ERK}, \{\text{FRS2}\} = 1$.

The model is divided into two models $\mathcal{M}1.1$ and $\mathcal{M}1.2$ which differ in the settings of the rule $K_{FRS2}, \{\text{FGF}, \text{ERK}\}$. $\mathcal{M}1.1$ sets the rule to 0 which means that the ERK negative feedback is stronger than positive influence of FGF. $\mathcal{M}1.2$ sets the rule to 1 which means that the negative feedback is not strong enough to inactivate FRS2 if there is positive stimulation by FGF.

4.3.1 $\mathcal{M}1$ behaviour

Our analysis shows that the models $\mathcal{M}1.1$ and $\mathcal{M}1.2$ behave differentially as a result of the regulatory logic on the node FRS2 (Figure 4.1).

The model $\mathcal{M}1.2$ behaves as if there is not the feedback ERK \rightarrow FRS2 due to the steady FGF stimulation. Therefore, the ERK activity is sus-

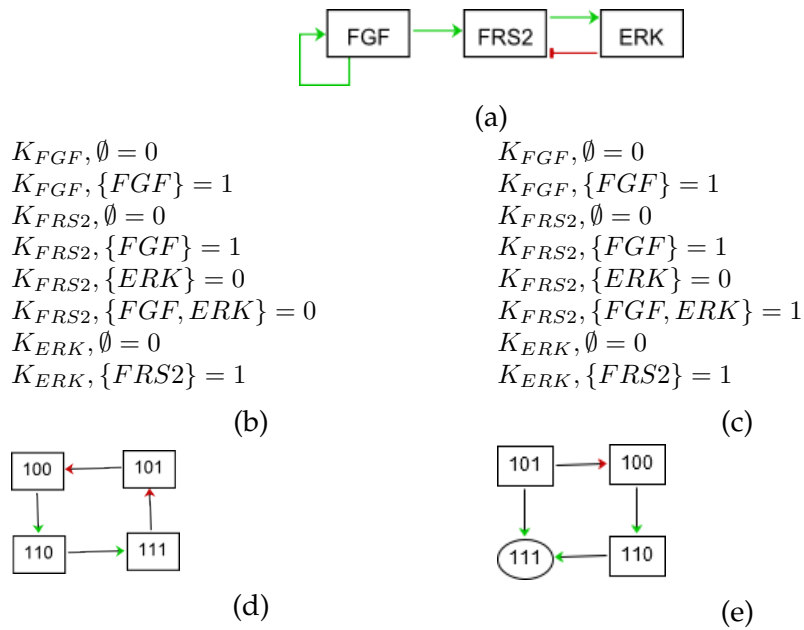


Figure 4.1: The regulatory graph of the model $\mathcal{M}1$ (a). The regulatory logic of the model (b) $\mathcal{M}1.1$ (c) $\mathcal{M}1.2$. The state transition graph of the model (d) $\mathcal{M}1.1$ (e) $\mathcal{M}1.2$. The components in a single state correspond to FGF, FRS2 and ERK, in this order.

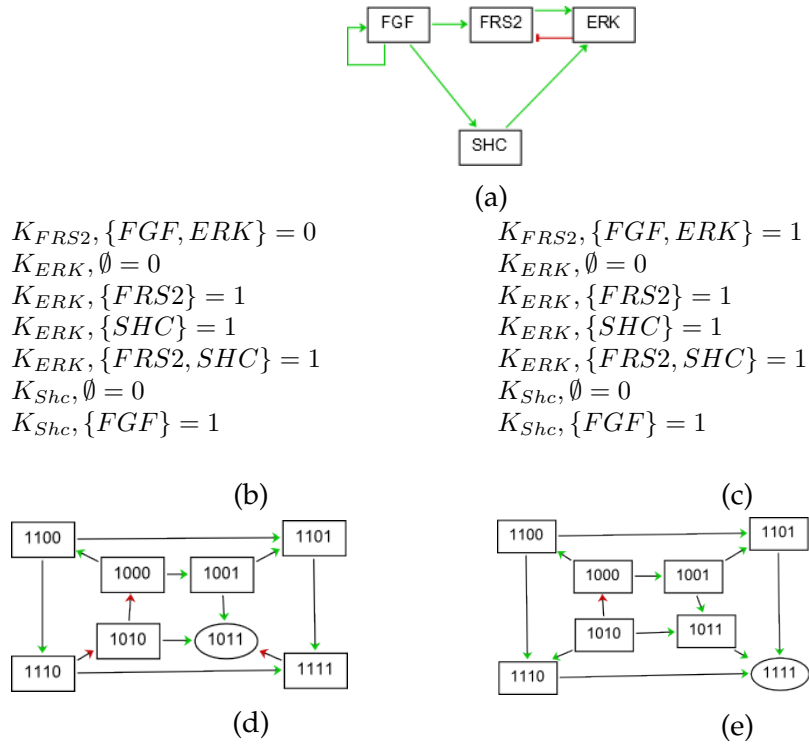


Figure 4.2: The regulatory graph of model $\mathcal{M}2$ (a). The regulatory logic of the model (b) $\mathcal{M}2.1$ (c) $\mathcal{M}2.2$. Rules which are not mentioned are the same as in $\mathcal{M}1$. The state transition graph of the model (d) $\mathcal{M}2.1$ (e) $\mathcal{M}2.2$. The components in a single state correspond to FGF, FRS2, ERK and SHC, in this order.

tained. On the contrary, the model $\mathcal{M}1.1$ allows the feedback $ERK \rightarrow FRS2$ which results in the transient ERK activity.

4.4 Model $\mathcal{M}2$

The model $\mathcal{M}2$ (Figure 4.2a) is an extension of $\mathcal{M}1$. There is one additional protein SHC which allows an alternative flow of a signal. The $\mathcal{M}2$ is again divided in two models $\mathcal{M}2.1$ and $\mathcal{M}2.2$ which differ in the rule $K_{FRS2, \{FGF, ERK\}} = 0/1$.

4.4.1 $\mathcal{M}2$ behaviour

Our analysis shows that the model $\mathcal{M}2.1$ allows both transient and sustained ERK activity (Figure 4.2). The transient activity is caused by the signal transduction via FRS2 which is in cycles switched off by the negative feedback ERK \neg FRS2. The sustained ERK activity results from the signal transduction via SHC.

Compared to $\mathcal{M}1.1$, the model $\mathcal{M}2.1$ gains new properties due to the SHC which allows an alternative flow of a signal.

Similarly as $\mathcal{M}1.1$, the model $\mathcal{M}2.2$, allows only sustained behaviour because the negative feedback on FRS2 cannot take place while there is positive stimulation by FGF.

4.5 Model $\mathcal{M}3$

The model $\mathcal{M}3$ is an extension of $\mathcal{M}2.2$ (Figure 4.3). There is one additional negative edge $\langle \text{ERK}, \text{SHC} \rangle$ which simulates the negative feedback ERK \neg Grb2-Sos [8].

4.5.1 $\mathcal{M}3$ behaviour

Our analysis shows that $\mathcal{M}3$ allows both the transient and sustained ERK activity. The transient/sustained ERK activity results from the signal transduction via SHC/FRS2, respectively.

In addition, the FRS2 activity is also sustained. Therefore, the model fits to our data (Chapter 3.6) which show that the ERK and FRS2 sustained activity occur at the same time.

4.6 Model $\mathcal{M}4$

The model $\mathcal{M}4$ extends the model $\mathcal{M}3$. It contains three new components Gab1, the complex Grb2-Sos (here as GS), and the complex Shp2-Grb2-Sos (Shp2-GS). Gab1 binds GS and Shp2-GS and SHC binds GS [17]. FRS2 creates a complex with both GS and Shp2-GS. However, the complex FRS2-GS does not affect the activation of ERK according to literature [18]. Therefore it is not model here.

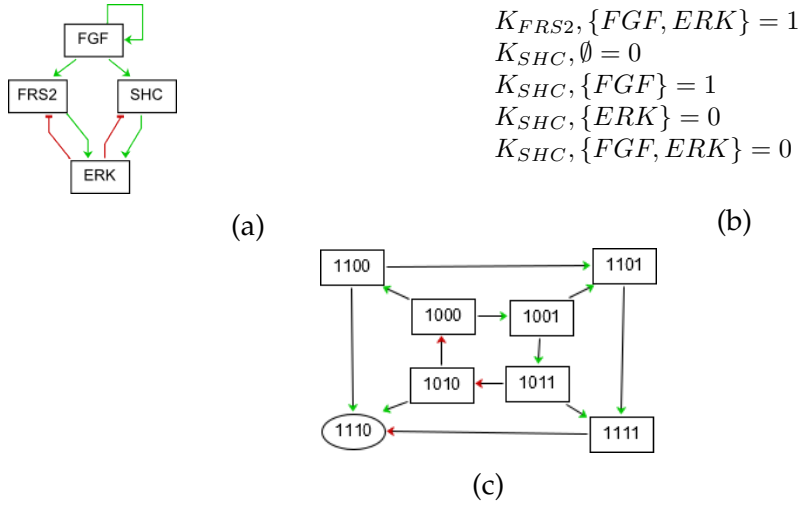


Figure 4.3: (a) The regulatory graph of the model \mathcal{M}_3 (b) The regulatory logic of \mathcal{M}_3 . Rules which are not mentioned are the same as in $\mathcal{M}_2.2$. (c) The state transition graph of the model \mathcal{M}_3 . The components in a single state correspond to FGF, FRS2, ERK and SHC, in this order.

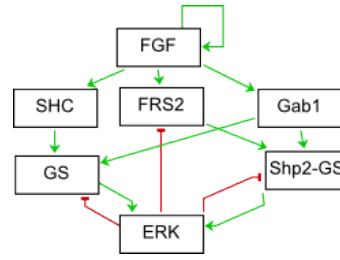
The state transition graph of the model \mathcal{M}_4 has 64 nodes therefore it is not shown here. We have analysed \mathcal{M}_4 properties using NuSMV.

4.7 Model \mathcal{M}_5

The model \mathcal{M}_5 is defined in the Figure 4.5. Compared to \mathcal{M}_4 , the model \mathcal{M}_5 contains four additional components FGFR, Ras, Raf and MEK which are known to be part of the FGF pathway. Therefore, the model \mathcal{M}_5 is less abstract and more realistic than \mathcal{M}_4 .

As initial, we choose all states where FGF and FGFR is 1 because we want to mimic the dynamics of a cell with persistently active receptors.

The additional components do not change the properties of \mathcal{M}_4 analysed in this theses. However, they might be important for future extension of the model and analysis of other properties of the FGF pathway.



(a)

$$\begin{aligned}
 K_{g_i, \emptyset} &= 0 \\
 K_{FGF, \{FGF\}} &= 1 \\
 K_{SHC, \{FGF\}} &= 1 \\
 K_{Gab1, \{FGF\}} &= 1 \\
 K_{FRS2, \{FGF\}} &= 1 \\
 K_{FRS2, \{FGF, ERK\}} &= 1 \\
 K_{FRS2, \{ERK\}} &= 0 \\
 K_{GS, \{SHC\}} &= 1 \\
 K_{GS, \{Gab1\}} &= 1 \\
 K_{GS, \{SHC, Gab1\}} &= 1 \\
 K_{GS, \{ERK, Gab1\}} &= 0 \\
 K_{GS, \{ERK, SHC\}} &= 0
 \end{aligned}$$

$$\begin{aligned}
 K_{GS, \{ERK, SHC, Gab1\}} &= 0 \\
 K_{GS, \{ERK\}} &= 0 \\
 K_{Shp2-GS, \{FRS2\}} &= 1 \\
 K_{Shp2-GS, \{Gab1\}} &= 1 \\
 K_{Shp2-GS, \{FRS2, Gab1\}} &= 1 \\
 K_{Shp2-GS, \{ERK, FRS2\}} &= 1 \\
 K_{Shp2-GS, \{ERK, Gab1\}} &= 1 \\
 K_{Shp2-GS, \{ERK, FRS2, Gab1\}} &= 1 \\
 K_{Shp2-GS, \{ERK\}} &= 0 \\
 K_{ERK, \{GS\}} &= 1 \\
 K_{ERK, \{Shp2-GS\}} &= 1 \\
 K_{ERK, \{GS, Shp2-GS\}} &= 1
 \end{aligned}$$

(b)

Figure 4.4: (a) The regulatory graph of the model $\mathcal{M}4$ (b) The regulatory logic of $\mathcal{M}4$.

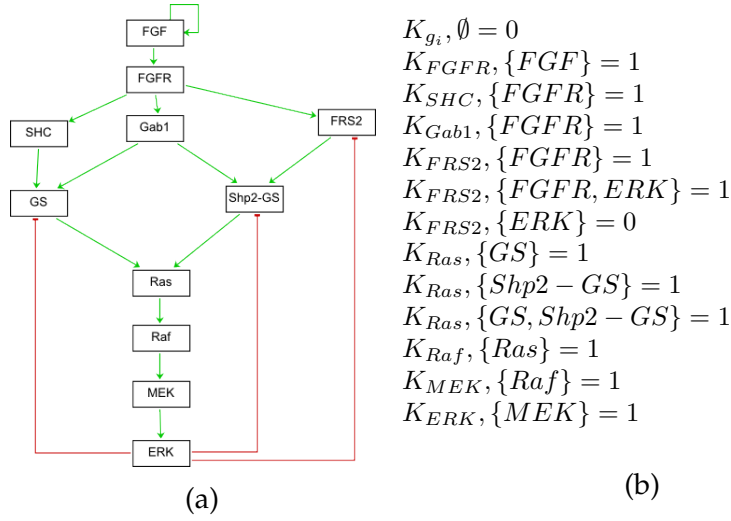


Figure 4.5: (a) The regulatory graph of the model $\mathcal{M}5$ (b) The regulatory logic of $\mathcal{M}5$. Rules which are not mentioned are the same as in model $\mathcal{M}4$.

4.8 Summary of analysis

The Table 4.1 summarises analysed properties of the models.

Model $\mathcal{M}1$ is a basic regulatory unit which consists of three nodes. It allows either sustained or transient behaviour based on the settings of logic rules on the node FRS2 which has incoming positive and negative interaction. $\mathcal{M}1$ cannot allow both types of behaviour.

In the model $\mathcal{M}2$ we have implemented additional node SHC which provides an alternative signal flow. As a result, the model $\mathcal{M}2.1$ allows both types of behaviour, however, there is not a run in $\mathcal{M}2.1$ such that it satisfies the property of sustained ERK and FRS2. Therefore, the model $\mathcal{M}2.1$ does not fit to our data that sustained ERK activity occurs together with the sustained FRS2 activity.

The model $\mathcal{M}3$ is the simplest complex model which allows transient and sustained ERK activity and which contains a run such that it satisfies the property of sustained ERK and FRS2. The model $\mathcal{M}4$ and $\mathcal{M}5$ extend $\mathcal{M}3$ based on the literature. They have the same properties as $\mathcal{M}3$. However, the models should be validated with respect to other properties beyond our analysis.

Model	ERK activity	FRS2 activity	sustained ERK and FRS2
$\mathcal{M}1.1$	transient	transient	no
$\mathcal{M}1.2$	sustained	sustained	yes
$\mathcal{M}2.1$	transient and sustained	transient	no
$\mathcal{M}2.2$	sustained	sustained	yes
$\mathcal{M}3$	transient and sustained	sustained	yes
$\mathcal{M}4$	transient and sustained	sustained	yes
$\mathcal{M}5$	transient and sustained	sustained	yes

Table 4.1: Properties of the models.

5 Discussion

Using GINsim we have built seven boolean models of the FGF pathway based on literature and unpublished western blot data. The models should be validated by additional *in vitro* experiments. We have expressed the properties of transient and sustained behaviour of the pathway by LTL formulae and analysed the models with an implementation of DFS and with the NuSMV model checker. The DFS program was useful to clearly visualize the model behaviour. However, it is not scalable and we used it for analysis of less complex models. The models \mathcal{M}_4 and \mathcal{M}_5 have been analysed only by NuSMV.

Even though the boolean modelling is highly abstract formalism it is suitable for analysis of the FGF pathway because the exact interactions in the pathway are unclear. To model this system with more expressive formalism, e.g. ODE, would require using enormous number of parameters based only on estimation. Using the boolean models we avoid this problem and clarify the basic questions about the FGF pathway behaviour without the knowledge of biochemical dynamics underlying the pathway. The models we have built can very well serve as a scaffold for creating more expressive models in ODE.

Our analysis implies that the sustained behaviour is established by reaching the terminal state without any outgoing edge and the transient behaviour is established by reaching a cycle of states where the pathway is periodically switched on and off. The increasing complexity of our models allowed us to identify the regulatory rules which model the negative feedback as a key element affecting model properties. By a change of the regulatory rules we can evoke a change between transient and sustained behaviour.

Based on our analysis, we suggest two hypotheses about the mechanism underlying the switch between transient and sustained behaviour of the FGF signalling pathway.

- Single cell is capable of both, transient and sustained behaviour, in response to FGF. The sustained behaviour is maintained by signalling via FRS2 and the transient behaviour is maintained by signalling via other adaptor, e.g. SHC or an unknown adaptor. We can model this case as a single model with two adaptor proteins which differ in the regulatory rules, as in the case of

the model $\mathcal{M}3$

- Single cell has only one behaviour profile in response to FGF, transient or sustained. The profile depends on the cell type. Specialized cells differ in proteins which they produce. Differences in protein concentration overstepping a threshold result in the change of regulation of signal transduction. We can model this case as two models with the same topology and different settings of the regulatory rules, as in the case of the model $\mathcal{M}1.1$ and $\mathcal{M}1.2$.

6 Conclusion

In this thesis, we study the behaviour of the FGF signalling pathway using boolean networks. In the Chapter 2 we have summarised a brief overview of the FGF pathway. It documents the biological background based on which we have implemented seven models in GINsim, a software for modelling and simulation of boolean networks.

The activity of the FGF pathway can have either transient or sustained profile. The molecular mechanism responsible for differences in the pathway activity is unknown. We have expressed this two qualities formally in LTL. Using the NuSMV model checker we have analysed our models for given qualities. To comprehensively visualize the model dynamics we have implemented DFS algorithm which takes as an input a model simulation exported from GINsim and an initial node and returns all paths reachable from the initial node depicted by gnuplot.

Analysis of our models implies that the sustained profile is established by reaching the stable state, the terminal state without any outgoing edge. The transient profile is caused by reaching a cyclic attractor, a cycle of states where the pathway is periodically switched on and off. The switch between transient and sustained profile can be induced by a change of the regulatory rules which model the negative feedback.

Based on our results, we suggest two hypotheses about the mechanism responsible for differences in the pathway activity. Firstly, for a single cell to show both profiles there has to be two or more signalling proteins each of them maintaining separate signal flow which is differentially affected by the negative feedback. Secondly, if a specialized cell type has only one profile, either transient or sustained, then all of the separate signal flows are affected to the same level by the negative feedback.

The models we have built can serve as scaffold for more complicated models for future analysis. The models are suitable for creating hypotheses about the mechanisms of signalling pathway in genetic diseases, as we are currently doing for the signal transduction in achondroplasia disease, most common form of human dwarfism.

Bibliography

- [1] I. Ahmad, T. Iwata, and H. Y. Leung. Mechanisms of FGFR-mediated carcinogenesis. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1823(4):850–860, 2012.
- [2] T. Araki, H. Nawa, and B. G. Neel. Tyrosyl phosphorylation of Shp2 is required for normal ERK activation in response to some but not all, growth factors. *Journal of Biological Chemistry*, 278(43):41677–84, 2003.
- [3] J. Barnat, L. Brim, S. Cerna, S. Drazan, J. Fabrikova, J. Lanik, and D. Safranek. BioDiVinE: A Framework for Parallel Analysis of Biological Models. *Electronic Proceedings in Theoretical Computer Science*, 15, 2009.
- [4] J. Barnat, L. Brim, A. Krejci, A. Streck, D. Safranek, M. Vejnar, and T. Vejpustek. On parameter synthesis by parallel model checking. *IEEE/ACM transactions on computational biology and bioinformatics*, 9(3):693–705, 2012.
- [5] T. Bondeva, A. Balla, P. Várnai, and T. Balla. Structural determinants of ras-raf interaction analyzed in live cells. *Molecular Biology of the Cell*, pages 2323–2333.
- [6] A. Cimatti, E. M. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Rovere, R. Sebastiani, and A. Tacchella. Nusmv 2: An opensource tool for symbolic model checking. *Proceeding CAV '02 Proceedings of the 14th International Conference on Computer Aided Verification*, 2002.
- [7] E. M. Clarke, O. Grumberg, and D. Peled. *Model Checking*. 1999.
- [8] T. Corbalan-Garcia, S. Yang, K. R. Degenhardt, and D. Bar-Sagi. Identification of the Mitogen-Activated Protein Kinase Phosphorylation Sites on Human Sos1 That Regulate Interaction with Grb2. *Molecular and cellular biology*, 16(10):5674–5682, 1996.

-
- [9] L. Erzsébet, S. Welte, and K. Scheffzek. Inhibition and Termination of Physiological Responses by GTPase Activating Proteins. *Physiological Reviews*, 92(1):237–272, 2012.
- [10] V. P. Eswarakumar, I. Lax, and J. Schlessinger. Cellular signaling by fibroblast growth factor receptors. *Cytokine & Growth Factor Reviews*, 16(2):139–149, 2005.
- [11] S. Foldynova-Trantirkova, W. R. Wilcox, and P. Krejci. Sixteen years and counting: the current understanding of fibroblast growth factor receptor 3 (FGFR3) signaling in skeletal dysplasias. *Human Mutation*, 33:29–41, 2012.
- [12] L. Grieco, L. Calzone, I. Bernard-Pierrot, F. Radvanyi, B. Kahn-Perles, and D. Thieffry. Boolean modeling of biological regulatory networks: A methodology tutorial. *Methods*, 62(1), 2013.
- [13] L. Grieco, L. Calzone, I. Bernard-Pierrot, F. Radvanyi, B. Kahn-Perles, and D. Thieffry. Integrative modelling of the influence of mapk network on cancer cell fate decision. *PLOS Computational Biology*, 9(10), 2013.
- [14] O. A. Ibrahim, B. K. Yeh, A. V. Eliseenkova, S. K. Zhang, F. ad Olsen, M. Igarashi, Aarson S. A., R. J. Linhardt, and M. Mohammadi. Analysis of mutations in fibroblast growth factor (FGF) and a pathogenic mutation in FGF receptor (FGFR) provides direct evidence for the symmetric two-end model for FGFR dimerization. *Mol Cell Biol*, 25:671–684.
- [15] H. Klarner, A. Streck, D. Safranek, J. Kolcak, and H. Sibert. Parameter Identification and Model Ranking of Thomas Networks. *Computational Methods in Systems Biology*, 2012.
- [16] P. Krejci, V. Bryja, J. Pachernik, A. Hampl, R. Pogue, P. Mikikian, and R. Wilcox. FGF2 inhibits proliferation and alters the cartilage-like phenotype of RCS cells. *Experimetal Cell Research*, 297:152–164, 2004.
- [17] P. Krejci, B. Masri, L. Salazar, C. Farrington-Rock, H. Prats, L. M. Thompson, and W. R. Wilcox. Bisindolylmaleimide I Suppresses

- Fibroblast Growth Factor-mediated Activation of Erk MAP Kinase in chondrocytes by Preventing Shp2 Association with the Frs2 and Gab1 Adaptor Proteins. *Journal of Biological Chemistry*, 282(5):2929–36, 2007.
- [18] I. Lax, A. Wong, B. Lamothe, A. Lee, A. Frost, J. Hawes, and J. Schlessinger. The Docking Protein FRS2 α Controls a MAP Kinase-Mediated Negative Feedback Mechanism for Signaling by FGF Receptors. *Molecular Cell*, 10:709–719, 2002.
- [19] B. J. Mayer and D. Baltimore. Signaling through sh2 and sh3 domains. *Trends in Cell Biology*, 3:8–13.
- [20] M. Mohammadi, S. K. Olsen, and O. A. Ibrahimi. Structural basis for fibroblast growth factor receptor activation. *Cytokine & Growth Factor Reviews*, 16(2):107–37, 2005.
- [21] P. Monteiro and Chaouiya. C. Efficient verification for logical models of regulatory networks. *Proc. 6th Intl. Conf. on Practical Applications on Computational Biology & Bioinformatics PACBB'12 (Salamanca, Spain)*, 154:259–267, 2012.
- [22] A. Naldi, D. Berenguier, A. Faure, F. Lopez, D. Thieffry, and C. Chaouiya. Logical modelling of regulator networks with GINsim 2.3. *Biosystems*, 97(2):134–9, 2009.
- [23] A. D. Sharrocks. Cell Cycle: Sustained ERK Signalling Represses the Inhibitors. *Current Biology*, 16:540–542, 2006.
- [24] D. Thieffry and R. Thomas. Dynamical Behavior of Biological Regulatory Networks. *Bulletin of Mathematical Biology*, 57(2):277–97, 1995.
- [25] K. Thobe, A. Streck, H. Klarner, and H. Siebert. Model Integration and Crosstalk Analysis of Logical Regulatory Networks. *Computational Methods in Systems Biology. Lecture Notes in Computer Science*, 8859:32–44, 2014.
- [26] R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks—I. Biological role of feedback

loops and practical use of the concept of the loop-characteristic state. *Bulletin of Mathematical Biology*, 57(2), 1995.

- [27] S. Yamada, T. Taketomi, and A. Yoshimura. Model analysis of difference between EGF pathway and FGF pathway. *Biochemical and Biophysical Research Communications*, 314:1113–1120, 2004.

A Attached files

The following directories are attached to this thesis.

- `NPF` contains an implementation of DFS algorithm.
- `Ginsim` contains models implemented in GINsim
- `layout` contains the state transition graphs exported from GINsim which can be used as input for `NPF.exe`.
- `Nusmv` contains models in the smv format.

B Signalling pathways

The following text summarises basic principles of signalling pathways. It is intended for educational purposes. It is written in Czech. The chapters are numbered separately from the rest of the thesis.

1 Signální dráhy

Tento výukový text poskytuje obecný přehled principů signálních drah.

1.1 Buněčná komunikace

Buněčná komunikace u mnohobuněčných živočichů udržuje stabilní vnitřní prostředí organismu a řídí jeho růst a vývoj. Buňky vysílají chemický signál do svého okolí, kde jej detekují ostatní buňky pomocí receptorů, specializovaných proteinů, které mají vazebná místa specificky uzpůsobená pro konkrétní signál. Signály jsou velmi různorodé. Patří sem např. proteiny, malé peptidy, aminokyseliny, nukleotidy, steroidy, retinoidy, deriváty mastných kyselin nebo i plyny oxid dusnatý a uhelnatý (Alberts *et al.*, 2008).

1.2 Receptory a příjem signálu

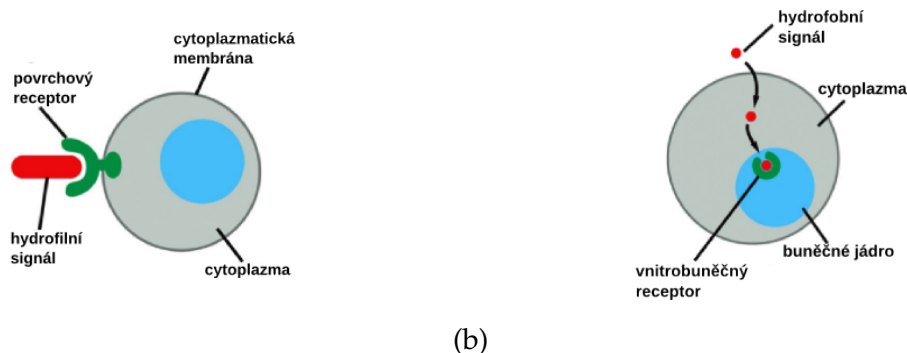
Rozhraní v komunikaci mezi buňkou a okolím tvoří cytoplazmatická membrána. Ta je tvořena hlavně z fosfolipidů, což ji činí nepropustnou pro velké nebo hydrofilní molekuly (obrázek 1a) (Alberts *et al.*, 2008). Většina signálů tedy nemůže projít cytoplazmatickou membránou a je přijímána povrchovými receptory. Povrchový receptor je tvořen z vnější části, která váže signální molekuly, transmembránové části, která receptor ukotvuje v membráně, a vnitřní části, která přenáší signál v cytoplazmě. Vazba signálu mění prostorovou konformaci receptoru, čímž se zpráva přenáší do buňky.

Naopak malé hydrofobní molekuly vstupují do cytoplazmy nebo až do jádra buňky, kde se váží na vnitrobuněčné receptory a aktivují je (obrázek 1b). Aktivní receptory následně podněcují změny v transkripci genů. Vnitrobuněčné receptory mají např. steroidní a thyroïdní hormony.

Existují tři hlavní typy povrchových receptorů, a to receptory spojené s iontovými kanály, receptory vázané na G-protein a receptory vázané na enzym. Receptory spojené s iontovými kanály po vazbě signálu umožňují průchod iontů přes plazmatickou membránu (obrázek 3a). Výsledkem je rychlá změna rozložení celkového náboje okolo membrány. Tyto receptory přenášejí informaci na nervových synapsích.

Receptory vázané na G-protein jsou nejpočetnější rodina receptorů. Uvnitř buňky receptor interaguje s G-proteinem, který je v neaktivním stavu tvořen ze tří podjednotek a váže GDP. V aktivním stavu G-protein váže GTP a rozpadá se na dvě části, které aktivují další složky signální dráhy. Signalizace přes G-protein probíhá cyklicky, protože po aktivaci se G-protein sám deaktivuje (obrázek 2).

Receptory spojené s enzymy obsahují část orientovanou do cytoplazmy, která má enzymovou aktivitu. Většina těchto receptorů jsou tyrozinkinázové receptory (RTK),



Obrázek 1: Převzato z (Alberts *et al.*, 2008). (a) Hydrofilní molekuly nejsou schopné procházet cytoplazmatickou membránou do buňky. Proto je detekují povrchové receptory ukotvené v cytoplazmatické membráně. Po vazbě na signál mění receptor konformaci, čímž se signál přenáší do cytoplazmy. (b) Naopak hydrofobní molekuly cytoplazmatickou membránou procházejí a váží se na vnitrobuněčné receptory, které tím aktivují. Aktivované vnitrobuněčné receptory regulují transkripci genů v jádře.

jejichž enzymovou aktivitou je schopnost katalyzovat připojení fosfátové skupiny na další proteiny v signální dráze (viz sekce 1.3.1) (Robinson *et al.*, 2000). RTK jsou receptory některých růstových faktorů např. EGF a FGF (*Epidermal/Fibroblast Growth Factor*). Je pro ně typické, že při vazbě signálu vytvářejí dimery (obrázek 3b) (Yarden a Ullrich, 1988).

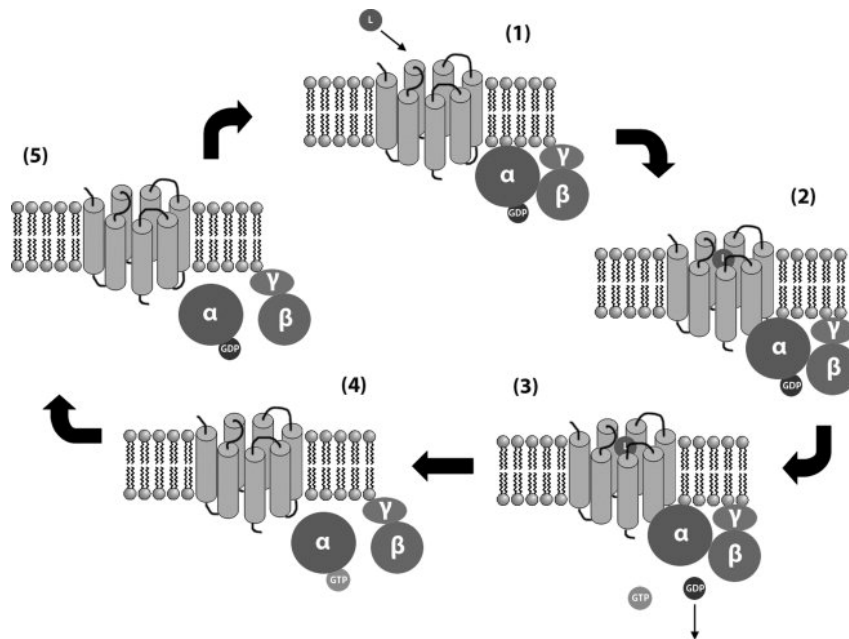
1.3 Přenos a zpracování signálu

V cytoplazmě přenášejí a zpracovávají signál hlavně signální proteiny. Tyto proteiny signál zesilují, integrují se signály z různých drah, distribuují do více paralelních toků a přenášejí ho až k efektorovému proteinu, který podnítl specifickou buněčnou odpověď (obrázek 4a). Efektory jsou transkripční faktory řídící přepis genů do mRNA, proteiny metabolismu nebo cytoskeletu (Alberts *et al.*, 2008).

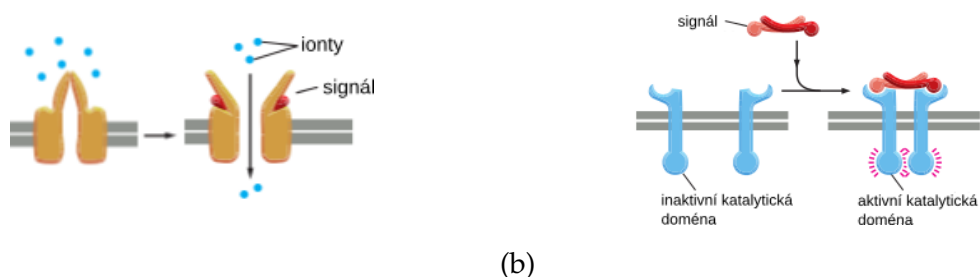
Informace šířící se po signálních drahách je navíc regulovaná pomocí negativních a pozitivních zpětných vazeb. Negativní zpětné vazby zajišťují vypnutí aktivní dráhy. Bez utlumení dráhy by buňka ztratila citlivost a schopnost reagovat na nově příchozí signál. Neschopnost utlumit některé signální dráhy může vést k nekontrolovanému dělení buněk a vzniku rakoviny.

Informace se šíří pomocí navození změn v signálních proteinech, nejčastěji v jejich konformaci. Díky těmto změnám se aktivuje konkrétní funkce proteinu. Např. se aktivuje schopnost fosforylovat jiné proteiny. Rovněž se tyto změny mohou promítnout ve spektru proteinů, které je signální protein schopen vázat a vytvářet s nimi komplex.

V následujících podsekcích jsou rozepsány nejčastější mechanismy, které se využívají k navození změn v proteinech signálních drah. Jsou to fosforylace/defosforylace,



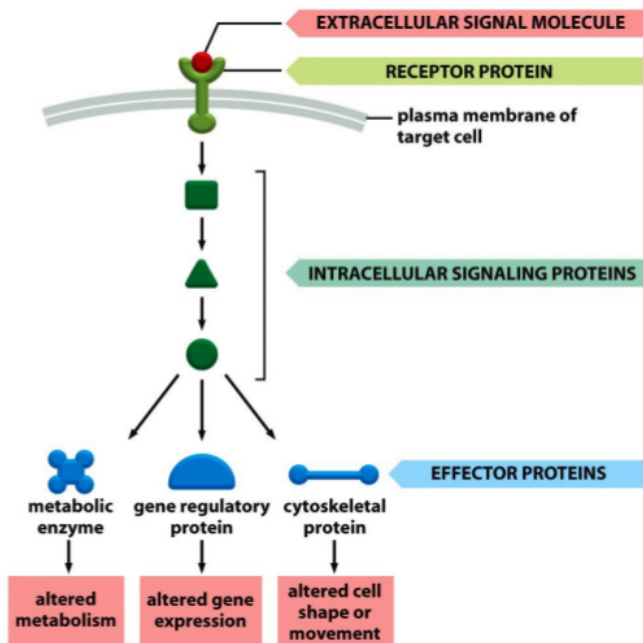
Obrázek 2: Cyklus signalizace přes G-protein. Převzato z (Khazifov a Lattazi, 2009). Receptor interaguje s G-proteinem, složeným z podjednotek α , β a γ . V neaktivním stavu jsou všechny tři podjednotky v komplexu a podjednotka α váže molekulu GDP. Změna konformace receptoru po detekování signálu vyvolá změnu konformace v G-proteinu. G-protein uvolní GDP, naváže GTP a rozpadne se na dvě části, podjednotku α -GTP a β - γ . Obě části difundují od receptoru a aktivují další složky v signální dráze. GTP samovolně hydrolyzuje na GDP a G-protein se opět deaktivuje spojením všech tří podjednotek.



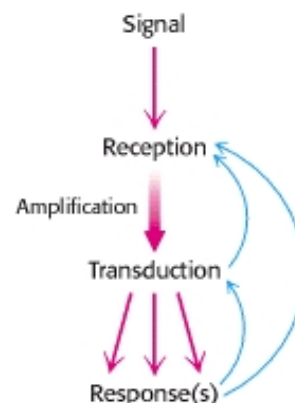
(a)

(b)

Obrázek 3: Převzato z (Alberts *et al.*, 2008). (a) Receptory spojené s iontovými kanály. (b) Receptory spojené s enzymovou (kinázovou) aktivitou.



(a) Převzato z (Alberts *et al.*, 2008).



(b) Převzato z (Berg *et al.*, 2002).

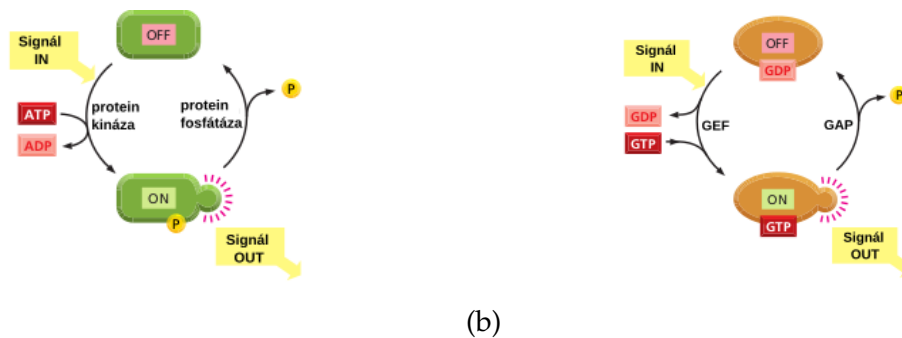
Obrázek 4: Princip signální dráhy. (a) Vnější signál je přes povrchový receptor přenesen do cytoplazmy, kde se pomocí signálních proteinů propaguje a distribuuje k efektorovým proteinům. Výstupem signální dráhy je specifická buněčná odpověď např. změna metabolismu, genové exprese nebo tvaru buňky. (b) Signál z prostředí např. růstový faktor či hormon je detekován nejčastěji povrchovým receptorem. Uvnitř buňky je zesílen a přeměněn na jiné chemické formy. Dráhu mohou regulovat pozitivní i negativní zpětné reakce.

vazba GTP/GDP, vazba jiných signálních proteinů, vazba na sekundární přenašeče signálu (cAMP, Ca^{2+} , diacylglycerol) a ubiquitinace.

1.3.1 Fosforylace a defosforylace proteinů

Běžným způsobem přenosu informace je fosforylace a defosforylace proteinů signální dráhy. Fosforylaci zprostředkovávají enzymy kinázy, které přenášejí fosfátovou skupinu z ATP na aminokyselinu v proteinu. Nejčastěji fosforylované aminokyseliny jsou serin, threonin a tyrosin (Hanks a Hunter, 1995). Kinázy lze rozdělit do dvou skupin podle aminokyseliny, kterou modifikují na serin/threonin kinázy a tyrosin kinázy. Antagonisté kináz jsou fosfatázy, které fosfátovou skupinu z proteinu odstraňují.

Fosforylace a defosforylace probíhají v řádu milisekund a umožňují tak rychlé aktivování a utlumení signální dráhy. Fosforylace nejčastěji aktivuje funkce proteinu, jak je to uvedeno na obrázku 5a, např. mění neaktivní kinázy na aktivní. Nicméně může na funkci proteinu působit i inhibičně. Fosforylace se často vyskytuje i na několika různých



Obrázek 5: Převzato z (Alberts *et al.*, 2008). Signální proteiny fungují jako molekulární přepínače mezi aktivním a neaktivním stavem (ON/OFF) pomocí (a) připojení/odstranění fosfátové skupiny (b) nebo pomocí vazby a hydrolýzy GTP. Hydrolýza GTP a navázání GTP na G protein probíhá samovolně, ale může být urychleno pomocí proteinů GEF a GAP.

místech daného proteinu, přičemž jedna fosforylace může být aktivační a jiná inhibiční. Běžně se v signálních dráhách vyskytují celé kaskády fosforylačních reakcí, kdy je první kináza aktivovaná fosforylací a následně fosforyluje druhou kinázu v kaskádě, která fosforyluje třetí kinázu atd. Velmi známá je např. MAPK fosforylační kaskáda (Kyriakis, 2014).

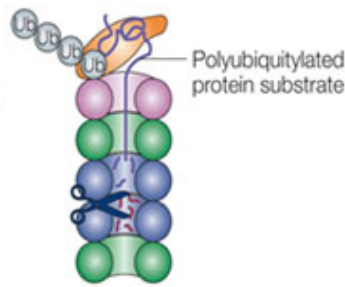
1.3.2 G proteiny v signálních dráhách

G proteiny přepínají mezi aktivní a neaktivní konformací pomocí vazby GTP/GDP (obrázek 5b). Proteiny GEF (*guanine nucleotide-exchange factor*) podporují uvolnění GDP z proteinu (Erzsébet *et al.*, 2012). K vazbě GTP dochází samovolně, protože je v cytoplazmě v nadbytku. Rychlost hydrolýzy GTP na GDP ovlivňují proteiny GAP (*GTPase-accelerating protein*).

V sekci 1.2 jsme si představili velké G proteiny složené ze tří podjednotek, které interagují s buněčnými receptory a disociují na dvě části. Kromě toho existuje i druhá skupina G proteinů. Jsou to malé proteiny tvořené jen jednou podjednotkou, která je podobná podjednotce α velkých G proteinů. Do této skupiny G proteinů patří proteinová rodina Ras, která aktivuje MAPK signální kaskádu.

1.3.3 Sekundární přenašeče a amplifikace signálu

Koncentrace receptorů v membráně je nedostatečná ke spuštění buněčné odpovědi. Proto musí být signál zesílen. K tomu signální dráhy často využívají malých nitrobuňčných mediátorů tzv. sekundárních posílů. Jsou vytvořeny ve velkém množství po aktivaci receptoru, difundují od místa svého vzniku a šíří signál dál do cytoplazmy. Mezi sekundární posly patří cyklický AMP, Ca^{2+} . Sekundární posel diacylglycerol je



Obrázek 6: Převzato z (Ciechanover, 2005). Proteazom je válcovitý molekulární komplex, který rozeznává polyubikvitinované proteiny a štěpí je na krátké peptidy.

rozpustný v tucích a difunduje v membráně. Sekundární poslové aktivují některé enzymy, např. proteinové kinázy. K zesílení signálu napomáhají i fosforylační kaskády. Důsledkem amplifikace signálu je, že už i nízká koncentrace ligandu spustí významnou odpověď uvnitř buňky, čímž se zvyšuje schopnost buňky reagovat na malé změny prostředí.

1.3.4 Ubikvitinace

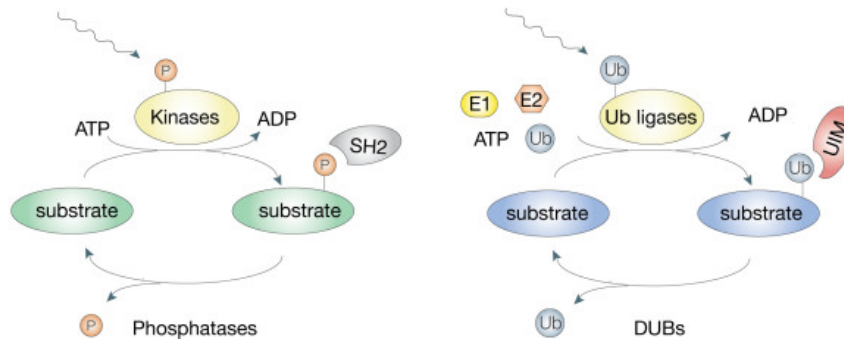
Ubikvitinace je stejně jako fosforylace dynamická, reverzibilní modifikace proteinu (Chen a Sun, 2009). Během této modifikace je na cílový protein připojen ubikvitin, krátký protein obsahující asi 70 aminokyselin. Ubikvitin obsahuje 7 aminokyselin, na které se mohou připojit další molekuly ubikvitinu tak, že vytvářejí různě propojené řetězce. Různé typy polyubikvitinace mají různý význam, protože jsou specificky rozeznávány jinými proteiny. Nejznámější je polyubikvitinace lysinu 48, která označuje protein k degradaci v proteazomu (obrázek 6).

Fosforylace stejně jako ubikvitinace umožňují vzájemné interakce proteinů. Modifikace proteinu slouží jako značka, kterou specificky rozeznávají a váží jiné proteiny, čímž dojde k vytvoření proteinových komplexů, které dále šíří stimul po signální dráze. Fosforylaci tyrosinu typicky rozeznává doména SH2 a ubikvitinaci rozeznává doména UIM (obrázek 7). Proteiny často obsahují více domén rozeznávajících různé značky, čímž se zvyšuje kombinatorický prostor pro vznik komplexů.

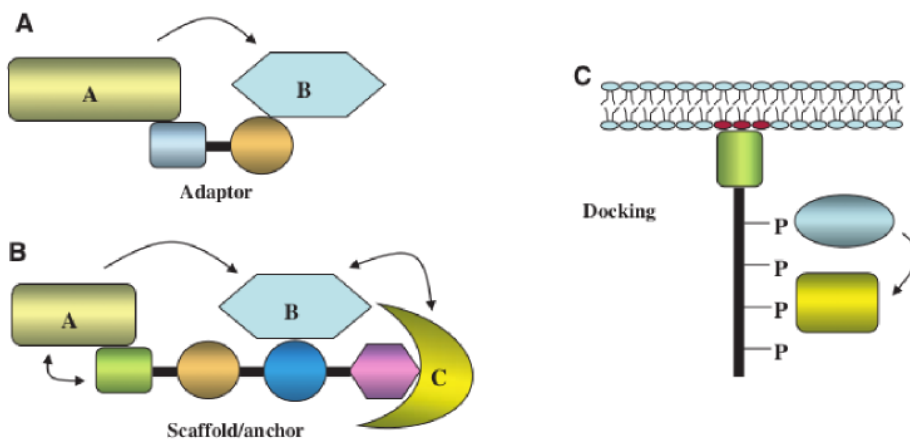
1.3.5 Proteinové komplexy v signálních dráhách

Interakce proteinů signálních drah často umožňují prostředníci, kteří udržují signální proteiny ve vzájemné blízkosti a ve správné orientaci (obrázek 8). Výsledkem je přesné a efektivní řízení směru toku informace v síti signálních drah.

Pro tyto proteiny existuje několik termínů, které se běžně zaměňují a používají synonymně. Jsou to proteinové adaptory, lešení, kotvy a doky (ang. *scaffold*, *adaptor*, *anchor*, *docking proteins*) (Buday a Tompla, 2010). Jako adaptory se obvykle označují malé pro-



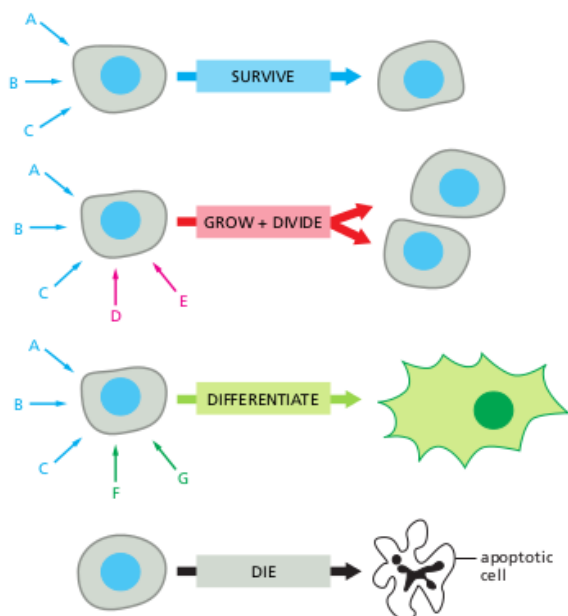
Obrázek 7: Srovnání fosforylace a ubikvitinace. Převzato z (Woelk *et al.*, 2007). Ubikvitinaci katalyzují ubikvitin ligázy E1, E2 a E3. Odstranění ubikvitinu katalyzují enzymy označované jako DUB. Fosfátovou skupinu rozeznávají interagující proteiny pomocí své SH2 domény, zatímco ubikvitin rozeznávají proteiny obsahující doménu UIM. Obě modifikace tak slouží jako lepidlo, které spojuje proteinové komplexy.



Obrázek 8: Prostředníci ve tvorbě proteinových komplexů jsou *adaptor*, *scaffold/anchor*, *docking* proteiny. Šipky označují vzájemné interakce a modifikace. Převzato z (Buday a Tompla, 2010)

teiny, které spojují dohromady dva interakční partnery např. proteinkinázu a její substrát. Proteinové lešení (či proteinová kotva) je větší než adaptor a je schopné vázat více než dva proteiny signální dráhy. Proteinové doky rovněž interagují s více než dvěma partnery, ale navíc jsou ukotvené v cytoplazmatické membráně.

Významnou skupinou adaptorů jsou adaptory složené z SH2/SH3 domén. Jsou to např. proteiny Grb2, Shc a Crk (Buday, 1999). Protein Grb2 obsahuje jednu doménu SH2, která rozeznává fosforylovaný tyrosin na aktivovaných proteinkinázových receptorech růstových faktorů (př. EGFR, FGFR) či na proteinových docích (Gab1, FRS2). Grb2 dále obsahuje dvě domény SH3, které interagují s motivem bohatým na prolin v proteinu Sos. Sos je protein GEF ativující Ras.



Obrázek 9: Příklad chování buňky v závislosti na kombinaci vstupních signálů. Převzato z (Alberts *et al.*, 2008). Buňka může odpovědět růstem a dělením nebo diferenciací v jiný typ buňky. Nepřichází-li k buňce žádné signály k přežití dojde k programované buněčné smrti (apoptóze).

1.4 Buněčná odpověď

Typicky jsou buňky mnohobuněčného organismu v daný moment vystaveny stovkám rozdílných signálů z prostředí, které na ně mohou působit stimulačně či inhibičně. Pro danou kombinaci a koncentraci vstupních signálů působících na konkrétní buňku je výstupem specifická buněčná odpověď (obrázek 9). Buněčná odpověď závisí na spektru receptorů, které jsou u buňky vytvořeny. Buňky různých tkání se liší v zastoupení receptorů a následně i ve schopnosti reagovat na vnější podněty.

Buňky se stejným receptorem mohou odpovědět jiným způsobem z toho důvodu, že u nich proběhne rozdílné zpracování signálu. Vstupní signál sám o sobě nese velmi málo informace o tom, jakou vyvolá buněčnou odpověď.

Poměrně rychlá buněčná odpověď (v řádu sekund až minut) nastává, v případě, že konečným efektoem signální dráhy je enzym metabolismu nebo protein buněčné kostry (cytoskeletu) (Alberts *et al.*, 2008). Takto působící signální dráhy pouze mění funkci již existujících proteinů v buňce. Pomalejší odpověď (v řádu minut až hodin) nastává u signálních drah, které mění genovou expresi. Signál se přenesse na transkripční faktor, který prostupuje do jádra buňky, kde mění genovou expresi.

2 Literatura

- Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P.** 2008. *Molecular Biology of the Cell*. Garland Science. 5th edition, chapter 15.
- Berg J. M., Tymoczko J., Stryer L.** 2002. *Biochemistry*. W H Freeman. 5th edition, chapter 15.
- Buday L.** 1999. Membrane-targeting of signalling molecules by SH2/SH3 domain containing adaptor proteins. *BBA*. 1422(2): 187–204.
- Buday L., Tompla P.** 2010. Functional classification of scaffold proteins and related molecules. *FEBS J*. 277: 4348–4355.
- Chen Z. J., Sun L. J.** 2009. Nonproteolytic Functions of Ubiquitin in Cell Signaling. *Mol Cell*. 33(3): 275–86.
- Ciechanover A.** 2005. Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol*. 6: 79–87.
- Erzsébet L., Welti S., Scheffzek K.** 2012. Inhibition and Termination of Physiological Responses by GTPase Activating Proteins. *Physiol Rev*. 92(1): 237–272.
- Hanks S. K., Hunter T.** 1995. In the Beginning, There Was Protein Phosphorylation. *JBC*. 9(8): 576–596.
- Khazifov K., Lattazi G.** 2009. G protein inactive and active forms investigated by simulation methods. *Proteins*. 75(4): 919–930.
- Kyriakis J. M.** 2014. In the Beginning, There Was Protein Phosphorylation. *JBC*. 289(14): 9460–9462.
- Reece J., Campbell N.** 2002. *Biology*. Benjamin Cummings. 5th edition.
- Robinson D. R., Wu Y., Lin S.** 2000. The protein tyrosine kinase family of the human genome. *Oncogene*. 19(49): 5548–5557.
- Woelk T., Sigismund S., Penengo L., Polo S.** 2007. The ubiquitination code: a signalling problem. *Cell Division*. 2(11): doi:10.1186/1747–1028–2–11.
- Yarden Y., Ullrich A.** 1988. Growth factor receptor tyrosine kinases. *Annu Rev Biochem*. 57: 443–478.