

Predicting Phenotype from Genotype through Automatically Composed Petri Nets

Mary Ann Blätke¹, Monika Heiner², and Wolfgang Marwan¹

¹ Magdeburg Centre for Systems Biology and Lehrstuhl für Regulationsbiologie, Otto-von-Guericke-Universität, Magdeburg, Germany

² Chair of Data Structures and Software Dependability, Brandenburg Technical University, Cottbus, Germany

`mary-ann.blaetke@ovgu.de`

Abstract. We describe a modular modelling approach permitting curation, updating, and distributed development of modules through joined community effort overcoming the problem of keeping a combinatorially exploding number of monolithic models up to date. For this purpose, the effects of genes and their mutated alleles on downstream components are modeled by composable, metadata-containing Petri net models organized in a database with version control, accessible through a web interface (www.biomodelkit.org). Gene modules can be coupled to protein modules through mRNA modules by specific interfaces designed for the automatic, database-assisted composition. Automatically assembled executable models may then consider cell type-specific gene expression patterns and the resulting protein concentrations. Gene modules and allelic interference modules may represent effects of gene mutation and predict their pleiotropic consequences or uncover complex genotype/phenotype relationships. Forward and reverse engineered modules are fully compatible.

Keywords: Biomodel engineering, formal language, data integration, high-throughput, quantitative trait loci.

1 Introduction

Systems biology witnesses the evolution of experimental high-throughput methods with steadily increasing power regarding the quantification of nucleic acids, proteins, covalent modifications, and metabolites. Soon, these methods will broadly allow *omics* scale analyses of molecules involved in cellular regulatory control that capture the time-resolved response to stimulation or (genetic) network perturbation [22]. These advances challenge the development of integrative and modular modelling frameworks that support the combination of findings obtained through different, qualitative and quantitative experimental approaches. To be useful, such models will need to be multi-level in terms of integrating multiple levels of abstraction in causally connected and experimentally well established processes at the molecular level with adjustable resolution of details for higher level phenomena. These may include cell fate decisions for

simulating the intrinsic heterogeneity of clonal populations of cells following individual trajectories during development. Petri nets are an ideal formalism for the formal description of processes at multiple levels of abstraction for systems biology purposes [13,17,18,28].

For the sake of creating realistic scenarios, it will presumably become indispensable to compare, on a regular basis, simulation results on one and the same network topology as obtained by employing continuous, stochastic, and hybrid paradigms. In the software tool Snoopy, a given Petri net graph can be interpreted and simulated as continuous (ODE), stochastic, hybrid, or simply as qualitative model with export option to SBML [29]. Interpretation as coloured Petri net, again in Snoopy, provides advanced options for biomodel engineering [19] as coloured Petri nets combine the formalism of Petri nets with the expressive power of a programming language. For this reason, and because of the intuitively accessible graphical representation, we have chosen Petri nets as framework for modelling and simulation.

Repeated iterations of experimental data acquisition, modelling, and simulation can evaluate the consistency of the interpretation of experimental results. This is especially true when high-throughput data come into play. However, conventional monolithic models usually represent certain aspects of a phenomenon at a certain resolution in detail and are restricted to a certain mathematical modelling paradigm. Such models can neither be easily combined with other models nor can they be easily updated by persons other than the author of the particular model. One solution to this limitation is to create a collection of Petri net modules that can be automatically linked in order to obtain and to update coherent models covering all or selected aspects of a biological process. Conceptually, such modules may be contributed, curated, and updated by individuals of the community with special expertise in certain aspects. Being organized within a database with version control, modules obtained by reverse engineering of experimental data can also be integrated and, as we will show below, help to import complex data sets into models that have been automatically generated by composition of pre-existing modules.

This paper makes three major contributions which fundamentally enhance the versatility of modular Petri net modelling by (1) linking regulated gene expression to protein concentration, (2) allowing the fully automated generation of models for the application in genome-wide (*omics*) approaches, and (3) linking gene mutation to complex phenotypic consequences in generating predictable models. Let us briefly elaborate on these claims.

1. The introduction of gene modules and mRNA modules allows to model regulated gene expression and protein biosynthesis. As the gene expression pattern of a cell is not constant and can drastically change dependent on cell type, physiological state, or experimental conditions, cells are definitely equipped with specific sets of proteins of variable relative abundance. By introducing gene modules into the model, differentially regulated gene activity and the resulting gene expression patterns translate into the marking of places of the protein modules. As the rates of biochemical reactions always

depend on both, the kinetic rate constants and the concentrations of the reactants, changed gene expression will also change the rates of biochemical reactions, which in turn may drastically alter the dynamic behaviour of a regulatory network. Moreover, changed concentrations in regulatory proteins (e.g. transcription factors) may feed back in a complex manner onto the gene level by changing gene expression profiles. This circuitry of interwoven regulatory control becomes systematically accessible through the model.

2. Gene and mRNA module prototypes permit the fully automated generation of modules by simply uploading lists of gene names. This allows the automatic creation of models representing hundreds or even thousands of genes, their mRNAs and the proteins they form. By importing transcriptomic or proteomic data sets obtained in high throughput experiments [30], one can infer rate constants and reverse engineer regulatory mechanisms with the help of the model and predict changes in the proteome in response to differential gene regulation. Such models will also support the interpretation of phenomena observed in systematic RNAi screens where individual genes are knocked down [25,10,9].
3. Introduction of allelic influence modules extends the modelling of gene activity to the modelling of the regulatory consequences, which gene mutations have on cellular processes. This sets the formal framework to reverse engineer biomodels from complex phenotype data sets resulting from genotypic variation e.g. by employing Petri net compatible algorithms [16,12,11]. It is obvious that such type of models have a high potential for the application to various areas from basic research to synthetic biology or personalized medicine.

We are not aware of any modelling framework providing a comparable versatility and integrative power in terms of combining forward and reverse biomodel engineering.

This paper is organised as follows. In the next section we briefly summarise relevant own previous work on modular Petri net modelling and the automatic composition of modules with the help of a module database. Then we will provide the rules for the generation of modules and modular models and introduce gene and mRNA modules in Section 3, the application of which is demonstrated by the first case study in Section 4. Section 5 explains the specific features of gene and mRNA modules required for modelling gene expression and its differential regulation in eukaryotes. We continue in Section 6 with a second case study on cell differentiation and eukaryotic gene regulation, define allelic influence modules, and explain how these work together with gene, mRNA, and protein modules in generating models that integrate forward and reverse biomodel engineering approaches. In Section 7 we conclude with discussing the versatility of our approach and provide future perspectives regarding the application to synthetic biology and *omics* approaches.

2 Previous Work

Initially, we developed our modular modelling approach to represent biochemical reaction networks made of protein-protein interactions. The core idea is to take an object-oriented approach where the objects correspond to the natural modular building blocks of life. We represent individual proteins as independent and self-contained hierarchically structured Petri nets, called modules. Thus, modules correspond to natural units, each of which comprises intramolecular regulatory mechanisms of the respective protein and of all its interactions with other molecules. Modules of interacting proteins can be coupled by logical nodes of identical shared subnets describing their interaction with each other. The assembly of models from a set of modules needs no further modifications at the module level. An essential advantage of this approach is the reusability of all constructed modules in arbitrary combination to obtain models representing specific pathways [4,5].

A crucial point of our modular modelling approach is that each module is self-contained and can be evaluated on its own. Composed networks which in terms of their behaviour correspond to the conventional, monolithic networks need never to be explicitly shown. Their validity is entirely assessable by understanding the individual modules and the connection rules.

Our modular modelling approach has been deployed to construct modular Petri nets for two non-trivial case studies: (1) the JAK/STAT pathway in IL-6 signalling [4] and (2) nociceptive network in pain signalling [5]. The JAK/STAT pathway is one of the major signalling pathways in multicellular organisms controlling cell development, growth and homeostasis by regulating the gene expression. The modular network of the JAK/STAT pathway in IL-6 signalling comprises 7 protein modules (IL6, IL6-R, gp130, JAK1, STAT3, SOCS3, and SHP2). Overall, the model consists of 92 places, 102 transitions spread over 58 pages with a nesting depth of 4. The nociceptive network in pain signalling consists of several crucial signalling pathways, which are hitherto not completely revealed and understood. The latest version of the nociceptive network consists of 38 modules, among them are several membrane receptors, kinases, phosphatases and ion-channels. So far, the model is made up by 713 places and 775 transitions spread over 325 pages, again with a nesting depth of 4.

To support convenient module selection and network composition, we developed a database prototype, which is accessible by a web-interface [4,6], see Figure 1. The database holds the qualitative Petri net structure of each module, as well as the kinetic information assigned to each transition. In addition, the database contains also meta-information about the corresponding proteins (extracted from UniProt) and information about the literature used to construct the modules (extracted from PubMed), which can be associated with each module. The organization of the modules in such a database enables the user to (1) search for individual modules, places, transitions, etc., (2) store modules in collections, and (3) assemble a modular network from a chosen collection.

(a)

ID	Gene Symbol	Protein Name	Accession	Approved	Release Date
1	IL6	Interleukin-6	P05231	★	2011-11-08
2	IL6R	Interleukin-6 receptor subunit alpha	P08887	★	2011-11-08
3	IL6ST	Interleukin-6 receptor subunit beta	P40189	★	2011-11-08
4	JAK1	Tyrosine-protein kinase	P23458	★	2011-11-08
5	JAK1	Tyrosine-protein kinase	P23458	★	2011-11-10
6	STAT3	Signal transducer and activator of transcription 3	P40763	★	2011-11-08
7	PTPN11	Tyrosine-protein phosphatase non-receptor type 11	Q08124	★	2011-11-08
8	PTPN11	Tyrosine-protein phosphatase non-receptor type 11	Q08124	★	2011-11-10
9	SOCS3	Suppressor of cytokine signaling 3	O14543	★	2011-11-08

(b)

MODULE: IL6★ (ID:1 , Release Date: 2011-11-08)

General Information

Submitter(s): Eliašik, Mary Ann
 Curator(s): Dittsch, Anna
 Schaper, Fred
 Protein: Interleukin-6 (Gene Symbol: IL6, Accession Number: P05231)
 Connectable Module(s): IL6R, IL6ST, JAK1, JAK1_HYP1

Module Version(s): No other versions in the database.
 Publication(s): Müller-Newton G, Heinrich PC, Behrmann I, Haan S, Hermans HM, Schaper F: Principles of Interleukin (L)-6-type cytokine signaling and its regulation. *Biochem J*. 2003; 374 (Pt 1): 1-20 (PMID: 12773955)
 Müller-Newton G: The cytokine receptor gp130: faithfully promiscuous. *Sci STKE*. 2003; 2003 (201): PE40 (PMID: 14500288)

Place Overview

Place Name	Description
IL6R_CBM	cytokine binding module (CBM) of Interleukin-6 receptor subunit alpha (IL6R)
IL6R_CBM_IL6_site1	cytokine binding module (CBM) of Interleukin-6 receptor subunit alpha (IL6R) bound to binding site 1 (site1) of Interleukin-6 (IL6)
IL6R_site1b	site 1b (site1b) of Interleukin-6 receptor subunit beta (IL6R)
IL6R_site1b	site 1b (site1b) of Interleukin-6 receptor subunit beta (IL6R)
IL6ST_CBM	cytokine binding module (CBM) of Interleukin-6 receptor subunit beta (IL6ST)
IL6ST_CBM_IL6_site1a_IL6R_site1b	cytokine binding module (CBM) of Interleukin-6 receptor subunit beta (IL6ST) bound to binding site 1a (site1a) of Interleukin-6 (IL6) and binding site 1b (site1b) of Interleukin-6 receptor subunit alpha (IL6R)
IL6ST_site1a	immunoglobulin-like domain (site1a) of Interleukin-6 receptor subunit beta (IL6ST)
IL6ST_site1a_IL6_site1a_IL6R_site1b	immunoglobulin-like domain (site1a) of Interleukin-6 receptor subunit beta (IL6ST) bound to binding site 1a of Interleukin-6 (IL6) and binding site 1b (site1b) of Interleukin-6 receptor subunit alpha (IL6R)
IL6_site1	binding site 1 (site1) of Interleukin-6 (IL6)
IL6_site1a	binding site 1a (site1a) of Interleukin-6 (IL6)
IL6_site1a	binding site 1a (site1a) of Interleukin-6 (IL6)
JAK1_FERM_IL6ST_Box2	4-1 acetyl radixin and moesin domain of Tyrosine-protein kinase JAK1 (JAK1) bound to Box1 and Box2 of Interleukin-6 receptor subunit beta (IL6ST)

Fig. 1. Selected screenshots of the prototype database [6]. (a) The web-interface enables the user to browse and/or search for modules of specific proteins. Modules can be stored in collections by selecting them. (b) Detailed information about each module can be shown on a separate page.

3 Petri Net Modules

3.1 How Modules Are Built and Composed

We use Petri nets structured in the form of modules that allow the automatic composition of executable models [4,5] based on Snoopy [29]. The modules and their associated metadata are organized in a database accessible through a web interface designed to manage multiple versions of each module and supporting the automatic composition of models from modules (interactively) selected from a potentially rich collection [4]. Initially, the approach was designed to model protein-protein interactions (see Section 2) in the context of signal transduction networks [4,5]. In this paper we go one step further by defining gene modules, mRNA modules and allelic influence modules which considerably enhance power and versatility of our modelling approach.

A module in general is centred around one entity describing all its interactions with other components to which the module can be linked. The entity can be

a protein, an mRNA, a gene, or the specific allele of a gene. Conceptually, it is possible to extend this definition to admit other entities as well. Before going into detail, let us first briefly explain how the modular modelling approach technically works and how modules are connected to each other.

The basic principle is explained taking the reaction of an autophosphorylating kinase with its substrate as example (Figure 2). The kinase X autophosphorylates to give XP. XP transfers its phosphate group to the substrate Y. The phosphate group of YP then hydrolyses spontaneously. The Petri net describing these reactions (Figure 2a) can be decomposed into two modules each representing the reactions of one of the two proteins involved, X and Y, respectively (Figure 2b, 2c). Places and transitions that occur in more than one module are defined as so-called logical nodes. Logical nodes appear as multiple graphical copies of a given place or transition. In this paper, logical nodes are shaded in grey. Braking a Petri net up into modules introduces redundancy in terms of the graphical display of nodes which might appear unnecessarily complicated at first. For complex modules or when many modules are involved, the benefits are indeed tremendous as we have previously shown ([4,5], and see Discussion). The biosynthesis and the degradation of a protein or a nucleic acid are modelled by separate biosynthesis or degradation modules, respectively. Accordingly, the user can choose for each protein or mRNA whether or not its biosynthesis and degradation should be considered in the assembled model.

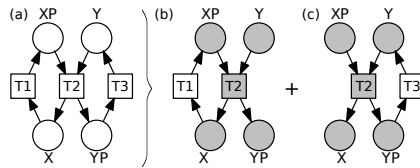


Fig. 2. The principle of modular modelling based on the use of logical nodes (connectors). (a) A Petri net model of the phosphorylation of protein Y by the autophosphorylating kinase X is split into modules for (b) protein X and (c) protein Y, respectively. Places and transitions that are shared by the two modules are implemented as logical nodes shaded in grey.

3.2 Definition of Modules Representing the Function of Genes

We now extend the initial approach constrained to protein modules by defining gene modules. A gene module considers the mechanisms of a gene being regulated through the reversible transition between its on and off state. Assigned metadata information includes the *Genbank* database accession number providing the DNA sequence information as a cross-reference. Logical nodes of a gene module are used to link the gene to other modules.

In general, multiple forms of each gene, so-called alleles, do exist that differ in one or more base pairs. These differences are due to mutations. Mutations can be silent in not altering the amino acid sequence of the encoded protein. Mutations can also be neutral in not changing the properties of the encoded protein

although its amino acid sequence is changed due to mutation. These sequence polymorphisms are commonly found in wild-type populations. Alternatively, a mutation can change the properties of the encoded protein due to its altered amino acid sequence or it may even prevent the formation of the protein at all, e.g. by introducing a stop codon. Following strictly the modularity principle in designing Petri nets, a separate gene module is created for each allele of a gene to represent mutations that change relevant properties of the gene products (RNAs and proteins) as compared to the wild-type. Entering a query for a gene, the module database will list all modules corresponding to alleles of that gene.

We will now use two case studies to demonstrate how gene modules work. The first case study (see Section 4) concerns metabolic regulation in bacteria. It is taken to explain the principle of gene modules with the help of a simple example. In Section 5, we explain the specific features of gene and mRNA modules required for modelling gene expression and its differential regulation in eukaryotes, which is the topic of our second case study (see Section 6). Here, we show a more complex scenario involving multiple layers of regulatory control. In addition, the second case study will provide an example of how modules can be obtained through reverse engineering of experimental data, it will introduce allelic influence modules and it will reveal the scalability of the approach.

4 Case Study: The Phosphate Regulatory Network

In the first case study, we consider the response of enteric bacteria like *Escherichia coli* to the limitation in inorganic phosphate which is required for the biosynthesis of nucleic acids and other cellular components (compare Figure 2).

When inorganic phosphate (P_i) becomes low in the environment, it may turn into a growth-limiting factor even if sufficient nutrients are available [26]. When present, inorganic phosphate is taken up from outside of the cell through an ABC transporter system, the PstSCAB transmembrane complex (Figure 3 [21]). With sufficient P_i outside, the PstSCAB complex actively pumps P_i across the cell membrane into the cytoplasm. Under this condition, the PhoU protein forms, according to the proposed mechanistic model [21], a complex with the pstSCAB transporter system and the PhoR histidine kinase. Complex formation prevents the kinase to autophosphorylate caused by the interaction with PhoU. PhoU is a chaperone-like PhoR/PhoB inhibitory protein. When external P_i is low and the PstSCAB complex is inactive, PhoU dissociates and allows the autophosphorylation of PhoR. PhoR then phosphorylates and thereby activates the transcription factor PhoB. The phosphorylated form of PhoB, namely PhoBP, then activates the transcription of at least 31 genes organised into 9 transcriptional units (*eda*, *phnCDEFGHIJKLMNOP*, *phoA*, *phoBR*, *phoE*, *phoH*, *psiE*, *pstSCAB-phoU*, and *ugpBAECQ*) [21]. One of the activated genes, *phoA*, encodes the PhoA protein which is a bacterial alkaline phosphatase. PhoA is exported across the membrane into the periplasm where it degrades organic phosphorous compounds to liberate P_i which is then taken up into the cell to overcome the limitation.

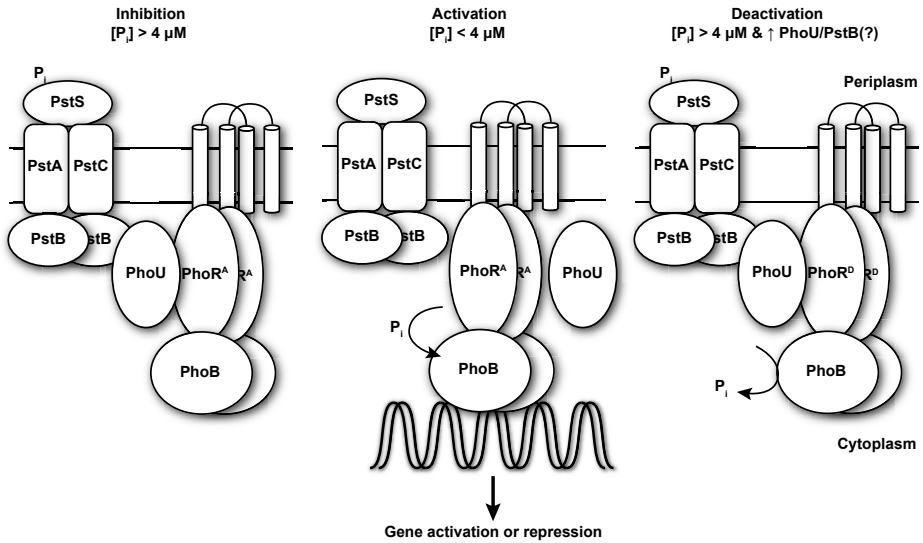


Fig. 3. Biochemical model for sensing extracellular inorganic phosphate (P_i) and transduction of the signal to control gene expression. The PstSCAB transmembrane complex serves as an ABC transporter for the uptake of environmental P_i . At high extracellular P_i concentration, the binding protein PstS is fully saturated and this signal is relayed to the cytoplasmic part of the receptor that forms an inhibitory complex with a second transmembrane protein, the PhoR kinase via the cytoplasmic protein PhoU. If P_i is low, the complex dissociates and the autophosphorylating kinase PhoR phosphorylates PhoB which, in its phosphorylated form, binds DNA to induce gene expression. When P_i subsequently increases, the complex with PhoU is formed again and PhoBP is dephosphorylated. The figure was redrawn from Hsieh and Wanner published in Current Opinion in Microbiology [21].

When enough P_i has been formed, this system is switched off again and PhoBP is dephosphorylated.

In [24], we gave a monolithic Petri net of a simplified version of the phosphate regulatory circuitry. To cover the entire functionality we now construct a modular Petri net model which is composed of three types of modules: (1) one protein module representing the reactions of the PhoB protein, (2) gene modules representing the regulated genes, and (3) mRNA modules representing the transcription of the gene, the translation of mRNA into the protein, and the degradation of mRNA. The degradation of the encoded proteins is represented by degradation modules which have been introduced previously [4].

4.1 The PhoB Module

The PhoB module (Figure 4) represents the reactions of the PhoB protein in its phosphorylated (PhoBP) and dephosphorylated (PhoB) states. It also represents the complex formation of PhoB with its regulatory proteins as well as the binding of PhoB to regulatory sequences in the DNA.

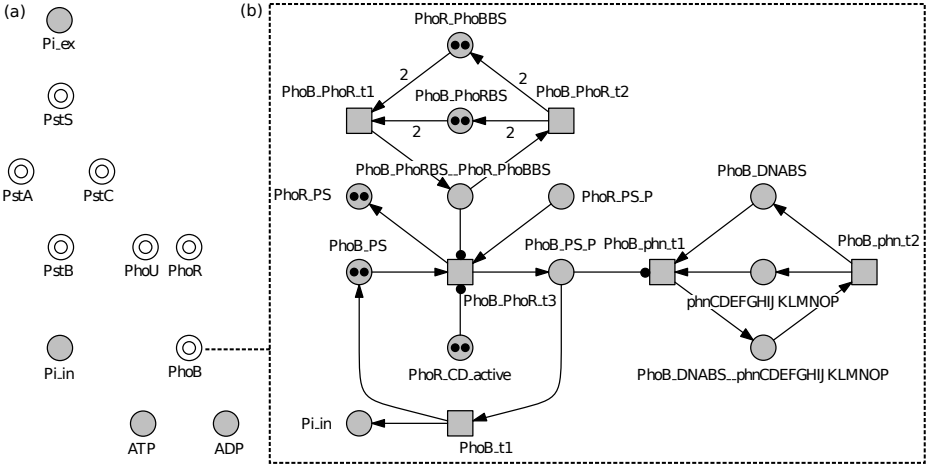


Fig. 4. Petri net representation of the phosphate regulatory network. (a) Top-level presentation of the modules of the phosphate regulatory system in the form of coarse places as they appear in Snoopy. (b) Protein module of the PhoB protein displaying the direct interactions with binding partners.(BS, binding site).

The PhoB module models the regulatory mechanism schematically shown in Figure 3. Binding and dissociation of PhoBP in its complex with PhoRP is represented separately for each transcriptional unit. Displaying the binding of PhoBP to each regulatory site on the DNA separately, keeps the Petri net graph clearly arranged and allows to reuse the structural motif of binding and unbinding reactions via copy/paste for the various regulated transcriptional units. Accordingly, PhoRP_PhoBP is declared as logical place.

To save space in Figure 4, we only show binding of PhoBP to one of the nine transcriptional units.

4.2 The Gene Modules

A gene module represents the regulation of the gene by other factors (e.g. transcription factors) through the transition between its on and off state (Figure 5). In the quantitative interpretation of the Petri net (as stochastic, continuous, or hybrid Petri net) the regulatory factors influence the equilibrium between the on and the off state of the gene. The on state means that the gene is transcriptionally active and that mRNA molecules can be accordingly formed as modeled in the mRNA module.

In prokaryotes, several functionally related genes can be organized into a single regulatory and transcriptional unit, a so-called operon. The genes of an operon are transcribed in the form of a single mRNA molecule, called a polycistronic message, which may encode several proteins at once. In the case of the *phnCDEFGHIJKLMNOP* transcriptional unit (Figure 6), one polycistronic message encoding 14 different proteins (PhoC to PhoP) is formed upon the

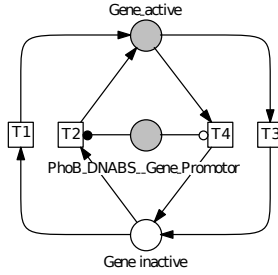


Fig. 5. Prototype of a gene module. The module may represent the activation of any gene responsive to the phosphorylated PhoB protein. Binding of PhoRP_PhoBP to the promotor renders the gene active. These regulatory interactions are modeled through a test arc activating transition T2 and an inhibitory arc blocking T4. The basal activity of the gene in the absence of PhoRP_PhoBP is maintained by T1 and T3.

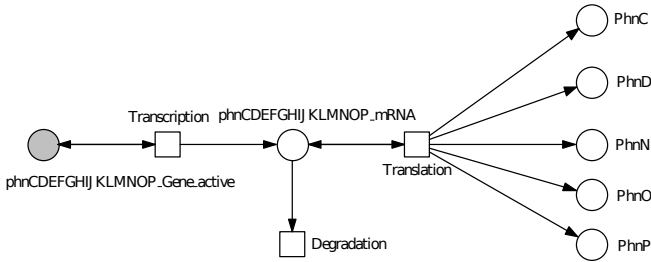


Fig. 6. mRNA module modelling the formation of the polycistronic message synthesized by transcription of the active *phnC-P* gene. The mRNA is translated into the proteins PhnC . . . PhnP. For simplicity, the places for only five of the 14 proteins that are formed are shown.

initiation of transcription. The probability per unit of time for the initiation of transcription to occur depends on binding of PhoRP_PhoBP to the regulatory region of the *phnCDEFGHIJKLMNOP* operon on the DNA. From the biological point of view, polycistronic messages provide a simple mechanism for co-regulation of genes encoding proteins that work together in a cellular process. In order to obtain a model including the transcriptional regulation of all 31 genes organized into the 9 transcriptional units that are under control of the PhoB protein, 9 gene modules (*phoA*, *phoBR*, *phoE* etc.) of analogous structure (as shown in Figure 5) are required. For large scale modelling approaches, these network structures could be generated automatically or semi-automatically.

4.3 The mRNA Modules

The mRNA modules models the reactions of the respective mRNA species, namely its biosynthesis by transcription, its degradation, and its translation into the encoded proteins (Figure 6). An mRNA module may in addition represent the binding of regulatory proteins to the RNA, the binding of antisense

RNA influencing the stability of the message, or the processing (e.g. splicing) of the transcript, as it occurs in eukaryotes (shown in Figure 8 and discussed in Section 7.2). The transcription of the bacterial *phnCDEFGHIJKLMNOP* transcriptional unit leads to the formation of a polycistronic message, which encodes for 14 proteins (PhoC . . . PhoP). The reactions (e.g. the catalytic activity) of the encoded proteins might then be considered in separate protein modules.

5 Modelling Eukaryotic Gene Regulation with Gene and mRNA Modules

5.1 Eukaryotic Gene Modules

The gene modules designed to model the regulation of eukaryotic genes are very similar to the models of prokaryotic genes as presented in the previous section for the phosphate regulatory network. However, the regulation of eukaryotic genes is typically more complex than in prokaryotes as more protein factors and more binding sites for regulatory proteins on the DNA may be involved. All these factors together may control the on state of a gene.

Gene regulation may involve protein binding sites on the DNA functioning as enhancers or silencers that are located several thousand base pairs distant from the genes they regulate. Proteins bound to these sites may influence the probability for transcription to be initiated through physical interactions with the protein complexes bound to the promotor of the regulated gene. These regulatory sites and the binding of regulatory proteins to these sites are represented as part of the gene module.

A prototype of a module representing the regulatory control of a eukaryotic gene is shown in Figure 7. Making regulatory sites part of the gene module comes with the advantage that potentially cooperative effects in protein binding and gene activation can be considered as part of the module.

5.2 Eukaryotic mRNA Modules

In addition to the biosynthesis of proteins there are several RNA-dependent processes that may be of regulatory importance especially in eukaryotic cells. Typically, the occurrence of these mechanisms depends on the considered RNA species and may also depend on physiological conditions as well as on developmental states. Each of the mechanisms described in the following has been implemented in a basic form using the mRNA module prototype shown in Figure 8.

Alternative Splicing. Primary transcripts in eukaryotes are processed during the maturation of the final protein-encoding mRNA. Processing includes the splicing of the RNA where non-coding introns are excised from the primary transcripts. Due to the occurrence of alternative splicing sites, it may be that differently spliced mRNAs are formed from one and the same primary transcript

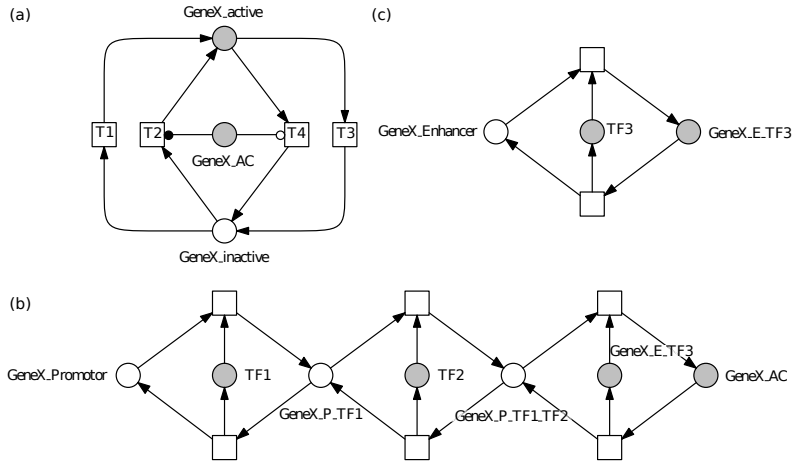


Fig. 7. Prototype of a eukaryotic gene module. The regulation of the eukaryotic gene (a) depends on more protein factors than the regulation of the prokaryotic gene shown in Figure 5 does. By binding to the promoter region of the gene, these factors form a multimeric protein complex, as shown for transcription factors TF1 and TF2, both of which directly bind to the promoter (b). The third transcription factor, TF3, binds first to an enhancer sequence distant from the promoter (c) and subsequently can bind to the promoter-TF1-TF2 complex to form the gene activating complex GeneX_AC (b) which switches the gene into its transcriptionally active state (a). As in Figure 5, gene activation by protein binding to the promoter is modeled by control arcs and the basal level of gene activity occurs through firing of transitions T1 and T3 (a). According to our module notion the gene module displays all direct molecular interactions of GeneX with the proteins that bind to its regulatory sequences. Note that binding of regulatory proteins may involve cooperative mechanisms which would be represented accordingly in the context of the gene module.

giving rise to proteins of partially different amino acid sequence. This splicing depends on protein factors that may be regulated depending on the physiological or developmental state of the cell. When necessary, the biochemical reactions of these slicing factors (like posttranslational modification or protein-protein interaction) may be represented in the context of protein modules with the help of logical places.

RNA-Binding Proteins. The half-life of mRNA species may vary between minutes and months. This variation may have different reasons in addition to the specific secondary structure of the RNA. One mechanism influencing the half-life of a given mRNA species is the binding to specific RNA-binding proteins that may store or degrade the RNA. Being bound to an RNA-binding protein, like Pumilio for example, the mRNAs can be stored in the cell while being translationally inactive. Upon release from the RNA-binding proteins the mRNA may suddenly become translationally active and hence become available at relatively high concentration [14].

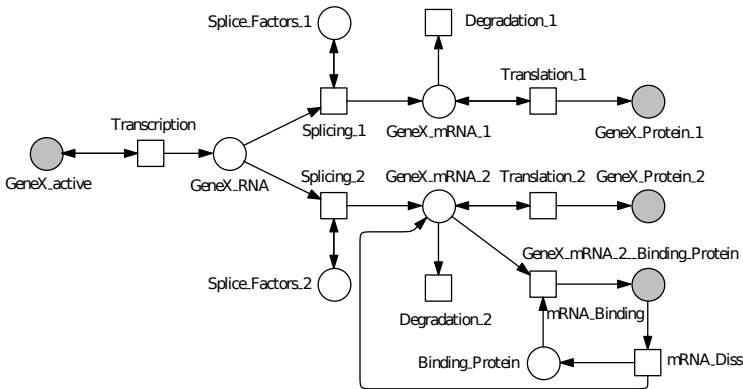


Fig. 8. Prototype of an eukaryotic mRNA module. Transcription of the active Gene (GeneX) leads to the formation of a primary transcript which is processed. In the example shown, the primary transcript is spliced into two alternative mRNAs, GeneX_mRNA_1 and GeneX_mRNA_2, respectively that are subsequently translated into the corresponding proteins. The mature mRNAs may bind to RNA binding proteins that regulate the stability of the mRNA and its availability for the translational machinery. For simplicity, this reaction is shown for one of the two mRNAs only.

RNA Interference. RNA interference is a natural mechanism for the specific inactivation of the expression of eukaryotic genes, e.g. by small interfering RNAs that bind to the target RNA. The degradation of the RNAs depends accordingly on specific protein factors and on the availability of the interfering RNA [8,33].

These mechanisms may have to be considered in the context of mRNA modules and receive regulatory input from specific cellular proteins.

5.3 Why Is the Integration of Bottom-Up and Top-Down Models Essential, Especially in Eukaryotes?

In some respects systems biology appears to be more difficult for eukaryotic as compared to prokaryotic cells. This may in part be due to the occurrence of fundamentally different regulatory processes in the two domains of life with non-obvious consequences of certain eukaryotic regulatory phenomena.

There is indeed highly detailed knowledge on the canonic pathways of eukaryotic signal transduction which allows the formulation of well-structured bottom-up models representing the biochemical interactions of regulatory components like e.g. the MAP kinase cascade. Such models can be tremendously useful in understanding mechanisms in health and disease [23,27,31] and in finding new and powerful drugs.

However, it is also true that many experimental findings on canonic pathway components seem to be contradictory. This may in part be due to the fact that gene expression patterns in different experimental systems and under different physiological conditions are different leading to a different composition of bio-

chemical reactants within the cell at a given time point or between replicas of a particular experiment.

This certainly restricts the current practical value of bottom-up models without disclaiming their general usefulness. When a certain cellular process, for example the differentiation phenomenon in a eukaryotic cell or the progression through the cell cycle, is to be rigorously analysed at the transcriptomic or the proteomic level, the changes in many perhaps in most of the observed components and their consequences can currently not be explained on the basis of the established bottom-up models of the canonic pathways. This suggests that there are tremendous gaps in our current understanding.

On the other hand, it seems for the time being impossible that the thousands of components accessible through *omics* approaches can be analysed with such experimental effort as invested for the exploration of canonic pathway components. Therefore it seems self-evident that *omics* data are used to rigorously reverse-engineer models. Then, a next and essential step is to integrate these top-down models with relevant bottom-up models to obtain integrated models with predictive power.

6 Case Study: The Sporulation Control Network in *Physarum polycephalum*

6.1 Sporulation Is Controlled by a Gene Regulatory Network

Physarum polycephalum is a unicellular eukaryote belonging to the amoebozoa group of organisms [1,2,15]. During its relatively complex and branched life cycle, *Physarum* develops into various cell types. These cell types occur in temporal order and differ in morphology (shape), molecular composition, and physiological function ([7] and references therein).

One of these cell types is the plasmodium, a multinucleate macroscopic single cell. Differentiation of the plasmodium can be easily studied under lab conditions as the response can be experimentally triggered by applying a brief pulse of far-red light. The light pulse sets a defined starting point on the time axis on which subsequent events are observed. During about 18 hours after the trigger, the entire plasmodial cell is extensively remodeled and fruiting bodies are formed that give rise to mononucleate haploid spores that are precursor cells of amoebal gametes which will develop at later stages of the cycle [7]. This process is called sporulation. Please note that sporulation in bacteria, although the name is the same, is biochemically a completely different process than in eukaryotic cells.

Five to six hours after an inductive far-red pulse, the cell is irreversibly committed to sporulation. The associated molecular events are of scientific interest as the plasmodium loses some stem cell-like capabilities during commitment. The expression pattern of hundreds of genes changes [15,7]. These changes in gene expression that normally occur can be compared to the changes that are seen in mutant cells that have lost their ability to be committed to sporulation [3,20].

A widely-used method in the biosciences is the genetic dissection of gene regulatory networks by generating mutants which are altered in the regulatory control and analysing the phenotype which a mutation produces. Mutants may be produced through forward and/or reverse genetic approaches. In forward genetics, randomly mutated cells or organisms are screened for phenotypes of interest and the gene which causes the phenotype is identified subsequently. In the reverse genetic approach, a gene of interest is mutated and the phenotypic consequence of the generated mutation is analysed. Mutation of a gene, both in forward and reverse genetics, may either cause the loss of a protein or a change in its function. Mutation may change the activity of a protein (up or down) or it may change the specificity of its catalytic activity. In many cases, the molecular mechanisms of how a given mutation translates into the observed phenotype remain unknown for a number of years. Despite this ambiguity, the genetic approach is powerful as it rigorously establishes causal dependencies within the living organism. In most cases biochemistry alone could not fulfill this task.

A powerful way to employ genotype/phenotype relationships for modelling and simulation is the reverse engineering of genetic data. Reverse engineering of gene expression data provides a direct link to bottom-up models of protein-protein interactions. By reverse engineering one can establish effects, which a mutated gene (the allele of a gene) exerts on a cellular process. We define allelic influence modules to represent these influences. We will now show how allelic influence modules are built and how they are useful for the integration of top-down and bottom-up model parts into one coherent model.

6.2 Linking Genotype to Phenotype: Allelic Influence Modules

As gene modules, allelic influence modules are centred around the allele of a given gene. However, allelic influence modules differ from gene modules in representing the regulatory influences exerted by the allele on cellular processes by controlling the firing activities of respective transitions through read or inhibitory arcs. In reality, these influences can be rather indirect by involving numerous other, potentially unknown components. Accordingly, the allelic influence module may represent the control of molecular events like the biosynthesis of RNA by transcription or even more complex processes of potentially unknown molecular mechanism as inferred from functional studies. To make this more clear, let us consider the case study.

In response to far-red light, *Physarum* plasmodia differentially express a large number of genes several hours before the cell is irreversibly committed to sporulation [20]. Genes with both up- and down-regulated RNAs have been identified at a genomic scale [15,3], and the precise expression kinetics of some of them have been investigated in detail [20]. Currently, we do neither know the molecular mechanisms through which these genes are controlled nor do we know which causal consequences the change in expression level in detail have. However, the majority of the differentially expressed genes encodes proteins with high sequence similarity to proteins of important regulatory function.

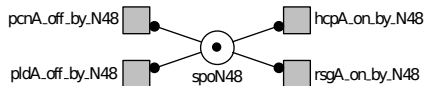


Fig. 9. Allelic influence module. The module represents the differential regulation of four genes, *pcnA*, *pldA* (down regulated), *hcpA*, and *rsgA* (up regulated) by the *spoN48* allele of the *spoN* gene in *Physarum polycephalum*. The logic transitions shown here are also part of the gene modules of the four differentially regulated genes (not shown). Genes and names of their orthologs in the *UniProt* database: *pcnA*, proliferating cell nuclear antigen; *pldA*, Phosphatidylinositol-glycan-specific phospholipase D; *hcpA*, Histone chaperone ASF1A; *rsgA*, Regulator of G-protein signalling 2.

We have genetically dissected the underlying regulatory network with the help of mutants that are altered in the photocontrol of sporulation, as isolated in phenotypic screens ([32]; Rätzel et al., unpublished results). Representatives of one class of these mutants have lost their ability to be committed to sporulation and remain forever in a proliferative state. Although these mutants do not respond to far-red light by sporulation, they clearly respond at the transcriptional level. However the pattern of genes that are differentially expressed in response to the stimulus significantly differs in the mutants as compared to the wild-type and also differs between mutants. The altered gene expression patterns clearly reveal the regulatory influence of the mutated genes and can be used to infer the network of regulatory control in which the different regulators inactivated in each of the mutants are interwoven. The changed pattern of differentially expressed genes can be used to reverse engineer the regulatory influence of the gene mutations. In Figure 9 the allelic influence module obtained for the *spoN48* allele of the *spoN* mutant gene is shown. The proteins encoded by some of the differentially expressed genes have well-known biochemical functions in the developmental control of eukaryotic cells. In a bottom-up approach, these genes can be linked to corresponding protein modules in terms of changed protein concentrations as predicted by the model.

7 Versatility of the Approach and Future Perspectives

We have described a strictly modular approach to Petri net modelling based on clearly defined types of modules corresponding to the different types of molecular entities around which each module is centred: genes, RNAs, and proteins. The small number of module types and the few easy-to-follow rules for creating a module are expected to encourage community efforts in creating a collection of modules in analogy to how Wikipedia collects pages. In Petri net modules, cross references to other modules allow the automatic composition of large models that are directly executable [4]. A web-accessible database was constructed to manage different versions of each module which is available as a prototype [4]. It allows that different explanations of molecular mechanisms directly translate into alternative computational models that predict experimental findings. Moreover,

working with modules provides several options for the engineering of biomodels and their scalable application to systems and synthetic biology.

7.1 Regulatory Interactions Appear in Clear Graphical Structure

Because each module summarizes all functional interactions of a given molecular component including its influences on other components, even complex regulatory interactions can be always displayed in the form of an easy to perceive graphical layout. Certainly, not all of the functional interactions that appear in a module necessarily have to be part of a composed model. With the support of a database, modules can be selected according to user-defined criteria and then automatically linked to give a functional model. Those interactions that do not find a counterpart in one of the selected modules remain inactive because the respective places remain unmarked in the composed Petri net.

7.2 Modules May Be Added, Removed, or Exchanged: *in silico* Mutation of Networks

The modular structure allows to remove or exchange modules when automatically composing a model without touching or even considering modules that remain unchanged. This is a crucial advantage as compared to the manual re-engineering of monolithic models which requires careful consideration of how the components are wired up with each other since overlooking connections might accidentally introduce modelling errors. With the modular approach and the built-in version control of the database, different versions of a given module can be easily exchanged. This can be very helpful to analyse how alternative kinetic mechanisms of molecular interactions would influence the overall behaviour of the system.

For example for the phosphate regulatory network, one might wish to analyse whether or not different activation mechanisms of the PhoB protein would change the gene expression response and the performance of the feed-back loop. When working with really complex models, the modeller investigating a local mechanism has not to care about the inner life of all the numerous modules, in analogy to programming where the procedures of an approved library of subroutines do not have to be reconsidered each time they are used for building a new program.

At the moment where the database will contain a relatively high number of modules, automatically generated models might reveal nonobvious regulatory interactions of molecular components, bring them into a quantitative context and predict nonobvious and eventually counterintuitive consequences of network activation or perturbation. This will definitely be the case when regulatory interactions at a genome-wide scale change gene expression levels that translate into an updated marking of the places of protein modules due to the change in cellular concentration of perhaps many proteins. Even without modelling, this is already obvious by just looking at the gene expression data mentioned in the *Physarum* case study.

One might systematically probe components for their global role in the biomolecular system by simply removing modules from a model. This is the *in silico* complement to systematic mutant screens that are performed in genetic model organisms. *In silico* mutational studies may turn out to be of great benefit for synthetic biology in all cases where systematic mutant screens cannot be applied for what reasons soever.

7.3 Modules Can Be Automatically Generated at Large Scale

For genome scale models where the regulatory control of hundreds or thousands of genes or proteins is to be considered, automatic generation of models becomes an issue. By creating multiple copies of the module prototypes described here, modules can be generated fully automatically simply by assigning names to places and transitions. This is especially straightforward for the gene and mRNA modules but also for protein degradation modules. We expect that automatically generated large scale Petri nets will transmute into helpful tools for the reverse engineering of models from transcriptomic and proteomic data sets.

7.4 Allelic Influence Modules Integrate Forward and Reverse Approaches to Biomodel Engineering

Allelic influence modules were designed to represent regulatory influences of mutated genes (the alleles of a gene) onto the system. Unlike in the other module types, these influences may be rather indirect and may involve a number of potentially unknown components. As we have shown, defining these modules allows to reverse engineer networks from data collected on mutants. These reverse engineered networks are indeed fully compatible with the molecule-centred modules through transitions that control the active states of a gene as shown in the case study.

Acknowledgement. We thank Mostafa Herajy, Fei Liu, Christian Rohr and Martin Schwarick for their continuous support in developing Snoopy.

References

1. Baldauf, S.L., Doolittle, W.F.: Origin and evolution of the slime molds (Mycetozoa). *Proc. Natl. Acad. Sci. USA* 94, 12007–12012 (1997)
2. Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., Doolittle, W.F.: A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972–976 (2000)
3. Barrantes, I., Glöckner, G., Meyer, S., Marwan, W.: Transcriptomic changes arising during light-induced sporulation in *Physarum polycephalum*. *BMC Genomics* 11, 115 (2010)
4. Blätke, M.A., Dittrich, A., Rohr, C., Heiner, M., Schaper, F., Marwan, W.: JAK/STAT signalling - an executable model assembled from molecule-centred modules demonstrating a module-oriented database concept for systems- and synthetic biology (submitted 2012)

5. Blätke, M.A., Meyer, S., Marwan, W.: Pain signaling - A case study of the modular Petri net modeling concept with prospect to a protein-oriented modeling platform. In: Proceedings of the 2nd International Workshop on Biological Processes & Petri Nets (BioPPN 2011), Newcastle upon Tyne, United Kingdom, pp. 1–19 (2011)
6. Blätke, M.A.: BioModelKit (2012), <http://www.biomodelkit.org>
7. Burland, T.G., Solnica-Krezel, L., Bailey, J., Cunningham, D.B., Dove, W.F.: Patterns of inheritance, development and the mitotic cycle in the protist *Physarum polycephalum*. *Adv. Microb. Physiol.* 35, 1–69 (1993)
8. Cerutti, H., Casas-Mollano, J.A.: On the origin and functions of RNA-mediated silencing: from protists to man. *Current Genetics* 50(2), 81–99 (2006)
9. Chia, N.Y., Chan, Y.S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.S., Huss, M., Soh, B.S., Kraus, P., Li, P., Lufkin, T., Lim, B., Clarke, N.D., Bard, F., Ng, H.H.: A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* 468(7321), 316–320 (2010)
10. Cuomo, A., Bonaldi, T.: Systems biology "on-the-fly": SILAC-based quantitative proteomics and RNAi approach in *Drosophila melanogaster*. *Methods in Molecular Biology* 662, 59–78 (2010)
11. Durzinsky, M., Marwan, W., Ostrowski, M., Schaub, T., Wagler, A.: Automatic network reconstruction using ASP. *Theory and Practice of Logic Programming* 11, 749–766 (2011)
12. Durzinsky, M., Wagler, A., Marwan, W.: Reconstruction of extended Petri nets from time series data and its application to signal transduction and to gene regulatory networks. *BMC Systems Biology* 5(1), 113 (2011)
13. Fisher, J., Henzinger, T.A.: Executable cell biology. *Nature Biotechnology* 25(11), 1239–1249 (2007)
14. Gerber, A., Luschig, S., Krasnow, M., Brown, P.: Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* 103(12), 4487–4492 (2006)
15. Glöckner, G., Golderer, G., Werner-Felmayer, G., Meyer, S., Marwan, W.: A first glimpse at the transcriptome of *Physarum polycephalum*. *BMC Genomics* 9, 6 (2008)
16. Hecker, M., Lambeck, S., Toepfer, S., Van Someren, E., Guthke, R.: Gene regulatory network inference: data integration in dynamic models—a review. *Bio Systems* 96(1), 86–103 (2009)
17. Heiner, M., Gilbert, D., Donaldson, R.: Petri Nets for Systems and Synthetic Biology. In: Bernardo, M., Degano, P., Zavattaro, G. (eds.) SFM 2008. LNCS, vol. 5016, pp. 215–264. Springer, Heidelberg (2008)
18. Heiner, M., Lehrack, S., Gilbert, D., Marwan, W.: Extended stochastic Petri nets for model-based design of wetlab experiments. *Transactions on Computational Systems Biology XI*, 138–163 (2009)
19. Heiner, M., Herajy, M., Liu, F., Rohr, C., Schwarick, M.: Snoopy – A Unifying Petri Net Tool. In: Haddad, S., Pomello, L. (eds.) PETRI NETS 2012. LNCS, vol. 7347, pp. 398–407. Springer, Heidelberg (2012)
20. Hoffmann, X.K., Tesmer, J., Souquet, M., Marwan, W.: Futile attempts to differentiate provide molecular evidence for individual differences within a population of cells during cellular reprogramming. *FEMS Microbiology Letters* 329(1), 78–86 (2012)
21. Hsieh, Y.J., Wanner, B.L.: Global regulation by the seven-component Pi signaling system. *Current Opinion in Microbiology* 13(2), 198–203 (2010)

22. Ideker, T., Krogan, N.J.: Differential network biology. *Molecular Systems Biology* 8, 565 (2012)
23. Kolch, W., Calder, M., Gilbert, D.: When kinases meet mathematics: the systems biology of MAPK signalling. *FEBS Letters* 579(8), 1891–1895 (2005)
24. Marwan, W., Rohr, C., Heiner, M.: Petri nets in Snoopy: a unifying framework for the graphical display, computational modelling, and simulation of bacterial regulatory networks. *Methods in Molecular Biology* 804, 409–437 (2012)
25. Moffat, J., Sabatini, D.M.: Building mammalian signalling pathways with RNAi screens. *Nature Reviews Molecular Cell Biology* 7(3), 177–187 (2006)
26. Neidhardt, F., Ingraham, J., Schaechter, M.: *Physiology of the Bacterial Cell. A Molecular Approach*. Sinauer Associates, Sunderland (1990)
27. Orton, R., Adriaens, M., Gormand, A., Sturm, O., Kolch, W., Gilbert, D.: Computational modelling of cancerous mutations in the EGFR/ERK signalling pathway. *BMC Systems Biology* 3(1), 100 (2009)
28. Pinney, J.W., Westhead, R.D., McConkey, G.A.: Petri net representations in systems biology. *Biochem. Soc. Trans.* 31, 1513–1515 (2003)
29. Rohr, C., Marwan, W., Heiner, M.: Snoopy—a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics* 26(7), 974–975 (2010)
30. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M.: Global quantification of mammalian gene expression control. *Nature* 473(7347), 337–342 (2011)
31. Sturm, O.E., Orton, R., Grindlay, J., Birtwistle, M., Vyshemirsky, V., Gilbert, D., Calder, M., Pitt, A., Kholodenko, B., Kolch, W.: The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Science Signaling* 3(153), 90 (2010)
32. Sujatha, A., Balaji, S., Devi, R., Marwan, W.: Isolation of *Physarum polycephalum* plasmodial mutants altered in sporulation by chemical mutagenesis of flagellates. *Eur. J. Protistol.* 41, 19–27 (2005)
33. Uhlmann, S., Mannsperger, H., Zhang, J.D., Horvat, E.A., Schmidt, C., Küblbeck, M., Henjes, F., Ward, A., Tschulena, U., Zweig, K., Korf, U., Wiemann, S., Sahin, O.: Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer. *Molecular Systems Biology* 8, 570 (2012)