Charmi Panchal | Sepinoud Azimi | Ion Petre

Generating the logicome of a biological network

Turku Centre *for* Computer Science

TUCS Technical Report
No 1139, June 2015

Generating the logicome of a biological network

Charmi Panchal
Computational Biomodeling Laboratory
Turku Centre for Computer Science
Åbo Akademi University, Joukahaisenkatu 3-5A, 20520 Turku, Finland
cpanchal@abo.fi
Sepinoud Azimi
Computational Biomodeling Laboratory
Turku Centre for Computer Science
Åbo Akademi University, Joukahaisenkatu 3-5A, 20520 Turku, Finland
sazimi@abo.fi
Ion Petre
Computational Biomodeling Laboratory
Turku Centre for Computer Science
Åbo Akademi University, Joukahaisenkatu 3-5A, 20520 Turku, Finland
ipetre@abo.fi

TUCS Technical Report
No 1139, June 2015

Abstract

There has been much progress in recent years towards building larger and larger computational models for biochemical networks, driven by advances both in high throughput data techniques, and in computational modeling and simulation. Such models are often given as unstructured lists of species and interactions between them, making it very difficult to understand the *logicome* of the network, i.e. the logical connections describing the activation of its key nodes. The problem we are addressing here is to predict whether these key nodes will get activated at any point during a fixed time interval (even transiently), depending on their initial activation status. We solve the problem in terms of a Boolean network over the key nodes, that we call the logicome of the biochemical network. The main advantage of the logicome is that it allows the modeler to focus on a well-chosen small set of key nodes, while abstracting away from the rest of the model, seen as biochemical implementation details of the model. We validate our results by showing that the interpretation of the obtained logicome is in line with literature-based knowledge of the EGFR signalling pathway.

Keywords: Biomodeling, Boolean circuit, Logicome, EGFR pathways, ODE models

TUCS Laboratory
Computational Biomodeling Laboratory

1 Introduction

One of the central topics of interest in systems biology is to identify the functionalities of a living cell and to understand how the huge number of interactions within a cell facilitate such functionalities. The set of complex and involved interactions lead to obtaining a large number of collected experimental data as well as complex networks. These broad sources of information can prove to be very useful in providing a realistic life picture of the phenomenon under study, but can also make it difficult to analyze the system and can cause inaccuracy in predicting the system's behavior. Identifying the main players within a network and understanding how they activate each other can help to overcome these difficulties.

There have been many studies on the logical modelling of biological networks; for example, [29], [6], [5], [4] discuss the correspondence between Boolean networks and ODEs; for an introduction to Boolean networks and ODEs we refer to [13] and [14] respectively. Fuzzy logic was used in [19] to yield the logical models corresponding to the biological networks. As a different approach, [26] build the Boolean logic models by training a literature-based prior knowledge network against biochemical data. These studies mainly proposed approaches where the full understanding of the biological aspects of the phenomenon under study was crucial and the goal was to obtain a mathematical model reproducing that understanding. Our study goes in the reverse direction: it starts from an existing mathematical model and aims to obtain an abstract, high-level understanding of the functionality of the biological network underlying the model. Our goal is to obtain a logical description of the activation conditions between the key nodes of the network; even in the case when one starts from a detailed biological model going towards the mathematical model, our reverse engineering approach brings a new higher-level understanding of the functionality of the biological model we started from. The result of our approach is formulated as a Boolean network whose nodes are the key species we focus on; we coin the term *logicome* to name this network.

Extracting a Boolean network model from a given ODE-based model is a well-studied topic with many different solutions, see, e.g., [29] for a recent new solution and a good overview of the topic. Typically, the Boolean network model is seen as a companion of the ODE-based model, compensating for the lack of detailed kinetic-level data for the model, or allowing for alternative global analysis of model dynamics, such as attractor- or multi-stability- analysis, see [29]. A key step going from an ODE model to its corresponding Boolean network model is the discretization scheme allowing to replace continuous variables with their corresponding 0/1 variables. This is typically done by sampling the numerical integration of the continuous variables at different time points and by discretizing their values at those points. This leads to the dynamics of the Boolean model being interpreted in terms of discrete time series reflecting the behavior of the original ODE model. Our approach is coarser: we aim to capture the activation of the key nodes of the model over the whole time interval (to be thought of

as much larger than those involved in the discretization of ODE models). This includes capturing the transient activation of a node over that interval, even if at the extremities of the interval the node may be inactive. The result is a Boolean network that accompanies the starting ODE model in terms of describing asynchronous cause-effect relationships among its key nodes over a fixed time interval.

As a case study we focus on the *EGFR (epidermal growth factor receptor) signaling pathway*. Epidermal growth factors are key players in cell proliferation, survival, migration and differentiation. EGFR signaling also has a major role in EGFR-dependent signal transduction, see [28]. Therefore, understanding their behavior is crucial in any cancer related studies, see [20]. For more information on EGFR signaling pathways we refer to [31], [2], and [28].

This paper is organized as follows. In Section 2, we present our methodology to infer the logicome of biochemical networks. In Section 3, we introduce the case study we used in this paper. In Section 4 we present the results of applying the method to the case study and analyze the produced results and finally we conclude with some discussions in Section 5. All the models and data files used in this paper can be found at: `http://combio.abo.fi/research/logicome-models-2/`.

2 Methodology

In this section we present our method to infer the logicome of an ODE-based model. The steps are described in a generic way – their detailed implementation is up to the modeler and it depends on the case study. In the next section we discuss one particular way in which we used this method in the case of the EGFR pathway.

Step 1 – Setup. We start with an ODE model for a biochemical network. We assume also to have a set of "key nodes" whose influences over each others' activation we aim to capture. The choice of the key nodes from among the variables of the ODE model depends on the modeler and on the network under study.

Step 2 – Discretization. To be able to describe the logicome of a network in terms of Boolean network, we need to translate continuous simulation data to a Boolean, "on/off"-based language. Therefore, as the second step we incorporate a discretization algorithm into our method. Many discretization methods exist, see for example [18], and [25]. In this study our discretization step is based on a threshold-based approach in which we assign "1" to a species if its simulation value is above a given threshold, and "0" otherwise. The precise choice of the threshold depends on the network under study.

Step 3 – Simulation. We simulate all possible knock-out mutants; in other words, all models where the key species are turned on/off in all possible combinations. We then apply to each simulation result the discretization step to

obtain the Boolean results corresponding to each mutant. In this way we produce a truth table describing the output of each simulation as a Boolean function with the key nodes as its Boolean variables. Translating the input Boolean values of the key nodes to absolute numerical values to be used in the simulation can be done in several different ways, depending on the case study. For example, the 0 value for a Boolean key node may be translated to value 0 for the corresponding variable(s) in the original model, while value 1 may be translated to the threshold value chosen for that variable in Step 2.

Step 4 – Logicome generation. In this step we generate the logicome corresponding to the given biochemical network from the produced truth table in the previous step. Different algorithms can be used to implement this step, see for example [1], [16], [11], [21]. In this paper we use the *Logic Friday* tool which incorporates the *Espresso algorithm* proposed in [21].

3 Case-study: the EGFR pathway

We focus in this paper on a signaling network that is strongly associated with the development of cancer processes: the *EGFR signaling pathway*. In the following subsections we provide a brief biological background and some computational details of this model.

3.1 Biological background

The epidermal growth factor receptor (EGFR) pathway regulates several important cellular processes including cell proliferation, survival, differentiation and development, see [20]. Because of its association with the various types of cancer processes, this pathway is a widely investigated signal transduction system. The EGFR pathway can be seen as a union of several smaller pathways, also called *modules*, see [30] and [3]. The proteins situated at the intersection between these modules are called *interface species*. The analysis presented in [10] identifies the locations of oncogenes and essential components of the EGFR signaling cascade that define most of the interface regions. Our model is adopted from [30] that uses the model originally presented in [27] and implements it in the stochastic pi-calculus language together with the results identified by [10].

We follow the approach of [30] and their modularization of the EGFR signaling pathway in the following 7 modules: EGF, Grb2, Ras-Shc-Dependent /Independent, Raf, MEK, and ERK. These modules communicate with each other through the following 8 *interface species*: (EGF-EGFR*)2-GAP, (EGF-EGFR*)2-GAP-Grb2-Sos, (EGF-EGFR*)2-GAP-Shc*-Grb2-Sos, Ras-GTP, Ras-GTP*, MEK-PP, Raf* and ERK-PP. We adopt these interface species as the key nodes in our approach.

We briefly describe the functionality of the EGFR pathway focusing mainly on the signal propagation within the interface species, as suggested in [10]; the

modules of the pathway are considered as black-boxes communicating to each other through the interface species. The EGFR is situated on the extracellular surface of the cell and signal transduction begins upon binding of ligand EGF (epidermal growth factor) to EGFR. The EGF-bounded receptor induces dimerization and autophosphorylation of several members of intracellular domains, which leads to the recruiting of several cytoplasmic enzymes and adaptor proteins. This initiates to the activation of two principal pathways, one Shc-dependent and another Shc-independent, that play a significant role in the activation of downstream signaling processes like hydrolyzation of Ras-GDP and activation of Ras-GTP that follows by dissociation of Ras-GTP from the receptor complex. Further dissociation of Ras-GTP makes it inactive and promotes the intrinsic activity of Ras protein regulated by the GTPase activating protein (GAP) that is involved in several crucial cellular processes see [24, 10]. It is assumed that the dissociated Ras-GTP molecule causes phosphorylation of the Raf protein that in-turn double phosphorylates MEK (turning it to MEK-PP) and ERK (turning it to ERK-PP) proteins. The final result of the signaling cascade is the double phosphorylated ERK-PP that further regulates a number of transcription factors and essential proteins for cell differentiation and growth.

A systematic analysis of control mechanisms (including positive/negative feedback loops) underlying EGFR pathway are presented in [30] and [10]. We aim to represent the functional relationships associated with the interface species through a Boolean network – the *logicome* of the EGFR signaling pathway.

3.2 Mathematical model, simulation and discretization

We associated a mass-action ODE-based model, see [14, 8], to the reaction based model of [10]. Each of the 103 variable molecular species of the model in [10] gets a variable in our mathematical model. We wrote the reaction-based model of the EGFR pathway in the COPASI software, see [9], and used its feature to automatically generate the mass-action-based system of ordinary differential equations associated to the model. We call the resulting model our *basic model*.

Following the approach of [30], we simulated in COPASI this model for an EGF stimulus of 4981 molecules/pl which is enough to phosphorylate 50000 EGF-receptors. The simulation was run for 6000 seconds and the time series results of each interface species were collected.

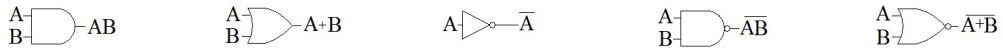
For our method we are interested in analyzing all knock-out mutants where the interface species are active/inactive in all possible combinations. In the knock-out mutants the initial values of the inactive interface species are set to the value 0, while the active interface species are set to a specific threshold value of 1% of that species' maximum value in the simulation of the basic model up to 6000 seconds. Since we considered 8 interface species, we have $256 = 2^8$ knock-out mutant simulations.

3.3 Generating the logicome

Each knock-out mutant can be seen as a particular truth assignment over the 8 Boolean variables standing for the interface species. The results of the 256 knock-out simulations were discretized as follows.

Collecting the outputs of all knock-out mutants can be done in the form of a Boolean function with 8 inputs and 8 outputs.

We used the *LogicFriday* software to generate the Boolean function associated to the EGFR pathway based on the Boolean table collected above. We then used the 5 types of Boolean gates illustrated in Figure 1 to generate the logicome associated to the EGFR signaling pathway.



(a) (b) (c) (d) (e)

Figure 1: The Boolean gates for the logical outcome: (a) AND : AB, (b) OR : $A + B$, (c) NOT : $\overline{A}$, (d) NAND: $\overline{AB}$, (e) NOR : $\overline{A + B}$, where we denote the negation of A with $\overline{A}$, the disjunction of A and B with $A + B$, and the conjunction of A and B with AB.

4 Results

The interface species are denoted in the logicome as the nodes of the Boolean network in the way explained in Table 1. The Boolean functions generated as the result of the steps described in Section 3 are shown in Table 2. We repeated the same experiment where we set the initial values of the active key nodes to 10% (rather than 1%) of their maximum value in the simulation of the basic model; the corresponding Boolean formulation is presented in Table 3.

Table 1: The notation used for the interface species in the Boolean network.

Node	Interface species
G_0	(EGF-EGFR*)2-GAP
G_1	Raf*
G_2	MEK-PP
G_3	Ras-GTP*
G_4	ERK-PP
G_5	(EGF-EGFR*)2-GAP-Shc*-Grb2-Sos
G_6	Ras-GTP
G_7	(EGF-EGFR*)2-GAP-Grb2-Sos

Table 2 shows G_1 as getting activated in all knock-out models and thus, being set to constant 1. This means that for all combinations of active/inactive key nodes (even those where G_1 is initialized as inactive), G_1 gets eventually activated in the time interval $[0, 6000]$ sec. This can be interpreted as G_1 being insensitive to (relatively) small changes in the levels of the other key nodes; indeed, all the key nodes are 0 in the basic model, leading to activation of G_1; setting the initial values of the key nodes to 1% of their maximum level in the basic model does not change the situation. This result also suggests that in the case of small perturbations in the initial values of key nodes, the activation of G_1 is driven by other factors, outside the set of key nodes. The situation is different if we look into bigger changes in the initial values of the key nodes, e.g., setting them to 10% of their maximum values in the basic model; as shown in Table 3, G_1 is in this case non-constant and influencing the behavior of G_6.

Table 2: The Boolean functions describing the logicome of the EGFR signaling pathway for the threshold of 1%. An overline over a variable's name denotes its negation, the plus denotes disjunction, while the concatenation of two variables denotes their conjunction.

Boolean functions
$G_0 := \overline{G}_3 + G_5 + G_0\overline{G}_4 + \overline{G}_4 G_7 + G_0\overline{G}_6 G_7;$
$G_1 := 1;$
$G_2 := G_2 + \overline{G}_3 + G_5 + G_6;$
$G_3 := G_0 + \overline{G}_2 + G_3 + G_4 + G_5 + G_6 + G_7;$
$G_4 := G_2 + \overline{G}_3 + G_4 + G_6 + G_0 G_5 G_7;$
$G_5 := G_0 G_5 + \overline{G}_3 G_5 + \overline{G}_3\overline{G}_6 + G_5\overline{G}_6 + G_5 G_7 + G_0\overline{G}_3 G_7;$
$G_6 := \overline{G}_3 + G_5 + G_0 G_6 + G_6 G_7;$
$G_7 := \overline{G}_3 + G_5.$

Another interesting observation of the logicome in Table 2 is that all key nodes get activated in the case of G_3 starts inactive and G_5 starts active. The same observation is found in the results obtained for the threshold of 10%, see Table 3, and even for 20% and 30% see Tables 4 and 5. This is consistent with the observation of [7, 10, 23, 30] about the role played by the shc*-dependent component (denoted by G_5) and the Ras subfamily protein (denoted by G_3) in the activation of several pathway components, including all of our key nodes.

It is also interesting to note that the EGFR signaling pathway has an internal mechanism for compensating the potential failure of G_5 by G_7. Based on [7, 10, 30], G_0 mediates the activation of both G_5 and G_7; in case G_5 fails while G_3 remains inactive then G_7 gets activated and this is enough to activate all key nodes. This is seen in Table 3, if $G_0 = \overline{G}_3 = \overline{G}_5 = G_7 = 1$, then all key nodes get

Table 3: The Boolean functions describing the logicome of the EGFR signaling pathway for the threshold of 10%.

Boolean functions
$G_0 := G_5 + G_0\overline{G}_3\overline{G}_4 + \overline{G}_3\overline{G}_4\overline{G}_6 + G_0\overline{G}_3G_7 + \overline{G}_3\overline{G}_4G_7;$ $G_1 := \overline{G}_3 + G_5 + G_6;$ $G_2 := G_2 + \overline{G}_3 + G_5 + G_6;$ $G_3 := G_0 + \overline{G}_2 + G_3 + G_5 + G_6 + G_7;$ $G_4 := G_2 + \overline{G}_3 + G_4 + G_6 + G_0G_5G_7;$ $G_5 := G_0G_5 + \overline{G}_3G_5 + \overline{G}_3\overline{G}_6 + G_5\overline{G}_6 + G_5G_7 + G_0\overline{G}_3G_7;$ $G_6 := G_5 + G_0\overline{G}_3 + \overline{G}_1\overline{G}_3 + G_0G_6 + \overline{G}_3G_6 + \overline{G}_3G_7 + G_6G_7;$ $G_7 := \overline{G}_3G_5 + \overline{G}_3\overline{G_6} + \overline{G}_3G_7 + G_0G_5\overline{G}_6 + G_0G_5G_7 + G_5\overline{G}_6G_7.$

activated.

4.1 Sensitivity to the numerical setup of the model

To investigate the sensitivity of our method to changes in the numerical setups of the underlying ODE model, we re-ran all simulations for different values of EGF and EGFR. We first experimented with different concentrations of EGF stimulus keeping the same EGFR concentration of 50000 molecules and then with different concentrations of EGFR keeping the same EGF stimulus of 4981 molecules. We observe that the obtained logicomes are almost identical to the previous result presented in Table 2.

To investigate the sensitivity of our method to different threshold criteria, we repeated the experiments above with a threshold value of 30% of each interface species' maximum value. By comparing results, we note that the logicome results obtained with the threshold value of 10%, 20%, and 30% (see Tables 3, 4, and 5) are much more complex than the previous one.

4.2 Incomplete availability of the knock-out mutants

In the way we described our method in Sections 2 and 3, we implicitly assume the full availability of the simulation results of all knock-out mutant models. We considered the case when the data on several knock-out mutants is in fact not available and compared the results to the case when all data is available. We considered the simulations results of only 186 knock-out mutants and assumed that the data on the other 70 knock-out mutants is unavailable. We used the threshold value of 1% and the numerical setups of EGF and EGFR as 4981 and 50000 molecules, respectively.

Table 4: The Boolean functions describing the logicome of the EGFR signaling pathway for the threshold of 20%.

Boolean functions
$G_0 := G_5 + G_0\overline{G}_3\overline{G}_4 + \overline{G}_3\overline{G}_4\overline{G}_6 + G_0\overline{G}_3G_7 + \overline{G}_3\overline{G}_4G_7;$ $G1 := \overline{G}_3 + G_5 + G_6;$ $G_2 := G_2 + \overline{G}_3 + G_5 + G_6;$ $G_3 := G_0 + \overline{G}_2 + G_3 + G_5 + G_6 + G_7;$ $G_4 := G_2 + \overline{G}_3 + G_4 + G_6 + G_0G_5G_7;$ $G_5 := G_0G_5 + \overline{G}_3G_5 + \overline{G}_3\overline{G}_6 + G_5\overline{G}_6 + G_5G_7 + G_0\overline{G}_3G_7;$ $G_6 := G_5 + G_0\overline{G}_3 + \overline{G}_1\overline{G}_3 + G_0G_6 + \overline{G}_3G_6 + \overline{G}_3G_7 + G_6G_7;$ $G_7 := \overline{G}_3G_5 + \overline{G}_3\overline{G}_6 + \overline{G}_3G_7 + G_0G_5G_7 + G_5\overline{G}_6G_7.$

Table 5: The Boolean functions describing the logicome of the EGFR signaling pathway for the threshold of 30%.

Boolean functions
$G_0 := G_5 + G_0\overline{G}_3\overline{G}_4 + \overline{G}_3\overline{G}_4\overline{G}_6 + G_0\overline{G}_3G_7 + \overline{G}_3\overline{G}_4G_7;$ $G_1 := \overline{G}_3 + G_5 + G_6;$ $G_2 := G_2 + \overline{G}_3 + G_5 + G_6;$ $G_3 := G_0 + G_3 + G_5 + G_6 + G_7 + \overline{G}_1\overline{G}_2 + \overline{G}_2G_4;$ $\mathbf{G}_4 := G_2 + \overline{G}_3 + G_4 + G_6 + G_0G_5G_7;$ $G_5 := G_0G_5 + \overline{G}_3G_5 + \overline{G}_3\overline{G}_6 + G_5\overline{G}_6 + G_5G_7 + G_0\overline{G}_3G_7;$ $G_6 := G_5 + G_0\overline{G}_3 + \overline{G}_1\overline{G}_3 + G_0G_6 + \overline{G}_3G_6 + \overline{G}_3G_7 + G_6G_7;$ $G_7 := \overline{G}_3G_5 + \overline{G}_3\overline{G}_6 + \overline{G}_3G_7 + G_0G_5G_7 + G_5\overline{G}_6G_7.$

The result obtained in this case is shown in the Table 6 and it is almost the same as the result in Table 2 obtained by using the full data. This shows that in this case the logicome extraction method was robust to the missing data; this may of course be different for other models and for other missing data.

5 Discussion

We propose in this article an addition to the rich field of logic modeling of biological networks, see, e.g., [15], [19] and [4]. We start from a mathematical model of the network, taking advantage of the growing availability of mathematical models. The logicome approach proposed in this article allows the modeler to focus on a selected set of key nodes, important for the network under study, while abstracting

Table 6: The Boolean functions associated with the logicome of the model where the data of 70 knockout mutants are not available. The result is almost identical to that in Table 2 where all data was available, showing that the method in this case was robust to missing data.

Boolean functions
$G_0 := \overline{G}_3 + G_5 + G_0\overline{G}_4 + \overline{G}_4 G_7 + G_0\overline{G}_6 G_7;$ $G_1 := 1;$ $G_2 := G_2 + \overline{G}_3 + G_5 + G_6;$ $G_3 := \overline{G}_2 + G_3 + G_4 + G_5 + G_6 + G_7;$ $G_4 := G_2 + \overline{G}_3 + G_4 + G_6 + G_0 G_5 G_7;$ $G_5 := G_0 G_5 + \overline{G}_3 G_5 + \overline{G}_3\overline{G}_6 + G_5\overline{G}_6 + G_5 G_7 + G_0\overline{G}_3 G_7;$ $G_6 := \overline{G}_3 + G_5 + G_0 G_6 + G_6 G_7;$ $G_7 := \overline{G}_3 + G_5.$

away from the rest of the network; the output is a description of their influence on each other (even transient) activation over a fixed time interval.

The bottom-up modeling approaches (e.g., large-scale modeling [17], automatic knowledge extraction [22], data-driven network construction [12], etc.) have been very popular due to their ability to provide a very detailed picture, to explain the data, and to reproduce the behaviour of the phenomenon under study. The logicome is a companion to such detailed models; it gives a more abstract, systematic and objective description of the functionalities of the model. This is especially relevant in the case of big models built from many different sub-models and for which a full global "blueprint" does not exist. The logicome aims to be such a blueprint, deduced a-posteriori, based on an existing detailed view of the model.

The output of the logicome approach depends on the numerical setup of the method: both on the numerical setup of the basic mathematical model, and on the choice of the threshold values in the discretization step. This is natural since the method is dependent on the numerical ODE-based simulations of the basic model and of the knock-out mutants; this suggests choosing an already well-fitted and -validated model for the network under study. The choice of the threshold value is in fact a decision on how a species of the model can be labeled as 'active'; we suggested using a percentage of the maximum value reached by that species in the simulation of the basic model, but other choices may also be appropriate depending on the case study.

The computational efficiency of the method is dependent on the number of key nodes selected in the analysis: with more key nodes selected, exponentially more knock-out mutant models should be analyzed. Eliminating some of the knock-out mutants is possible, and the result of the method will be in this case an only-partial

description of the logical dependencies between the key nodes. On the other hand, the method scales up very well in the size of the basic model: as long as the ODE-based models may be simulated efficiently, the method will be practical; this means that networks with thousands of nodes may be analyzed, as long as the number of key nodes n is so that it remains practical to run 2^n simulations.

References

[1] T. Akutsu, S. Miyano, and S. Kuhara. Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. In Roland Somogyi and Hiroaki Kitano, editors, *Pacific Symposium on Biocomputing*, volume 4, pages 17–28. Citeseer, 1999.

[2] D.J. Britton, I. R. Hutcheson, J. M. Knowlden, D. Barrow, M. Giles, R. A. McClelland, J. M. Gee, and R. I. Nicholson. Bidirectional cross talk between ERα and EGFR signalling pathways regulates tamoxifen-resistant growth. *Breast Cancer Research and Treatment*, 96(2):131–146, 2006.

[3] F. J. Bruggeman, H. V. Westerhoff, J. B. Hoek, and B. N. Kholodenko. Modular response analysis of cellular regulatory networks. *Journal of Theoretical Biology*, 218(4):507–520, 2002.

[4] M. Chaves, E. D. Sontag, and R. Albert. Methods of robustness analysis for Boolean models of gene control networks. *IEE Proceedings-Systems Biology*, 153(4):154–167, 2006.

[5] M. I. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PloS one*, 3(2):e1672, 2008.

[6] L. Glass and S. A. Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *Journal of theoretical Biology*, 39(1):103–129, 1973.

[7] Y. Gong and X. Zhao. Shc-dependent pathway is redundant but dominant in mapk cascade activation by egf receptors: A modeling inference. *FEBS Letters*, 554(3):467–472, 2003.

[8] D.E. Gratie, B. Iancu, and I. Petre. ODE analysis of biological systems. In Marco Bernardo, Erik de Vink, Alessandra Di Pierro, and Herbert Wiklicky, editors, *Formal Methods for Dynamical Systems*, volume 7938 of *Lecture Notes in Computer Science*, pages 29—62. Springer, 2013.

[9] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, Pe. Mendes, and U. Kummer. COPASI – a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[10] J. J. Hornberg, B. Binder, F. J. Bruggeman, B. Schoeberl, R. Heinrich, and H. V. Westerhoff. Control of MAPK signalling: from complexity to what really matters. *Oncogene*, 24(36):5533–42, 2005.

[11] H. R. Hwa. A method for generating prime implicants of a Boolean expression. *IEEE Transactions on Computers*, 23(6):637–641, 1974.

[12] K. A. Janes and M. B. Yaffe. Data-driven modelling of signal-transduction networks. *Nature reviews Molecular cell biology*, 7(11):820–828, 2006.

[13] S. Kauffman. Homeostasis and differentiation in random genetic control networks. *Nature*, 224:177–178, 1969.

[14] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems Biology in Practice: Concepts, Implementation and Application*. John Wiley & Sons, 2008.

[15] N. Le Novere. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics*, 16(3):146–158, 2015.

[16] S. Liang, S. Fuhrman, and R. Somogyi. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. In Barbara Bryant, Aleksandar Milosavljevic, and Roland Somogyi, editors, *Pacific Symposium on Biocomputing*, volume 3, pages 18–29. Citeseer, 1998.

[17] D. N. Macklin, N. A. Ruggero, and M. W. Covert. The future of whole-cell modeling. *Current opinion in biotechnology*, 28:111–115, 2014.

[18] S. Martin, Z. Zhang, A. Martino, and J. Faulon. Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics*, 23(7):866–874, 2007.

[19] M. K. Morris, J. Saez-Rodriguez, P. K. Sorger, and D. A. Lauffenburger. Logic-based models for the analysis of cell signalling networks. *Biochemistry*, 49(15):3216—3224, 2010.

[20] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, 1(1), 2005.

[21] P. Pantel and M. Pennacchiotti. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In Marine Carpuat and Kevin Duh, editors, *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 113–120. Association for Computational Linguistics, 2006.

[22] E. Pitkänen, P. Jouhten, J. Hou, M. F. Syed, P. Blomberg, J. Kludas, M. Oja, L. Holm, M. Penttilä, J. Rousu, and M. Arvas. Comparative genome-scale reconstruction of gapless metabolic networks for present and ancestral species. *PLoS Computational Biology*, 10(2):1–12, 2014.

[23] K. Rajalingam, R. Schreck, U. R. Rapp, and Š. Albert. Ras oncogenes and their downstream targets. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1773(8):1177–1195, 2007.

[24] S. Rajasekharan and T. Raman. Ras and ras mutations in cancer. *Central European Journal of Biology*, 8(7):609–624, 2013.

[25] J. Saez-Rodriguez, L. G. Alexopoulos, J. Epperlein, R. Samaga, D. A. Lauffenburger, S. Klamt, and P. K. Sorger. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Molecular Systems Biology*, 5(1), 2009.

[26] J. Saez-Rodriguez, L. G. Alexopoulos, M. Zhang, M. K. Morris, D. A. Lauffenburger, and P. K. Sorger. Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer research*, 71(16):5400–5411, 2011.

[27] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnology*, 20(118):370 – 375, 2002.

[28] S. Sebastian, J. Settleman, S. J. Reshkin, A. Azzariti, A. Bellizzi, and A. Paradiso. The complexity of targeting EGFR signalling in cancer: from expression to turnover. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1766(1):120–139, 2006.

[29] C. Stötzel, S. Röblitz, and H Siebert. Complementing ode-based system analysis using boolean networks derived from an euler-like transformation. *PLoS ONE*, 10(10):e0140954, 2015.

[30] D. Y. Wang, L. Cardelli, A. Phillips, N. Piterman, and J. Fisher. Computational modeling of the EGFR network elucidates control mechanisms regulating signal dynamics. *BMC Systems Biology*, 3(1):1–18, 2009.

[31] Y. Yarden. The EGFR family and its ligands in human cancer: signalling mechanisms and therapeutic opportunities. *European Journal of Cancer*, 37(4):3–8, 2001.

Joukahaisenkatu 3-5 A, 20520 TURKU, Finland | www.tucs.fi

University of Turku
Faculty of Mathematics and Natural Sciences
- Department of Information Technology
- Department of Mathematics and Statistics

Turku School of Economics
- Institute of Information Systems Sciences

Åbo Akademi University
- Computer Science
- Computer Engineering

ISBN 978-952-12-3250-3
ISSN 1239-1891