Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher, eds

A MODULE-BASED APPROACH TO BIOMODEL ENGINEERING WITH PETRI NETS

Wolfgang Marwan Mary Ann Blätke

Otto-von-Guericke University
Pfälzerstraße 5
39106 Magdeburg, GERMANY

ABSTRACT

Based on Petri nets as formal language for biomodel engineering, we describe the general concept of a modular modelling approach that considers the functional coupling of modules representing components of the genome, the transcriptome, and the proteome in the form of an executable model. The composable, metadata-containing Petri net modules are organized in a database with version control and accessible through a web interface. The effects of genes and their mutated alleles on downstream components are modelled by gene modules coupled to protein modules through RNA modules by specific interfaces designed for the automatic, database-assisted composition. Automatically assembled models may integrate forward and reverse engineered modules and consider cell type-specific gene expression patterns. Prospects for automatic model generation including its application to systems biology, synthetic biology, and to functional genomics are discussed.

1 INTRODUCTION

Since the One Gene – One Protein Hypothesis has originally been proposed by George Beadle and Edward Tatum (Beadle and Tatum 1941) we have learned that the building blocks of life, the genes, the RNAs, the proteins, and the metabolites all together form a complex network of regulatory interactions. This network is robust, adaptive, and to some extent self-healing, as it includes multiple regulatory feedback loops composed of interacting proteins that often involve other types of biomolecules (Figure 1A). The early view that the flow of information within a cell occurs from the genes to the proteins has been revisited by many exciting discoveries that have been made during the past decades. We meanwhile appreciate that in reality the flow of information, in terms of regulatory interactions, occurs back and forth between the components of the different classes of biomolecules (DNA, RNA, proteins, small molecules). We also understand that there is extensive information processing mediated by the network of interacting proteins and that many proteins seem to be just made for fulfilling these computational tasks.

Many qualitative models on molecular mechanisms as well as the corresponding computational (kinetic) models exclusively focus on protein-protein interactions. When working with such models one should keep in mind that the considered networks are not necessarily hard-wired but may change. Alterations in the wiring due to components that may be added, deleted, or modified may be brought about by changes in the pattern of expressed genes. The gene expression pattern in general is responsive to environmental (experimental) conditions, it may depend on the considered cell type, or even on the history of an individual cell and impact stimulus sensing and responses (see (Otomo et al. 1989) for example). Changes in gene expression patterns can be central to regulatory processes. For a given process, the importance of gene regulation may differ from species to species. In fission yeast for example, the cell cycle is regulated mainly through protein-protein interactions. In mammalian cells, the proteins regulating the cell cycle are similar. However, regulation in addition affects changes in the gene expression altering the concentration of proteins involved in cell cycle regulation (Lodish et al. 1996).

For technical reasons, (high-throughput) experimental techniques often reveal information restricted to one class of biomolecule at a time (the genome, the transcriptome, the proteome, the metabolome etc.) or to one type of molecular interaction (e.g. protein-protein or protein-DNA interactions). For a true systems level understanding which systems biology aims at, we have to integrate these data to

obtain a comprehensive view of cellular regulation based on dynamic models with predictive power. This calls for suitable computational frameworks that can fulfil this task.

In this context, we promote Petri nets as they offer several features the combination of which makes them a versatile framework for modelling and simulation in systems and synthetic biology (Fisher and Henzinger 2007; Heiner et al. 2008).

1. **Petri nets** are a **formal modelling language** with a strictly defined syntax and semantics.
2. There is a number of excellent **tools for analysis and simulation** of Petri nets provided by an active community.
3. By using an appropriate tool like Snoopy, one can choose to interpret a given Petri net as **discrete, stochastic, continuous, or hybrid** model and perform respective **simulations by** simply **executing the model** (Rohr et al. 2010).
4. The graphical representation of the model structure is automatically translated into the mathematical equations used for simulations. This **WYSIWYG** feature is especially welcome by biologists and helps to avoid misunderstandings between experimentalists and theoreticians.
5. Because Petri nets as a formal language are mathematically defined, their **structure** and their **dynamic behaviour** can be **treated mathematically** (e.g. (Durzinsky et al. 2012; Durzinsky et al. 2011; Marwan et al. 2008)).
6. **Coloured Petri** nets are an extremely powerful extension as they combine the strengths of the various types of Petri nets with the **expressive power of a programming language** (Liu and Heiner 2012). This becomes particularly relevant for building realistic models of complex biosystems at the molecular and the cellular level and their simulation in the continuous, stochastic, or hybrid world.
7. Last but not least, because the **graphical representation** of Petri nets is very similar to the way biologists usually draw molecular interactions and biochemical reactions, Petri nets are easily and **intuitively accessible** to wet lab scientists, even to those that lack any mathematical skills. On the other hand, the strictly **defined formalism** of Petri nets enforces bioscientists to be **consistent in the description** of a biological process in the form of a model.

Building on these features, we use Petri nets to design a strictly modular approach to biomodel engineering (Breitling et al. 2010). In this article, we assume basic knowledge on Petri nets. Readers not familiar with Petri nets or with the application of Petri nets to systems biology may find a brief overview in the review by Pinney (Pinney et al. 2003) and tutorial-like introductions elsewhere (Blätke et al. 2011a; Marwan et al. 2012).

2 MOTIVATION AND PREVIOUS WORK

2.1 Motivation

Kinetic models of regulatory reaction networks are a core component of systems biology. As the power of modelling and simulation of biological systems becomes increasingly evident, and the number of models describing regulatory networks is growing, some inherent disadvantages of conventional models, let us call them monolithic models, emerge. The description of molecular mechanisms in terms of a model and their implementation in the form of equations should be well documented (Waltemath et al. 2011) which often is not the case. Employing formal languages to describe reaction mechanisms and causal dependencies in the form of a strict and simple syntax may help to overcome this problem. Formal languages like Petri nets can be automatically translated into the corresponding list of equations, which are used to run simulations.

At least from our point of view, the formal language alone does not solve the problem completely. Especially for non-modellers, still the vast majority among the bioscientists, the structure of complex, monolithic models, even in formal language, appears neither obvious nor easily accessible. This may be one reason why modelling still is not widely appreciated in the biosciences. Despite from that, monolithic models cannot be easily curated, updated, or modified by persons other than the author of a model, simply due to the complexity of the overall structure. Moreover, most existing models cannot be easily combined with each other without making major adaptations to the model structure.

Instead of creating monolithic models, we therefore propose to systematically create modules made of Petri nets. From the very beginning, these modules are designed for the automatic composition of biomodels, contain searchable metadata for documentation, and are organised in a database.

2.2 Previous Work

The modular (object-centred) approach is based on three essential components: the definition of modules, a few and simple rules according to which modules are designed, and the database that organises the modules and manages different versions thereof. Initially, we developed the modular approach to model signalling networks composed of proteins (Blätke et al. 2012b; Blätke et al. 2011b; Blätke et al. 2010). More recently, we generalised the concept by the definition of modules representing genes or RNAs (Blätke et al. 2012c). This extension allows to consider the impact of gene expression patterns and mutations on protein interaction networks and consequently allows to predict phenotype from genotype.
By providing a biochemical example of cellular signalling we will demonstrate how modular Petri net modelling in principle works. We will then discuss potential applications of modular models to systems and synthetic biology, and prospects of their extension by functional genomics data sets. To start with, let us briefly explain how modules are composed and which module prototypes are defined in order to obtain integrative models.

3 MODULES: STRUCTURE, COMPOSITION, PROTOTYPES, AND GENERAL STRUCTURAL PROPERTIES

3.1 Structure and Composition

According to the fundamental definition, a module represents a corresponding biomolecule and its direct interactions with other biomolecules. As natural biochemical reaction networks occur through the mutual interaction of biomolecules as the natural building blocks of life, executable biomodels are composed of interacting Petri net modules. The modularity of any resulting biomodel mirrors the modular composition of the living system at the molecular level. As we will show later, modules may be used to visualise the molecular evolution of life in terms of families of similar or related modules. This is a truly natural approach to the engineering of biomodels.

Each module is composed of a Petri net and associated metadata consisting of commented lists of places and transitions and of additional documentation like literature citations or links to sequence database entries and further material useful to judge the validity of the Petri net (Blätke et al. 2012a; Blätke et al. 2012b). The Petri net of a module graphically displays subprocesses clearly arranged and spatially well separated from each other. With this clear layout, a module can be easily understood. Accordingly, modules adhere to simple and clearly defined design principles (Figure 1B). Because of the metadata, a module can also serve as a kind of mini-review on the reaction mechanisms of the represented biomolecule and its functional interactions with other components. Being publicly accessible through a database with web interface, modules may be curated by experts and their approval documented accordingly. As the database allows different versions of each module, contradicting views on mechanisms can be fully represented. By simply exchanging modules with a mouse click the impact of contradicting mechanisms on the behaviour of complex systems can be easily evaluated without the need of rebuilding monolithic models manually.

The database plays a central role in the management of model versions and helps selecting modules for the automatic composition of biomodels. Automatically evaluating the lists of places and transitions, the user receives suggestions, which modules may be coupled to each other and alerts if there exist alternative modules that may be incorporated. The database is also helpful in creating modules representing related proteins by providing templates that can be modified as necessary. The different transducer proteins functioning as sensory receptors in photo- and chemotaxis of prokaryotes provide an example (see below).

3.2 Module Prototypes

Originally, protein modules were defined for modelling and simulation of protein interaction networks (Blätke et al. 2012b; Blätke et al. 2011b; Blätke et al. 2010). For creating more comprehensive models of regulatory control, we defined additional module prototypes, each centred around one object (Blätke et al. 2012c). The object can be a protein, a gene or an RNA (Figure 1B). For those cases where a module does not describe known molecular interactions, we define allelic influence modules and causal interaction modules. These extensions are made with the notion that the introduction of

new module types should be restricted to a minimum to keep the system and the rules for module design as simple as possible.

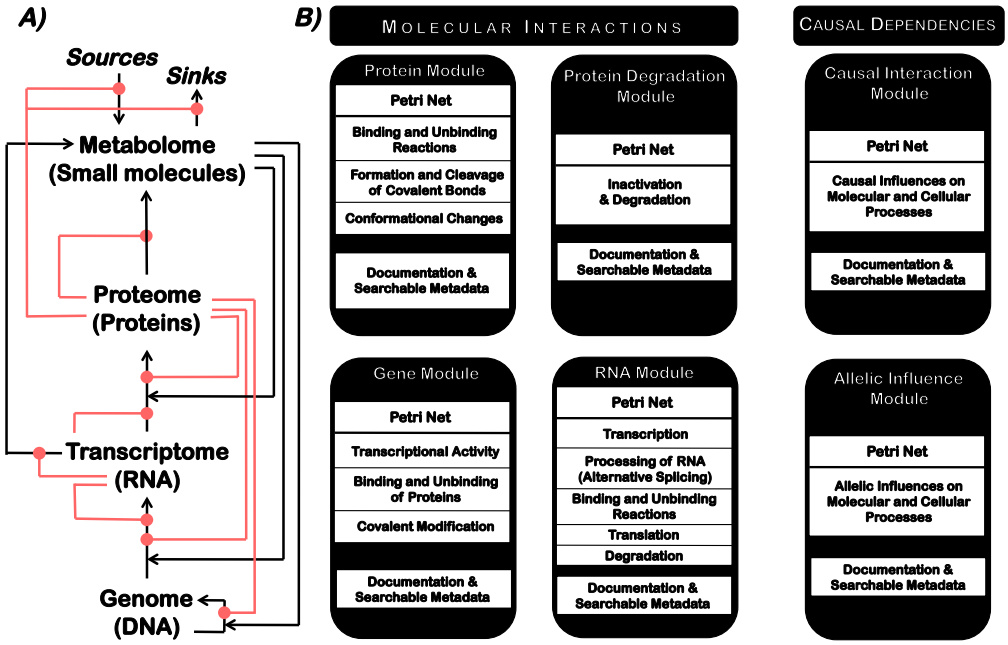

Figure 1: Major regulatory interactions between biomolecules (A) and structures of the modules prototypes (B). Panel (B) was taken from (Blätke et al. 2012c).

Protein modules are centred around one protein molecule as defined by an individual polypeptide chain. A protein module represents the binding of and dissociation from other molecules (e.g. the formation of multi-protein complexes), the formation and cleavage of covalent bonds including the catalysis of biochemical reactions (e.g. the phosphorylation of proteins), and conformational changes that may alter the activity or the properties of a protein.

Protein degradation modules represent the degradation of proteins, which may be regulated by other factors. Protein degradation modules are kept separate from protein modules so one can choose whether or not to consider protein degradation processes within a model.

Gene modules represent the transcriptional activity of a gene, the regulation of gene activity through binding or dissociation of protein factors, and its epigenetic covalent modification.

RNA modules represent the biosynthesis of RNA by transcription of a gene, the posttranscriptional modification of RNA including splicing and alternative splicing reactions, the binding and dissociation of proteins, the translation into the proteins encoded by the mRNA, and the degradation of the RNA including its potential control through proteins or RNAs. While gene modules, protein modules, and protein degradation modules are strictly centred around one molecule, the RNA module represents the primary transcript of a gene but in addition all of its reaction products. The RNA module merges the complex processes of transcription, RNA processing and modification, translation of the mRNAs (protein biosynthesis), and RNA degradation into single Petri net transitions, respectively. Eukaryotic RNAs may be alternatively spliced to form different mature mRNAs derived from a primary transcript. Prokaryotic mRNAs may be bicistronic or polycistronic transcripts, i.e. one RNA module is translated into two or several proteins, respectively. We currently feel that the comprehensive representation of all subsequent reactions of a primary transcript within a RNA module is best for providing information at a glance and does not spoil the benefits of the modular modelling approach.

Causal interaction modules represent causal influences of entities on molecular or cellular processes. Causal interaction modules may be required for reverse biomodel engineering approaches or to model the experimental addition or removal of factors or stimuli.

Allelic influence modules represent the effects of alleles (mutated versions of a gene) on molecular or cellular processes or on the system in general. While a gene module represents known molecular interactions, the allelic influence module is used to represent indirect causal influences. This causality may be mediated through an unknown and arbitrary number of steps and additional components. For-

mally, an allelic influence module could be considered as a subtype of a causal interaction module. Both, allelic influence modules and causal interaction modules represent dependencies and allow the reverse engineering of modules. They are specifically designed to allow the combination of reverse engineered modules with the forward engineered molecular-type modules into a composed biomodel.

4 THE COMPOSITION AND RECOMBINATION OF MOLECULAR MODULES IS A BASIC PRINCIPLE IN THE EVOLUTION OF REACTION NETWORKS: PROKARYOTIC TAXIS AS EXAMPLE

Natural proteins are polypeptide chains of amino acids arranged in strictly linear sequence. Considering the average length of a polypeptide chain of about 300 amino acids and the number of 20 amino acids that are found in proteins, the total number of atoms that would by required to synthesize all the combinatorially possible sequences is higher than the number of atoms in our universe (Alberts et al. 2008). However, only a very small fraction of these sequences is found in nature, presumably due to the constraint put by evolutionary selection that the polypeptide chain must fold into a stable structure in order to give a functional protein. Sequence comparisons show that many proteins are composed of modular parts, called domains that have been assembled into proteins during molecular evolution. Types of domains found in different proteins are defined by their sequence or structural similarity, indicating that functionally important sequences have been conserved through the constraints of structure-function relationships. The taxis of prokaryotes is a paradigm that convincingly demonstrates this principle.

Prokaryotes are simple cells that developed before evolution invented the cell nucleus in creating the eukaryotes. The prokaryotes are composed of two groups, bacteria and archaea (Elkins et al. 2008; Woese et al. 1990). Although bacteria and archaea are very similar in their cellular organisation, archaea have some molecular features in common with the eukaryotes. Therefore, the comparison of bacteria, archaea and eukaryotes gives valuable insights into molecular evolution (Koretke et al. 2003; Koretke et al. 2000; Schaller et al. 2011; Schlesner et al. 2009; Stewart 2010).

Many prokaryotes swim being propelled by flagella or move with the help of other structures (Chen et al. 2011; Herzog and Wirth 2012). These cells can move towards or away from environmental factors by specific sensory receptors that are coupled to the motor organelles via a central signal transduction system, a so-called two-component system (Falke et al. 1997; Schaller et al. 2011). This phenomenon is called taxis.

A typical taxis system of a prokaryotic cell is composed of (chemo-) receptors in the form of methyl-accepting taxis proteins: the central kinase CheA, the response regulator, CheY, enzymes (CheR and CheB) involved in chemosensory adaptation mediated by reversible methylation of the taxis receptors, and adapter proteins like CheW. In its phosphorylated form, CheY interacts with the proteins of the switch complex of the flagellar motor which control its rotational direction and thereby the swimming behaviour of the cell (Figure 2).

Comparison of the evolutionary related taxis systems of different prokaryotes provides fascinating insights into the way of how signal transducing regulatory networks evolved. Although the core system is preserved in the different prokaryotic species, the networks may be composed of additional nodes. These may emerge from proteins derived by gene duplication and mutation, e.g. like the two forms of CheY found in various species ((Schlesner et al. 2009; Szurmant and Ordal 2004) and references therein). In some systems, additional proteins essentially contribute to the dynamic function of the network like the CheV protein in *Bacillus subtilis* (Szurmant and Ordal 2004) or CheD in *Halobacterium salinarum* (Schlesner et al. 2009).

In addition to the molecular variety of the central signal processing core of the taxis network, there is even more variety at the level of the sensory receptors and great divergence at the level of the effector systems, the motor organelles. One example of gaining new functions by the modular recombination of proteins at the level of receptors is the sensory rhodopsin-transducer complex of *Halobacterium*. By combining modified proton pumps with truncated chemoreceptors, nature has created photoreceptors for colour vision (see below for details).

Comparative research on the chemotaxis system of prokaryotes has shown that the core signal transduction machinery with CheY as response regulator can control very diverse effector systems in terms of motor organelles. It controls bacterial-type flagellar motors, which in themselves are a group of structurally variable ion gradient-driven nano-machines of heterogeneous morphology and protein composition (Chen et al. 2011), but also motor structures involved in gliding motility. In archaea, the

core machinery controls a completely different type of flagellar rotary motor (Jarrell and Albers 2012), driven by ATP and obviously evolved from completely different proteins than the bacterial equivalent (Streif et al. 2008).

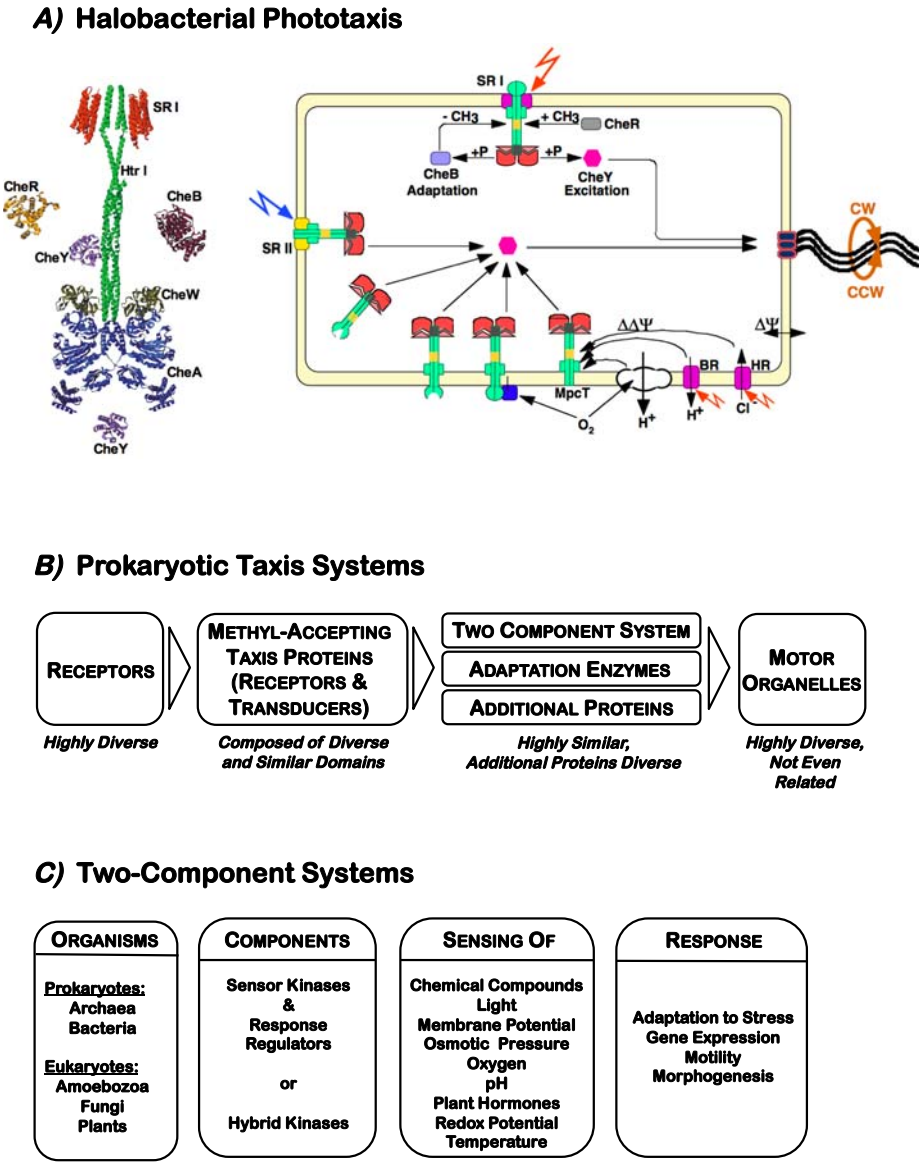

Figure 2: Two-component signalling through the evolutionary (re-) combination of molecular modules. The molecular model in (A) was taken from (Streif et al. 2010) and the scheme of halobacterial phototaxis was redrawn from (Koch and Oesterhelt 2005).

In addition to those two-component systems that are involved in the taxis of prokaryotes, there are many other two component systems involved in sensing environmental or internal factors, many of which control the differential expression of genes in order to adapt the organism to a changing environment (Jung et al. 2012; Schaller et al. 2011; West and Stock 2001). The two components that gave the name consist of a kinase, which is autophosphorylating in response to a stimulus and a response regulator, which is phosphorylated by the kinase and mediates the response of the cell, e.g. the differential expression of a set of genes or the motility response during taxis. There is a tremendous number of two-component systems found in bacteria, archaea, and eukaryotes. All of them are evolutionary related and function with a similar biochemical mechanism. Representing the function of these systems and their molecular interactions by protein module-type Petri nets is an optional formal alternative to the common verbal description of this biological variety and to the representation of these molecules in sequence databases. Such Petri net modules can be organised in the database which we re-

cently implemented (Blätke et al. 2012b) and linked to corresponding sequence data files. In addition to a comparative systematics of reaction mechanisms such modules are seeds for computational models of the many different, but still related molecular processes found in diverse organisms. Systematically creating such modules would provide in the medium term a valuable resource for systems and synthetic biology applications.

This excursus to the evolution and the modular nature of regulatory networks may demonstrate that Petri nets designed as modules indeed do reflect the natural modularity of regulatory networks in the form of graphically displayed computational models.

5 COMPOSING AN EXECUTABLE MODEL FROM MODULES: THE HALOBACTERIAL PHOTOTAXIS RECEPTOR SRI-HTRI-COMPLEX AS EXAMPLE

We will give the receptors mediating halobacterial phototaxis as example to briefly demonstrate the design of Petri net modules (Figure 3) and their composition into an executable model. We will not present a complete model of the phototaxis system as this would go far beyond the scope of this paper.

The cell membrane of *Halobacterium salinarum* contains four retinal proteins, bacteriorhodopsin (BR), halorhodopsin (HR), sensory rhodopsin I (SRI), and sensory rhodopsin II (SRII). These archaeal rhodopsins are evolutionary closely related as evident by their highly similar amino acid sequences. Through their covalently bound retinal chromophore, these proteins can absorb light of the visible or near uv range of the spectrum. Upon light absorption, the retinal chromophore photoisomerises from all-*trans* to 13-*cis* conformation and each of the four proteins proceeds through a so-called photocycle, a sequence of intermediate conformational states. Re-isomerisation of the chromophore to all-*trans* retinal is catalysed by the protein moiety and the directionality of the photocycle is thermodynamically driven by part of the energy of the absorbed photon, which is transiently stored in tensed protein conformational states. Because these photocycle intermediates absorb light of different wavelength, they can be detected by optical spectroscopy with high time resolution.

Bacteriorhodopsin and halorhodpsin are light-driven ion pumps that power the metabolism by energising the cell membrane, while sensory rhodopsin I and sensory rhodopsin II act as light sensors in conjunction with their cognate transducers. Let us first consider the reactions of the sensory rhodopsins and return to the ion pumps later.

In its initial state, SRI maximally absorbs orange light of 587 nm wavelength (Figure 3C). The metastable SRI_{373} photointermediate can return to the SRI_{587} initial state through two alternative ways. It may return through the relatively slow light-independent thermal re-isomerisation of the retinal chromophore or by a fast photochemical re-isomerisation through the SRI_{510} intermediate if the SR_{373} intermediate absorbs an ultraviolet photon. These alternative photochemical pathways antagonistically control the switch complex at the flagellar motor in suppressing (orange light alone) or by activating (orange together with uv light) motor reversals resulting in the attraction of the cell towards or the repulsion away from the stimulus, respectively.

In living cells, SRI forms a stable complex with its cognate transducer HtrI (Krah et al. 1994). Based on sequence similarity, HtrI is closely related to methyl-accepting chemotaxis protein-like chemoreceptors. However in contrast to chemoreceptors, HtrI (likewise HtrII) lacks any domain at the extracellular side that could bind the chemoattractant as it is the case in corresponding chemoreceptors (Ferrando-May et al. 1993; Yao and Spudich 1992). Binding of the transducer HtrI increases the rates of the photochemical reaction cycle, as indicated by the introduction of respective transitions in the SRI protein module. In the HtrI module, the photoreactions of SRI are also considered as they change the activity of the cytoplasmic domain of HtrI in modulating the activity of the CheA kinase. In addition, the module contains reversible methylation reactions of the transducer that mediate sensory adaptation to the stimulus background (not shown).

In normal (wild-type) cells, the expression of the genes *sopI* and *htrI* encoding SRI and HtrI, respectively, are controlled by the same operon (Yao and Spudich 1992). The activity of the gene depends on the oxygen tension in the growth medium and accordingly on the density of the culture (Otomo et al. 1989). Cells of freshly inoculated cultures express the SRI-HtrI complex only at low amount. During growth of the culture, the expression then increases drastically. In the gene module this is indicated by the oxygen tension influencing the probability that the gene is in its active state (Figure 3A). In the wild-type the two genes are transcribed together in the form of a single RNA molecule, a so-called bicistronic message, which is translated into the two proteins, SopI and HtrI (Fig-

ure 3B, upper part). Expression of the *sopI* gene from the plasmid yields a monocistronic mRNA which is transduced into the protein (Figure 3B, lower part).

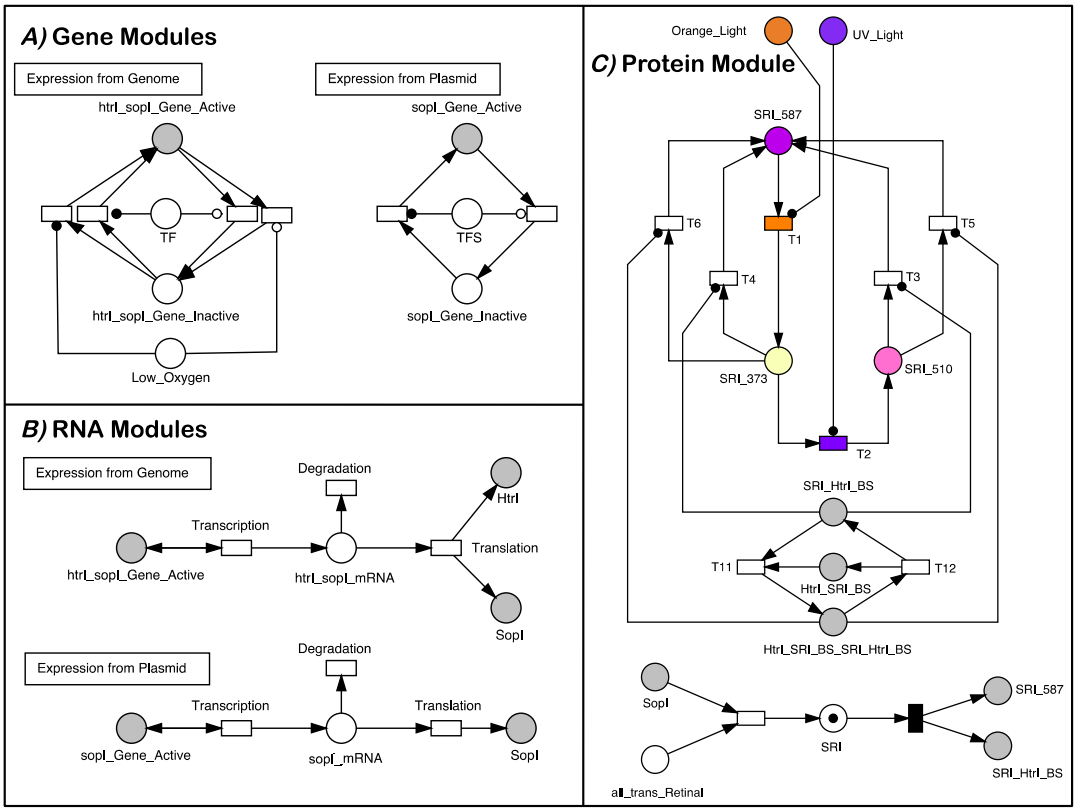

Figure 3: The halobacterial phototaxis receptor and its biosynthesis represented as Petri net modules. The composed modules form a coherent model through the logical places shaded in grey.

By overexpression of SRI as compared to HtrI it has been shown that SRI and HtrI form a stable complex (Krah et al. 1994). In these experiments, SRI was overexpressed five-fold as compared to the wild-type level from the *sopI* gene which was put under the control of a strong promoter and artificially introduced into the cell (Krah et al. 1994). As HtrI was expressed together with SRI from its natural gene through translation of the bicistronic message, there was a 5- to 6-fold excess of SRI as compared to HtrI. Measuring light-induced absorbance changes in the membrane fraction of the overexpressing strain revealed that the 587 nm absorbance signal of SRI_{587} decayed with a bi-exponential kinetic where the slow component corresponded to the overexpressed SRI. This indicated that (1) HtrI accelerates the photocycle of SRI five-fold and that (2) the two molecules form a stable complex. If HtrI would simply catalyse the decay of the SRI_{373} molecules by transiently forming enzyme substrate complexes, a mono-exponential kinetic would be observed.

Two separate gene modules are used to model the wild-type *sopI* gene and the artificially introduced *sopI* gene with its strong promoter (Figure 3A). These two different gene modules of the *sopI* gene use different RNA modules as the mRNA transcribed from the wild type *sopI* gene is bicistronic while the mRNA transcribed from the artificially introduced *sopI* gene is monocistronic (Figure 3B). As the protein molecules of SopI expressed from the wild type gene and the artificially introduced gene have the same amino acid sequence, only one protein module is required (Figure 3C). The bi-exponential decay observed in the spectroscopic experiments results from the fact that the relative amount of HtrI is limiting as indicated by the number of tokens in the respective HtrI places.

In addition to HtrI, *Halobacterium salinarum* has 15 other related transducer molecules with highly similar cytoplasmic signalling domain (Pfeiffer et al. 2008). Due to the modular composition of these Htr orthologs, these proteins are able to sense diverse stimuli (chemical, temperature, pH, membrane potential, oxygen, cytoplasmic arginine level and others). One of this group, HtrII, forms a complex with SRII (Wang et al. 2012) and thereby relays the light signal to the flagellar motor. Repre-

senting these 16 transducers in the form of Petri net protein modules can be performed by re-using a once designed module by copy/paste, as the biochemistry of receptor modification through reversible methylation of glutamate residues and the control of CheA activation through the cytoplasmic domain is the same in all transducers. In order to create a specific module for each of the transducers, only the part of the module concerned with the sensing of the specific stimulus has to be adapted accordingly and the number of methylation sites has to be taken into account. Adhering to templates in creating analogous modules (also in modelling the transducers of other species) not only saves time, but -more importantly- makes modules easily readable by adding a recognition value.

6 MODELS FROM MODULES: PROSPECTS OF AUTOMATIC MODEL GENERATION AND ITS APPLICATION TO SYSTEMS BIOLOGY, SYNTHETIC BIOLOGY, AND TO FUNCTIONAL GENOMICS

Petri net modules are specifically designed for automatic composition in variable combination and may be executed as continuous, stochastic, or hybrid models or merely be used to qualitatively simulate a causal sequence of discrete events. By choosing from the database which modules are to be composed allows to create desired models from a repertoire of pre-existing, exchangeable, re-useable, and curated (approved) modules (Figure 4). The different module prototypes (Figure 1B) provide extensive flexibility in considering components of the proteome, the transcriptome, and the genome. Following the pattern of a prototype as a template helps to create new modules that are easily readable when the underlying pattern of the prototype can be recognised.

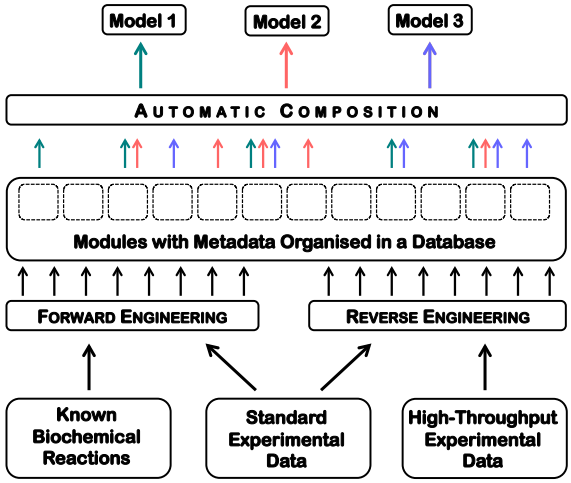

Figure 4: Composition of models through variable combination of modules

In building a model from modules, one may choose interactively and from case to case which biomolecules to include, whether or not to consider protein degradation, RNA stability, or the regulation of gene expression. During interactive model composition one may also choose whether or not to consider alternative regulatory mechanisms that are suggested by the database in the form of submitted alternative versions of a module.

Beyond this state of the art, modular modelling opens exciting prospects regarding algorithmic generation of new modules, the Petri net places of which may in turn be initialised by importing high-throughput data sets. The conserved prototype structure of gene and RNA modules supports the automatic generation of modules, which can be helpful when thousands of genes are to be considered in a model. By importing gene expression data for example, one may be able to generate models that specifically consider the protein composition of different cell types under given physiological condition. Gene expression patterns revealing the presence or absence of certain proteins can have a direct impact on structure and function of regulatory networks. Automatically generated models may provide a formal framework to systematically evaluate correlations between transcript and protein abundance on a global scale (Schwanhäusser et al. 2011). As a long-term prospect, such approaches may also have some relevance for personalised medicine, for example by evaluating consequences of changed gene expression patterns.

The automatic generation of models by combining modules from different organisms from a large repertoire may provide a powerful way to systematically search for synthetically created networks with desired properties. This could be an ongoing effort combining the steadily increasing biological knowledge with the steadily increasing (distributed) computing power. Upon incorporating known mechanisms of molecular evolution (including gene shuffling, etc.) one might be able to perform evolution of molecular networks *in silico* by setting selection criteria. For this purpose, the metadata of the modules are essential for recombining modules in the form of biochemically realistic scenarios.
Finally, automatic generation of models from modules may be used for *in silico* mutant screens. Types of mutations could be deletions or hyperactivations. Making essential use of the metadata of the modules, *in silico* mutations might even consider changes in the specificity of molecular interactions. In model organisms, *in silico* mutagenesis might complement experimental random mutagenesis screens to reveal how complex phenotypes may alternatively occur. For multicellular organisms, *in silico* mutagenesis might be employed to understand the consequences of somatic mutations or to overcome the restrictions of embryonic lethality.

ACKNOWLEDGEMENT

We thank Monika Heiner for the invaluable ongoing dialogue on Petri nets, and the members of her group, Christian Rohr, Martin Schwarick, Fei Liu, and Mostafa Herajy for continuous support in the development of Snoopy. We also thank David Gilbert for stimulating discussions regarding possible applications of modular modelling to synthetic biology.

REFERENCES

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. 2008. Molecular Biology of the Cell. *Garland Science*, New York.

Beadle, G.W., and Tatum, E.L. 1941. Genetic control of biochemical reactions in *Neurospora*. *Proceedings of the National Academy of Sciences of the United States of America*. **27**(11): 499-506.

Blätke, M.A., Dittrich, A., Heiner, M., Schaper, F., and Marwan, W. 2012a. JAK-STAT signalling as example for a database-supported modular modelling concept. *In Computational Methods in Systems Biology*, Springer LNCS/LNBI 7605: 362-365.

Blätke, M.A., Dittrich, A., Rohr, C., Heiner, M., Schaper, F., and Marwan, W. 2012b. JAK/STAT signalling - an executable model assembled from molecule-centred modules demonstrating a module-oriented database concept for systems- and synthetic biology. http://arxiv.org/abs/1206.0959v1; submitted.

Blätke, M.A., Heiner, M., and Marwan, W. 2012c. Predicting phenotype from genotype through automatically composed Petri nets. *In Compuational Methods in Systems Biology*, Springer LNCS/LNBI 760587-106.

Blätke, M.A., Heiner, M., and Marwan, W. 2011a. Tutorial - Petri Nets in Systems Biology. Technical Report. Otto von Guericke University and Magdeburg Centre for Systems Biology, Magdeburg. http://www.regulationsbiologie.de/Publications/BlaetkeTutorial.pdf

Blätke, M.A., Meyer, S., and Marwan, W. 2011b. Pain signaling - A case study of the modular Petri net modeling concept with prospect to a protein-oriented modeling platform. *Proceedings of the 2nd International Workshop on Biological Processes & Petri Nets (BioPPN2011), Newcastle upon Tyne,* United Kingdom: 1-19.

Blätke, M.A., Meyer, S., Stein, C., and Marwan, W. 2010. Petri net modeling via a modular and hierarchical approach applied to nociception. *Proceedings of the 1st International Workshop on Biological Processes & Petri Nets (BioPPN2010), Braga, Protugal.* p. 131.

Breitling, R., Donaldson, R., Gilbert, D., and Heiner, M. 2010. Biomodel engineering–From structure to behavior. T*ransactions on Computational Systems Biology XII*: 1-12.

Chen, S., Beeby, M., Murphy, G.E., Leadbetter, J.R., Hendrixson, D.R., Briegel, A., Li, Z., Shi, J., Tocheva, E.I., ller, A.M.u., Dobro, M.J., and Jensen, G.J. 2011. Structural diversity of bacterial flagellar motors. *The EMBO Journal* **30**(14): 2972-2981.

Durzinsky, M., Marwan, W., and Wagler, A. 2012. Reconstruction of extended Petri nets from time-series data by using logical control functions. *Journal of Mathematical Biology, Online First™*, 2 Februar 2012.

Durzinsky, M., Wagler, A., and Marwan, W. 2011. 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.

Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., Kunin, V., and Anderson, I. 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences of the United States of America* **105**(23): 8102.

Falke, J., Bass, R., and Butler, S. 1997. The two-component signaling pathway of bacterial chemotaxis: A molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu Rev Cell Dev Biol* **13**: 457-512.

Ferrando-May, E., Krah, M., Marwan, W., and Oesterhelt, D. 1993. The methyl-accepting transducer protein HtrI is functionally associated with the photoreceptor sensory rhodopsin I in the archaeon *Halobacterium salinarium*. *EMBO J*. **12**: 2999-3005.

Fisher, J., and Henzinger, T.A. 2007. Executable cell biology. *Nature Biotechnology* **25**(11): 1239-1249.

Heiner, M., Gilbert, D., and Donaldson, R. 2008. Petri nets for systems and synthetic biology. S*chool on Formal Methods*. Spriner, LNCS 5016: 215-264.

Herzog, B., and Wirth, R. 2012. Swimming behavior of selected species of archaea. *Applied and Environmental Microbiology* **78**(6): 1670-1674.

Jarrell, K.F., and Albers, S.-V. 2012. The archaellum: an old motility structure with a new name. *Trends in Microbiology* **20**(7): 307-312.

Jung, K., Fried, L., Behr, S., and Heermann, R. 2012. Histidine kinases and response regulators in networks. *Current Opinion in Microbiology* **15**(2): 118-124.

Koch, M.K., and Oesterhelt, D. 2005. MpcT is the transducer for membrane potential changes in Halobacterium salinarum. *Molecular Microbiology* **55**(6): 1681-1694.

Koretke, K., Volker, C., Bower, M., and Lupas, A. 2003. Molecular evolution of histidine kinases. *In Histidine Kinases. Edited by* M. Inouye, and R. Dutta. Academic Press.

Koretke, K.K., Lupas, A.N., Warren, P.V., Rosenberg, M., and Brown, J.R. 2000. Evolution of two-component signal transduction. *Mol. Biol. Evol.* **17**: 1956-1970.

Krah, M., Marwan, W., Verméglio, A., and Oesterhelt, D. 1994. Phototaxis of *Halobacterium salinarium* requires a signalling complex of sensory rhodopsin I and its methyl-accepting transducer HtrI. *EMBO J*. **13**: 2150-2155.

Liu, F., Heiner, M and Rohr, C. 2012. The manual for colored Petri nets in Snoopy–QPN C/SPN C/CPN C/GHPN C. Computer Science Reports Brandenburg University of Technology, Cottbus.

Lodish, H., Baltimore, D., Berk, A., Zipursky, S.L., Matsudaira, P., and Darnell, J. 1996. Molekulare Zellbiologie. *Walter de Gruyter*, Berlin.

Marwan, W., Rohr, C., and Heiner, M. 2012. Petri nets in Snoopy: a unifying framework for the graphical display, computational modelling, and simulation of bacterial regulatory networks. *In Methods in Molecular Biology*. *Edited by* J. van Helden, A. Toussaint, and D. Thieffry. Humana Press, Clifton, NJ. pp. 409-437.

Marwan, W., Wagler, A., and Weismantel, R. 2008. A mathematical approach to solve the network reconstruction problem. *Mathematical Methods of Operations Research* **67**(1): 117-132.

Otomo, J., Marwan, W., Oesterhelt, D., Desel, H., and Uhl, R. 1989. Biosynthesis of the two halobacterial light sensors P480 and SR and variation in gain of their signal transduction chains. *J. Bacteriol*. **171**: 2155-2159.

Pfeiffer, F., Broicher, A., Gillich, T., Klee, K., Mejía, J., Rampp, M., and Oesterhelt, D. 2008. Genome information management and integrated data analysis with HaloLex. A*rchives of Microbiology* **190**(3): 281-299.

Pinney, J.W., Westhead, R.D., and McConkey, G.A. 2003. Petri Net representations in systems biology. *Biochem. Soc. Trans.* **31**: 1513-1515.

Rohr, C., Marwan, W., and Heiner, M. 2010. Snoopy - a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics* **26**(7): 974-975.

Schaller, G.E., Shiu, S.-H., and Armitage, J.P. 2011. Two-component systems and their co-option for eukaryotic signal transduction. *Current Biology* **21**(9): R320-R330.

Schlesner, M., Miller, A., Streif, S., Staudinger, W.F., Müller, J., Scheffer, B., Siedler, F., and Oesterhelt, D. 2009. Identification of Archaea-specific chemotaxis proteins which interact with the flagellar apparatus. *BMC Microbiology* **9**: 56.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**(7347): 337-342.

Stewart, R.C. 2010. Protein histidine kinases: assembly of active sites and their regulation in signaling pathways. *Current Opinion in Microbiology* **13**(2): 133-141.

Streif, S., Oesterhelt, D., and Marwan, W. 2010. A predictive computational model of the kinetic mechanism of stimulus-induced transducer methylation and feedback regulation through CheY in archaeal phototaxis and chemotaxis. *BMC Systems Biology* **4**: 27.

Streif, S., Staudinger, W.F., Marwan, W., and Oesterhelt, D. 2008. Flagellar rotation in the archaeon *Halobacterium salinarum* depends on ATP. *Journal of Molecular Biology* **384**(1): 1-8.

Szurmant, L., and Ordal, G. 2004. Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiology and Molecular Biology Reviews* **68**(2): 301-319.

Waltemath, D. et al. 2011. Minimum information about a simulation experiment (MIASE). *PLoS Computational Biology* **7**(4): e1001122.

Wang, J., Sasaki, J., Tsai, A., and Spudich, J.L. 2012. HAMP domain signal relay mechanism in a sensory rhodopsin-transducer complex. Journal of Biological Chemistry **287**: 21316-21325.

West, A.H., and Stock, A.M. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**(6): 369-376.

Woese, C.R., Kandler, O., and Wheelis, M.L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *PNAS* **87**(12): 4576-4579.

Yao, V.J., and Spudich, J. 1992. Primary structure of an archaebacterial transducer, a methyl-accepting protein associated with sensory rhodopsin I. *PNAS* **89**: 11915-11919.