

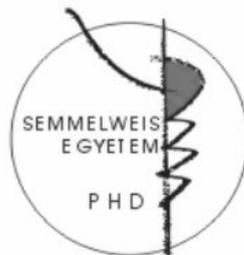
Dynamic analysis of complex biological networks

PhD Theses

Kristóf Zsolt Szalay

Semmelweis University

School of Molecular Medicine



Supervisor: Dr. Peter Csermely, member of the Hungarian Academy of Sciences

Official opponents: Dr. Peter Hamar, Associate Professor, Ph.D.

Dr. Beata Oborny, Associate Professor, Ph.D.

President of the exam committee: Dr. Andras Falus, member of the Hungarian Academy of Sciences

Members of the exam committee: Dr. Peter Enyedi, Professor, D.Sc.

Dr. Gergely Palla, Research Professor, Ph.D.

Budapest

2014

Introduction

Networks are all around us. Our friendships, human societies, the Internet, economical and ecological systems – and of course, our bodies themselves – are all composed of different networks. The concept of networks is an abstract model for describing complex systems, based on the idea that complex systems become more understandable if we analyze the system based on its individual elements and their interactions rather than trying to analyze the whole system as a "black box".

The immense knowledge acquired in different fields by network science could be fruitfully exploited in the life sciences observing the deluge of data in the last 10-15 years. In fact, most research topics in the life sciences are concerned with observing complex biological systems, thus gaining a deep understanding of the innate behavior of such systems is a crucial task.

In order to create a network model capable of giving strong predictions on the behavior of the observed system, it is essential to "turn on" the system, to have a dynamic, time-dependent model of the network besides the connectivity information itself.

Systems theory (or the theory of signals and systems) defines a system as a physically interconnected set of elements capable of producing an output signal in response to an input signal. A signal in this context could be any time-variant physical variable having the ability to transmit information.

Such a system could have one or more **inputs**, one or more **outputs** and may also include **state variables**, which contain the "memories" of the system enabling it to respond differently to the same input at different cases taking previous inputs into account. Inputs could affect both the state variables or the outputs directly, while state variables could affect each other or the outputs.

The presented systems theory approach is readily transferrable to the area of networks. The states of individual nodes could be mapped to state variables, while the outside perturbations affecting the network could be mapped to inputs. The

resulting time series data of arbitrarily selected interesting nodes may become the outputs.

There is another interesting area of research in dynamical systems theory which could be exploited in network analyses. A complex system without any inputs cannot have any arbitrary combination of state variables in equilibrium. Left alone, the state of the system will evolve to one of its steady states, called *attractors* in systems theory. A complex system – even one having a lot of attractors – still has far less attractors than the possible number of state variable combinations.

The first notion of attractors in biology could be attributed to Waddington, described by the "epigenetic landscape" in the 1957 book *The Nature of the Genes* (Figure 1). This attractor-based description of biological systems was only regarded as an interesting fact until Kauffman had shown in his seminal articles that biological systems *by design* tend to have a small number of attractors.

At that time, however, systems-level information in biology was scarce. Kauffman himself suggested that there could be 1 million different human genes. After sequencing most of the human genome in 2001 and publishing the complete sequence in 2003 in parallel with the exponential growth of data storage and computational speed, systems-level experiments in biology only became feasible in the mid-2000s.

The first attractor search experiment was conducted in 2006 by Fauré *et al.* In their 10-node model of the cell cycle, only two steady states were found, one quiescent and one proliferating. In the following years, attractors of other signaling networks were also described. The network of T-LGL leukemia was investigated in two separate papers, so I chose this network for the verification of my algorithms due to the large amount of previous analyses.

It was also previously shown in literature that attractors of signaling systems correspond to possible physiological or pathological states of the system. Cancer-related dysregulations and other complex diseases were also suggested to correspond

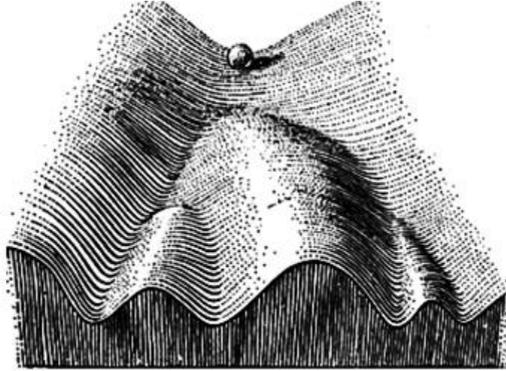


Figure 1: The 1957 image of Waddington representing the process of cell differentiation. The ball rolling down the "epigenetic landscape" symbolizes the states of the system during differentiation. The four valleys at the bottom represent the possible steady states, that is, the attractors of the differentiation process.

to changes in the attractor structure. This makes it feasible to think about cellular systems as working near their attractors in the general case.

Aims

1. In line with one of the most important targets of network science, the first aim of my work was analysing the dynamics of biological networks. In order to achieve this goal, I have created a fast, general dynamic network analysis software.
2. The second aim of my work was extending the software with a module capable of identifying the dynamic attractor structure of the networks. The attractor search module was successfully tested on multiple – including biological – networks, with and without noise.
3. The third aim of my work was creating a "designer" module for the software enabling the rational design of interventions which, when applied to the network

in a given starting state, could transfer the network to a given target state. This module was tested on signaling networks describing the connections of proteins implicated in certain types of cancer. The designed interventions could later serve as drug targets for combinatorial, or even personalized cancer treatments.

Methods

The analyses have been performed with a newly developed C++ program called Turbine, since there is no software currently available with similar capabilities.

The core of the program is a fast network simulator having the ability to calculate 1000 integration steps in 30 seconds on a network having 10 million nodes and links combined. Simulations of smaller networks only takes a fraction of a second.

By realizing that the behavior of such smaller networks could be simulated in less than a second, it became feasible to run an ensemble of simulations iteratively on a single network, enabling to perform new and useful analysis methods.

If we execute a set of simulations by randomly changing the initial state, we get an attractor search tool, capable of mapping the steady states of the system.

Turbine supports attractor search, and can also recognize limit cycle and some limit torus type steady states as well as fixed points, while using any type of continuous or discrete dynamics and also with the application of external constant or periodic perturbations.

If we not just randomly change the parameters of the simulation, but assign a governing artificial intelligence algorithm with this task, another new set of methods become available. A good example of this is the intervention design workflow of Turbine. Intervention design is capable of finding a near-minimal set of interventions, which, when applied to the network in a given undesirable "diseased" state, will transfer the system to a given, more desirable "healthy" state. The program uses a genetic algorithm for this purpose (Figure 2).

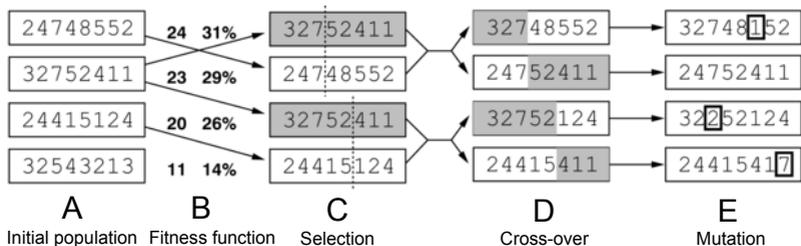


Figure 2: The steps of a genetic algorithm. From our population (A), we select pairs of parents using their fitness function, which describes the problem condition to solve (B). The selected parents (C) give rise to a new generation of offspring candidates using cross-over (D). The new generation is finalized by optionally applying mutations in the genome of some elements of the new population (E). The population in the case of Turbine are sets of interventions, and the fitness values are calculated by simulation. Original figure: http://project.mit.bme.hu/mi_almanach/books/aima/ch04s03

Results

General modeling

The first model created in the course of development was a generic dynamic model based on the idea of communicating vessels. This dynamic model can give a general, coarse approximation of the behavior of a network concerning perturbations if a specific model is not available due to the type of the network or data constraints. The literature suggests that changes in free protein concentrations can be described with similar exponentially declining functions as the ones given by the communicating vessels model. Using this model, I have defined a new type of *centrality measure* ranking the relative importance of the nodes in the network. A lot of such centrality measures are available, I named this dynamic-based new measure "perturbation centrality".

Protein structure network analysis. The first experiments were conducted on the *E. coli* metionil-tRNA synthetase (MetRS) networks described by Ghosh and Vishveshwara in 2007. I created the protein structure network from the original PDB

files using the RINerator software and calculated the perturbation centrality values for all nodes (amino acid residues). Afterwards, I have subtracted the perturbation centrality values of the nodes of the closed (tRNA-binding) conformation from the perturbation centralities of the corresponding nodes of the open conformation, and analyzed the top 20% of the resulting values. The most important points were grouped in two spots, one near the Met-binding active center, the other near the tRNA anticodon loop. This way, the perturbation centrality values successfully identified the two most important functional spots of the protein. Noteworthy, the tRNA was not included in the simulation itself, so the amino acid conformations of the protein itself may have changed during the transition in a manner that facilitates communication between the important spots in the closed conformation.

Protein-protein interaction network analysis. I have conducted the next experiment on another type of undirected network, the protein-protein interaction network of budding yeast (*S. cerevisiae*). For this experiment, I used differently stressed versions of the yeast protein-protein interaction network as described in the 2011 article of Mihalik and Csermely. After calculating the perturbation centrality values, I have compared the differences of the perturbation centrality values in the different stress types to the perturbation centralities measured in the resting network.

The resulting cell functions were in good agreement with known yeast stress responses. In all types of stress, the proteins with the largest increase of perturbation centralities were significantly enriched in the GO terms "response to stimulus", "carbohydrate metabolism", "trehalose metabolic process", "glycogen metabolic process", and in heat shock and osmotic stress, the term "response to stress" was also significantly enriched. This suggested change of the carbohydrate metabolism in stress is already described in the literature. Most of the enriched terms of proteins having the largest decrease of perturbation centrality values corresponded to the down-regulation of protein translation and ribosome synthesis, which are also well

known stress responses in yeast. Based on these results, the perturbation centrality measure also successfully identified important, known stress responses in yeast, so the network-based importance measured by the perturbation centrality measure and changes thereof seem to correlate well with the real-world functional importance of the network elements.

Specific dynamics for signaling systems

It is naturally expected that much more accurate and important predictions could be made about a network if we utilize a specific dynamic model, which more accurately describes the behavior of the network. I chose the process of protein signaling for this purpose, since there are multiple examples in the literature that even simple, relatively abstract models of signaling can result in accurate, experimentally verifiable predictions.

For the next experiments, I have created two dynamic signaling models for Turbine: a Boolean model akin to the one previously used in modeling T-LGL leukemia, and one recreating the behavior of the human cancer signaling network of Fumiã and Martins. I have tested both models on the corresponding networks found in the original articles, and found that the Turbine models perfectly reproduced the original results.

Verification on the T-LGL network. The T-LGL signaling network consists of proteins implicated in the activation and activation-induced cell death of cytotoxic T lymphocytes, containing full Boolean rulesets for each node describing the conditions of activation and inactivation. The authors also identified both the attractors of the system and possible interventions using a network simplification method in a follow-up article, so this system was a clear target for the verification of the artificial intelligence algorithms of Turbine.

In the resting T-LGL network Turbine found only a single, apoptotic attractor in 100,000 simulations. In agreement with the original article, by activating the IL15,

Stimuli and PDGFR inputs, two other proliferating attractors (very similar to each other) appeared corresponding to the LGL leukemic state.

Next, I have designed interventions using Turbine which could transfer the LGL attractor to the apoptotic attractor. 19 of the 20 single-target interventions found matches the 19 consensus interventions described in the original article, the 20th (CREB activation) is also presented as a high confidence intervention in the source paper. Individual elements of the combination interventions found by Turbine also appear in the source article as high-confidence intervention targets (sFas, TBET, Fas, FLIP).

Thus, we can state that both the attractor search and the intervention design tools of Turbine were capable of both reproducing the original results and give new combinatorial targets, verifying the functionality of the new methods.

Effect of noise on attractors. Observing the wide range of individual, effective treatments raises a suspicion that the presence of the proliferative attractors in the network without any special dysregulations may only be an interesting effect and actual leukemic conditions may be different. This notion is supported by the literature, where mutations of the Fas pathway were implicated in T-LGL leukemia.

We know that no method of communication is perfect, errors and artifacts always happen under real-world conditions. My hypothesis was that the important attractors of the system need to have good noise tolerance, so the important attractors should be large enough to resist the effects of system noise, otherwise the system could easily transfer to a pathological attractor.

I have conducted these experiments on a signaling network consisting of proteins implicated in cancer where a continuous dynamic model could be applied. I have used the only microenvironment where two types of attractors are present, a physiological and a pathological one. In this environment, the cell is under genotoxic stress, with abundant growth factors and nutrients, without TNF- α , and in normoxia. The physiological response in this case is apoptosis, while the pathological

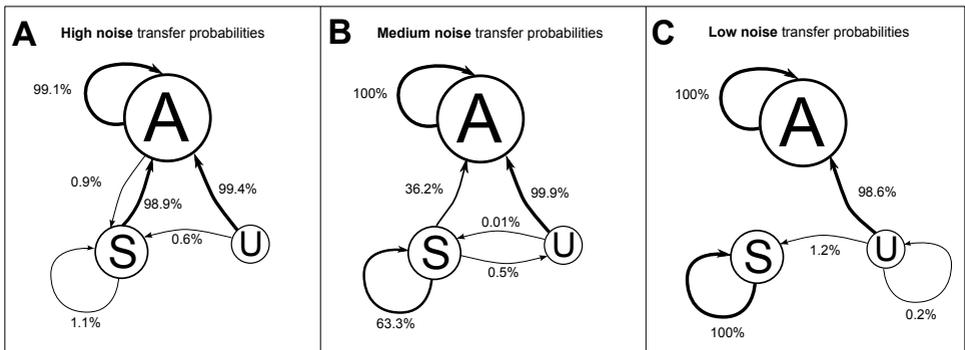


Figure 3: The attractors of the cancer signaling network and their transition probabilities under different amounts of noise. Using a high or medium level of noise (Panels A and B) only the largest, apoptotic attractor (**A**) was stable, while in low-noise condition (Panel C), the stable proliferating attractor (**S**) was also stable. The natural noise condition is likely most similar to the medium noise level (Panel B).

one is proliferation. In my experiments with Turbine, I have found two separate proliferating attractors. In the attractor called "stable proliferating" only the cyclins and the related proteins showed a cyclic change of activity levels, while in the "unstable proliferating" attractor, other unrelated proteins also changed their activity levels during the cell cycle.

Next, I have conducted a series of experiments where the three different attractors were set as starting states, and applied a different amount of pink noise to them, 1000 times for each attractor and noise level (small: 5 steps of 0.001 units, medium: 10 steps of 0.1 units, large: 30 steps of 0.4 units). After running the simulations, I checked if the resulting steady state is in a different attractor than the starting point. The results are shown on Panels A, B and C of Figure 3, where **A** marks the apoptotic attractor, **S** the stable proliferating, and **I** the unstable proliferating one. The physiological apoptotic attractor was considerably stable due to its size, only the large noise levels were enough to cause a transfer to the stable proliferating state in 0.9% of the cases. The unstable proliferating attractor was unstable even using the lowest noise level and was mostly transferred in the direction of the apoptotic

attractor (98.6%), and in a small part towards the stable proliferating attractor (1.2%). The stable proliferating attractor proved to be stable under the low noise condition, but was often transferred in the direction of the apoptotic attractor under higher noise levels (36.2%). Thus, if the physiological attractors are also much larger than the pathological attractors in the general case, the natural noise could *help* the regulation of signaling. The results are also consistent with the hypothesis that knowledge of the largest attractors of a system could be enough to describe its behavior, since smaller attractors seem to be unstable at natural noise levels.

Identifying driver mutations. While the identification of attractors and the analysis of the data of attractor states could prove to be valuable in research, in my case the goal of finding attractors was to find full state descriptions of the system that could become logical start and target states for the intervention design process and thus allow predicting specific interventions.

However, the landscape of a cell suffering from mutations is likely different from the landscape of the healthy cell. It was proposed that the cancerous attractors may not be available or present in a normal cell, and the mutations change the landscape in a way that these "disease attractors" become available.

This assumption has a lot of merit. A coding mutation, when it changes the activity level of a protein is a *continuous* perturbation, so it is able to change the attractor landscape of a cell.

The significant resistance encountered in tumors led me to the next hypothesis, that the important, driver mutations of a tumor may change the attractor landscape of a cell in a way that the pathologically proliferating attractor becomes much larger than the physiological attractor, at which point the previously helpful noise becomes a hindrance, as it now makes leaving the proliferating attractor as difficult as it made leaving the physiological attractor difficult in a healthy cell.

Thus, as a next step, I tried to identify mutations or combinations thereof in the Fumiã and Martins network where the proliferating attractor became significantly

Table 1: Mutation sets leading to a dominant proliferating attractor; *off* means a loss-of-function, *on* means a constitutive activation-type mutation. The function loss of P53, ATM/ATR and NF- κ B were all individually enough to make the proliferating attractor dominant. NF- κ B was not shown, because the "quasi-proliferative" attractor obtained with an NF- κ B function loss has a different cyclin activation sequence than the consensus proliferating activation sequence described in the original article.

Number	Mutations
#1	P53 <i>off</i>
#2	BCL-2 <i>on</i> , PTEN <i>off</i>
#3	S-NAIL <i>on</i> , CHK1/2 <i>off</i>
#4	ATM/ATR <i>off</i>
#5	CHK1/2 <i>off</i> , PTEN <i>off</i>
#6	BCL-2 <i>on</i> , AKT <i>on</i>
#7	E-CADH <i>off</i> , CHK1/2 <i>off</i>
#8	PTEN <i>off</i> , IKK <i>on</i>
#9	AKT <i>on</i> , CHK1/2 <i>off</i>
#10	PTEN <i>off</i> , BAX <i>off</i>
#11	CHK1/2 <i>off</i> , AKT <i>on</i>
#12	AKT <i>on</i> , E-CADH <i>off</i>
#13	AKT <i>on</i> , BAX <i>off</i>

larger than the apoptotic one, or those, where the apoptotic one even disappeared completely. I have identified 13 such mutation sets as shown in Table 1.

The results are promising. P53 is a well-known tumor suppressor, while the mutations of ATM/ATR, PTEN, AKT, BCL-2, BAX, CHK1/2 and IKK with a mutation type corresponding to the ones found in Table 1 have all been described in tumor development in the literature. This makes it very likely that the pathological processes leading to tumor formation could be described meaningfully using attractor landscapes.

Intervention design. After mapping the changed attractor landscapes caused by the driver mutations, the next important question to answer was the possibility of designing a specific intervention for the modified landscapes that can – during the course of the treatment – move the network to an apoptotic state.

Table 2: Mutation-specific interventions. The *inh.* mark means inhibition, *act.* means activation of a given protein. The *off* and *on* means function loss or constitutive activation type mutations of a given protein, respectively. Interventions marked with italics were predicted to be toxic based on simulations using a healthy cell model, while the ones marked with an asterisk shows that the intervention did not result in a complete, steady apoptosis, only the cyclic on/off switching of the caspase proteins.

Number	Profile	Interventions
#1	P53 <i>off</i>	<i>RTK inh.</i> AND (<i>JNK act.</i> or <i>TGF-β act.</i> or <i>BAX act.</i>)
#2	BCL-2 <i>on</i> , PTEN <i>off</i>	<i>AKT inh.</i> AND (<i>TGF-β act.</i> or <i>BAX act.</i> or <i>JNK act.</i>) *
#3	S-NAIL <i>on</i> , CHK1/2 <i>off</i>	<i>RTK inh.*</i> or <i>RAS inh.*</i> or <i>PI3K inh.*</i> or <i>P53 act.</i>
#4	ATM/ATR <i>off</i>	<i>P53 act.</i>
#5	CHK1/2 <i>off</i> , PTEN <i>off</i>	<i>PDK1 inh.</i> * or <i>AKT inh.</i> *
#6	BCL-2 <i>on</i> , AKT <i>on</i>	N/A
#7	E-CADH <i>off</i> , CHK1/2 <i>off</i>	<i>AKT inh.</i> * or <i>PDK1 inh.</i> * or <i>RTK inh.</i> * or <i>RAS inh.</i> * or <i>PI3K inh.</i> *
#8	PTEN <i>off</i> , IKK <i>on</i>	<i>NF-κB inh.</i> or <i>BCL-2 inh.</i>
#9	AKT <i>on</i> , CHK1/2 <i>off</i>	<i>P53 act.</i> or (<i>IKK inh.</i> AND (<i>VHL inh.</i> or <i>P14 act.</i> or <i>ROS act.</i>))*
#10	PTEN <i>off</i> , BAX <i>off</i>	<i>AKT inh.</i> or <i>PDK1 inh.</i>
#11	CHK1/2 <i>off</i> , AKT <i>on</i>	<i>P53 act.</i>
#12	AKT <i>on</i> , E-CADH <i>off</i>	<i>IKK inh.</i> or <i>NF-κB inh.</i> or <i>BCL-2 inh.</i>
#13	AKT <i>on</i> , BAX <i>off</i>	N/A

This experiment was conducted by setting the starting state of the network to the largest attractor encountered in a given mutation profile, and any apoptotic state as the target. Since we now had to modify the attractor landscape itself, the applied perturbations were also constant changes in the activity levels of certain proteins.

The interventions designed for each of the mutation profiles are shown in Table 2. It is reassuring that a lot of results contain a receptor tyrosine kinase inhibition element (*RTK inh.*), since the first and the most widely used targeted cancer therapy treatments are receptor tyrosine kinase inhibitors (cetuximab (Erbix), panitumumab (Vectibix): EGFR, sunitinib (Sutent): wide spectrum RTK inhibitor,

sorafenib (Nexavar): VEGFR, PDGFR). Other frequently occurring intervention targets have also been identified in the literature as promising treatment targets such as inhibitors of BCL-2, PDK1, IKK, AKT, PI3K and NF- κ B.

Designing for non-toxicity. While individual elements of interventions seem to be supported by literature, the full intervention sets could be predicted to be toxic in some cases (#1, which as the P53 loss case is crucially important, and also the #2, #4 and #11 profiles; in the last two cases the toxicity could be attributed to the direct activation of P53) when examining the effect of the treatment in a cell without any mutations, in a microenvironment where only a single quiescent (non-proliferating, non-apoptotic) attractor is present. The mentioned interventions also move these cells into apoptosis, suggesting that the intervention is not specifically targeting erroneously proliferating cells but will rather kill most affected cells.

Thus, the intervention design algorithm needed to be overhauled. I have extended the algorithm in a way that each intervention candidate was simulated twice, one on the erroneously proliferating cell, with apoptosis being the target state, and on a resting cell without any mutation where an apoptotic or proliferating result incurred a fitness penalty. Figure 3 shows the resulting intervention sets.

An interesting observation is that multiple intervention sets contained von Hippel-Lindau protein inhibition elements (VHL *inh.*). This is a counter-intuitive result, since the loss of function of the VHL protein leads to the appearance of a certain type of renal tumor in von Hippel-Lindau syndrome. However, on closer examination, it was revealed that these interventions were never required for the efficacy of the intervention, but for protecting the healthy cell against its harmful effects. This could mean that a temporary inhibition of the VHL protein could dampen the toxicity of targeted cancer treatments, or allow the utilization of compounds previously abandoned on the grounds of their toxicity.

Table 3: Non-toxic interventions. The *inh.* mark means inhibition, *act.* means activation of a given protein. The *off* and *on* means function loss or constitutive activation type mutations of a given protein, respectively. These intervention sets do not harm healthy cells based on the simulation results while at the same time they move the mutated cells into apoptosis.

Number	Mutations	Non-toxic intervention
#1	P53 <i>off</i>	BCL-XL <i>inh.</i> AND VHL <i>inh.</i> AND (PDK <i>inh.</i> or AKT <i>inh.</i> or PI3K <i>inh.</i>)
#2	BCL-2 <i>off</i> , PTEN <i>off</i>	TGF-B <i>act.</i> AND P90 <i>inh.</i> AND PDK1 <i>inh.</i>
#4	ATM/ATR <i>off</i>	VHL <i>inh.</i> AND (RAS <i>inh.</i> or PHD <i>inh.</i>)
#11	CHK1/2 <i>off</i> , AKT <i>on</i>	BCL-2 <i>inh.</i> AND PHD <i>inh.</i> AND SMADE2F <i>inh.</i>

Passenger mutations. When a malignant tumor becomes detectable, it already contains many more mutations in addition to the drivers. The mutated genome could contain passenger mutations which make the cell resistant to certain types of targeted treatments. The best example for this is the constitutive activation of the kRAS gene which renders the tumor extremely resistant to EGFR inhibitor therapies.

Therefore, I examined the mutation profiles having a corresponding intervention containing a receptor tyrosine kinase inhibition element (#1, #3, #7) to see if their effectiveness is indeed lost when an additional RAS activating mutation is applied to the cell.

Indeed, this was exactly what happened. All receptor tyrosine kinase based intervention sets lost their apoptotic effect. However, intervention sets targeting other pathways remained efficient, for example the inhibition of PI3K was still effective with the added RAS mutation in the profiles #3 and #7, along with the PDK1 and the AKT inhibition in the case of profile #7.

However, in the case of profile #1, all simple intervention sets contained RTK inhibition. Hence, because the function loss of P53 happens in the majority of

tumors, I chose to design a new intervention set for this special case (P53 *off*, RAS *on*).

The simplest such intervention set contained three individual elements: the inhibition of RAF, activation of BAX and the inhibition of either PI3K or PDK. This intervention set moves the mutated cell completely into apoptosis, but is also predicted to be toxic for healthy cells. It seems noteworthy to state that the first BRAF inhibitor called Vemurafenib was launched to market in August 2011 with indication for treating late-stage melanoma. The simplest non-toxic intervention for the P53 function loss, RAS activated case was predicted to have lower (but still significant) efficacy, and was the same as the intervention set identified for the case without the additional RAS mutation: BCL-XL inhibition, VHL inhibition and (for example) AKT inhibition.

Conclusions

The most important results of my work can be summarized as follows:

- I have created a fast, generic network simulator software named Turbine, which can efficiently simulate extremely large networks, while at the same time allows performing higher-order analysis types such as attractor search and intervention design on smaller networks.
- I have created a general dynamic model for Turbine, which I have termed communicating vessels and a specific one for more accurate analysis of signaling networks.
- I have used the aforementioned generic network to define a centrality measure which was able to highlight both tRNA binding spots on the *E. coli* Met-tRNA synthetase enzyme unlike other widely used centrality measures. I also demonstrated that the perturbation centrality values follow the ordered-

unordered transition of protein secondary structure. The same measure was also able to identify the most important stress reactions in the budding yeast.

- I have shown that the identification of only the largest attractors of a system could be enough to describe their behavior, since small attractors are likely unstable under noise levels observable in natural systems.
- I have demonstrated that many dysregulations common to tumors change the attractor landscape of the signaling network in a way that the pathological attractor becomes the largest one.
- Finally, after applying these dysregulations on the used signaling network, I demonstrated that the intervention design tool is capable of rationally designing intervention sets which shift the dysregulated cells into apoptosis, while at the same time causing smaller or no harm to the healthy cells according to the simulations.

Acknowledgements

First and foremost, I would like to thank **Prof. Péter Csermely** Member of the Hungarian Academy, my advisor, and by now also my friend, for his invaluable advice sharing his experience and knowledge with me, and also for working hard on making a scientist of me.

I express my gratitude towards **Prof. József Mandl** Member of the Hungarian Academy and **Prof. Gábor Bánhegyi** for providing an opportunity for me to work at the Institute.

Special thanks to **Prof. Csaba Sőti** for giving the chance to try myself at wet-lab work and **Milán Somogyvári** for teaching me *C. elegans* experimental methods.

I would also like to thank **Ákos Szőts** for writing the CUDA extension of Turbine and **Adam Portschy** for finding a way to make Turbine work in the BOINC system.

A lot of thanks for **Ágoston Mihalik** for the stressed yeast dataset and **Dr. Amit Ghosh** and **Prof. Saraswathi Vishveshwara** for the MetRS protein structure network data.

I am thankful for all members of the **LINK-Group** for their ideas, advice and the work together.

I am extremely grateful for my mother, **Éva Szalay** and my father **Zsolt Szalay** for supporting and trusting me on my way until now.

My warmest thanks for my wife, **Eszter Pintér** for standing beside me in the last years and for her unbroken support. I hope I will be able to reciprocate in the following years.

List of publications

Publications related to the thesis

Szalay **KZ**, Csermely P (2014) Attractor structures of signaling networks: consequences of different conformational barcode dynamics and their relations to network-based drug design. *Mol Inf* 33: 463–468. **IF: 2.013**

Szalay **KZ**, Csermely P (2013) Perturbation centrality and Turbine: a novel centrality measure obtained using a versatile network dynamics tool. *PLoS ONE* 8: e78059. **IF: 3.534**

Farkas IJ, Korcsmáros T, Kovács IA, Mihalik Á, Palotai R, Simkó GI, **Szalay KZ**, Szalay-Bekő M, Vellai T, Wang S, Csermely P (2011) Network-based tools for the identification of novel drug targets. *Science Signaling* 4: pt3.

Other publications

Veres DV, Gyurkó MD, Thaler B, **Szalay KZ**, Fazekas D, Korcsmáros T, Csermely P (2014) ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res*, doi: 10.1093/nar/gku1007 **IF: 8.808**

Csermely P, Hódsági J, Korcsmáros T, Módos D, Perez-Lopez AR, **Szalay K**, Veres DV, Lenti K, Wu LY, Zhang XS. (2014) Cancer stem cells display extremely large evolvability: alternating plastic and rigid networks as a potential mechanism. *Semin Cancer Biol* in press, doi: 10.1016/j.semcancer.2013.12.004.

Hegedűs T, Gyimesi G, Gáspár ME, **Szalay KZ**, Gangal R, Csermely P (2013) Potential application of network descriptions for understanding conformational changes and protonation states of ABC transporters. *Curr Pharm Des* 19: 4155–4172. **IF: 3.288**