Report from Dagstuhl Seminar 11151

Formal Methods in Molecular Biology

Edited by

Rainer Breitling[1], Adelinde M. Uhrmacher[2], Frank J. Bruggeman[3], and Corrado Priami[4]

1 **University of Glasgow, GB,** `rainer.breitling@glasgow.ac.uk`
2 **Universität Rostock, DE,** `lin@informatik.uni-rostock.de`
3 **CWI – Amsterdam, NL**
4 **Microsoft Research – University Trento, IT**

Abstract

This report documents the program and the outcomes of the Seminar 11151 'Formal Methods in Molecular Biology' that took place in Dagstuhl, Germany, on 10.–15. April, 2011. The most recent advances in Systems Biology were discussed, as well as and the contribution of computational formalisms to the modeling of biological systems, with the focus on stochasticity. About 30 talks were given. The participants formed 5 teams that worked on selected case studies. Two teams were awarded prizes, for their efforts in analyzing and further elucidating published biological models.

Seminar 10.–15. April, 2011 – http://www.dagstuhl.de/11151
1998 ACM Subject Classification J.3 Biology and Genetics.
Keywords and phrases bioinformatics, systems biology, formal modeling, computational biology, stochastic, model, simulation, checking, verification, abstraction, petri nets, process algebra
Digital Object Identifier 10.4230/DagRep.1.4.41
Edited in cooperation with Elzbieta Krepska

1 Executive Summary

Rainer Breitling
Adelinde M. Uhrmacher

The second Dagstuhl Seminar on Formal Methods in Molecular Biology took place from 10–15 April, 2011. 35 participants from 8 countries gathered to discuss the most recent advances in Systems Biology and the contribution of computational formalisms to the successful modeling of biological systems. Major recurrent themes were the description of stochastic phenomena in biology, the modeling of spatial aspects of cellular behavior, and the robustness of cellular switches in the face of molecular noise and uncertainty of parameter inference. The computational modeling approaches applied to these challenges were particularly diverse, ranging from differential equation-based models to various flavors of rule-based languages, Petri Nets and process algebras.

A central component of the seminar was the Second International Biomodeling Competition. Teams formed during the first day and worked on biological case studies using a variety of modeling formalisms and analysis methods; the results were presented on Thursday afternoon and the winner determined by a joint vote of the audience.

The 1st prize went to the team of Kirill Batmanov, Antje Beyer, Matthias Jeschke and Carsten Maus, for their work on 'Synchronization of cell populations'.

The 2nd prize went to the team of Andrea Bracciali, Mostafa Herajy, Pietro Lió, Chris Myers, Brett Olivier, and Natal van Riel for their work on 'A bistable gene switch'.

Special prizes were awarded to the team of Chiara Bodei, Luca Bortolussi, Davide Chiarugi, Maria Luisa Guerriero, Jane Hillston Ivan Mura, Alberto Policriti, and Alessandro Romanel (for 'Critical Analysis'), the team of Mary Ann Blätke, Qian Gao, David Gilbert, Simon Hardy, Monika Heiner, Andrzej Kierzek, Fei Liu and Wolfgang Marwan (for 'Innovative use of Petri Nets', and the team of Maciej Dobrzyński, Mathias John, Céline Kuttler, Bartek Wilczyński and Verena Wolf (for 'A pure stochastical approach').

2 Table of Contents

3 Overview of Talks

3.1 Spatial modeling of the community effect

Kirill Batmanov (Université de Lille I, FR)

We investigated some spatial aspects of a phenomenon called community effect, first described in [1]. The community effect is said to be present in a system if it shows some behavior only when the number of cells in the system is large enough. To study the effects of diffusion on the community effect, we added one-dimensional space constraints on the diffusion to the model of the community effect described in[2]. We found the conditions which are necessary for the community effect in the new spatial model.

To study whether it is possible for the community effect to restrict itself in space through self-regulation, we built a simplified model based on Turing reaction-diffusion theory [3]. Using the gene network for the community effect in sea urchin described in [4] combined with the conditions for formation of Turing patterns [5], we showed by stochastic simulations that such a system indeed can become self-regulated and restrict its area of activation in space.

References

1 Gurdon JB, Tiller E, Roberts J, Kato K. *A community effect in muscle development*. Current biology, 3(1):1-11, 1993.
2 Saka Y, Lhoussaine C, Kuttler C, Ullner E. *Theoretical basis of the community effect in development*. To appear in BMC Systems Biology.
3 Turing AM. *T*The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 237(641):37, 1952.
4 Bolouri H, Davidson EH. *The gene regulatory network basis of the "community effect", and analysis of a sea urchin embryo example*. Developmental biology, 340(2):170-8, 2010.
5 Kondo S, Miura T. *Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation*. Science, 329(5999):1616-1620, 2010.

3.2 Inductive Analysis when State Enumeration Explodes

Giampaolo Bella(Universita di Catania, IT)

Current computer science research in the domain of biology has helped establishing a suitable formalisation of biological networks as molecular-scale autonomous programmable computers that operate synchronous and asynchronous 'logical' control of biological processes. Each biological process can then be represented as a set of states amenable to model checking.

The computational paradigm underlying model checking is imperative in the sense that each state of the target system is modelled explicitly as an abstract state, that is a set of variables and their current values (also termed an algebra). The reasoning technique is the enumeration of all the possible states to check whether each satisfies a stated property. Clearly, the enumeration requires the system to be finite-state, but even so, very high number of states cannot be practically handled. Clever techniques exist to face this issue, and are led by symbolic model checking, which copes with up to 10^{20} states.

An effective reasoning technique for systems of unbounded size is mathematical induction. When a system is defined by induction, its properties can be assessed by the corresponding inductive proof principle. For example, induction is the natural way to define the set of the natural numbers, and inductive proofs can establish invariant properties of the naturals, such as the sum of the first n numbers. Remarkably, such proofs are independent from the size of n, and are also efficient because a functional computational paradigm is adopted: computation in fact proceeds by term rewriting, which uses the inductive hypothesis as an admissible rewriting rule. The efficiency can be practically realised by proving the mentioned $S_n = \frac{n(n+1)}{2}$ inductively as opposed to calculating S_n for a large n by enumerating the temporary states, which are sums in this case.

Theorem proving tools, also said proof assistants, offer great support to inductive reasoning. The properties of the genetic toggle switch have been studied inductively [1,2]. A theorem establishes stability, that both genes cannot prevail (in terms of concentration levels) at the same time. Attempting to prove that either gene cannot prevail over the other fails, hence enforcing the opposite, desirable property. This sort of meta-proving, together with the stability theorem, provide formal guarantee that the toggle is bistable. The potential to tackle much more complex biological systems, where the combinatorics becomes hardly manageable, appears huge and yet unexplored.

References

1 Bella, G., Liò, P.: *Formal analysis of the genetic toggle*. In Degano, P., Gorrieri, R., eds.: Proc. of the 9th Conference on Computational Methods in Systems Biology (CMSB'09). LNBI 5688, Springer (2009), 96–110.

2 Bella, G., Liò, P.: *Analysing the microRNA-17-92/Myc/E2F/RB Compound Toggle Switch by Theorem Proving*. In Romano, P., eds.: Proc. of the 9th Workshop on Network Tools and Applications in Biology (Nettab'09). Liberodiscrivere (2009), 59-62.

3.3 JAK/STAT-Pathway - A Case Study of the Modular Petri Net Modeling Concept

Mary Ann Blätke (Universität Magdeburg – IBIO, DE)

Here, we present a computational model of the JAK-STAT pathway in IL6-signaling, which describes the molecular mechanisms in a high resoultion. In addition, we introduce our modular modeling concept, which is based on the Petri net formalism and a modular approach. In our modeling concept, every protein, its interactions and intramolecular changes, are represented by an independent submodel, called module. Therefore, each module integrates the wide-spread information about a protein. A comprehensive model can be assembled from a set of modules of interacting proteins. Advantageously, our concept itself is not at all limited to the JAK-STAT pathway. We also propose a platform to organize approved curated modules in order to allow the generation of molecular networks. This platform might be useful in bridging the gaps between experimental bioscientists and theoretically oriented systems biologists.

References

1 Blätke, M.-A., Marwan W. Modular and Hierarchical Modeling Concept for Large Biological Petri Nets Applied to Nociception. German Workshop on Algorithms and Tools for Petri Nets. Cottbus, Germany, 2010.

2 Blätke, M.-A. Petri-Netz Modellierung mittels eines modularen and hierarchischen Ansatzes mit Anwendung auf nozizeptive Signalkomponenten. Otto von Guericke University Magdeburg. 2010 (Diploma thesis).

3 Blätke, M.-A. et al. Petri Net Modeling via a Modular and Hierarchical Approach Applied to Nociception. Int.Workshop on Biological Processes & Petri Nets (BioPPN), satellite event of Petri Nets 2010. Braga, Portugal, 2010.

3.4 Towards a Taxonomy of Causality-Based Biological Properties

Chiara Bodei (University of Pisa, IT)

Main reference C. Bodei, A. Bracciali, D. Chiarugi, R. Gori, "A Taxonomy of Causality-Based Biological Properties," Proc. Third Workshop From Biology To Concurrency and back (FBTC), 2010, pp. 116–133, EPTCS 19.
URL http://dx.doi.org/10.4204/EPTCS.19.8

We formally characterize a set of causality-based properties of metabolic networks. This set of properties aims at making precise several notions on the production of metabolites, which are familiar in the biologists' terminology. From a theoretical point of view, biochemical reactions are abstractly represented as causal implications and the produced metabolites as causal consequences of the implication representing the corresponding reaction. The fact that a reactant is produced is represented by means of the chain of reactions that have made it exist. Such representation abstracts away from quantities, stoichiometric and thermodynamic parameters and constitutes the basis for the characterization of our properties. Moreover, we propose an effective method for verifying our properties based on an abstract model of system dynamics. This consists of a new abstract semantics for the system seen as a concurrent network and expressed using the Chemical Ground Form calculus. We illustrate an application of this framework to a portion of a real metabolic pathway. Some ideas on how to apply our approach to other models is presented.

The talk is mainly based on the paper [1].

References

1 C. Bodei, A. Bracciali, D. Chiarugi, R. Gori, "A Taxonomy of Causality-Based Biological Properties," Proc. Third Workshop From Biology To Concurrency and back (FBTC), 2010, pp. 116–133, EPTCS 19.

3.5 Hybrid approximation of stochastic process algebra models of biological systems

Luca Bortolussi (Universitá di Trieste, IT)

Models in systems biology tend to cluster around two families of mathematical tools: differential equations and stochastic processes. Even though, physically speaking, stochastic models have firmer grounds, their computational analysis is much more costly than that of their differential counterpart. In any case, ODE-based descriptions of biological systems are often valuable and provide deep insights. Indeed, it is known that, limiting to mass action models, ODE's are an approximation of the (average of) stochastic models, and the differences between the two vanish in the thermodynamic limit (i.e. when populations and system's size go to infinity). Recently, there have been many attempts to mix these two techniques, at least as far as simulation of biological systems is concerned, resulting in several hybrid simulation algorithms. Hybrid dynamical systems have also been a hot topic in the last two decades, with much research work spanning across the boundary between computer science and engineering control. The best known model among hybrid dynamical systems are hybrid automata. Stochastic extensions of such concept are also receiving recently much attention. In both cases, most of the interest is in the development of automated reasoning tools rather than in simulation. It is widely recognized that Computational Systems Biology can highly benefit from modeling approaches embodying some stochastic ingredient. A very popular line along which such incorporation is realized, is based on the use of stochastic process algebras, which are proposed as front-end languages to (automatically) generate mathematical models, usually Continuous Time Markov Chains(CTMC). Recently, such process algebra based languages have continuous, while keeping the other discrete and stochastic. The basic idea is to find the best trade off between accuracy and computational efficiency (stochastic simulations are much more expensive than ODE simulation).

In this talk we present a programme which aims to increase even more the flexibility of stochastic process algebras by providing them with a very general semantics based on (stochastic) hybrid systems, encompassing CTMC and ODE as special cases. Such an approach is motivated not only by the gain in flexibility, but also by the goal of exploiting, in a systematic manner, automated reasoning tools to provide as much information as possible from a given model. Our stochastic process algebra of choice is stochastic Concurrent Constraint Programming (sCCP), an extension of CCP in the stochastic setting. In addition to the standard CTMC-based semantics, we have also provided sCCP with an ODE-based semantics and with an hybrid automata based semantics [2]. In particular, we will present a semantics based on Stochastic Hybrid Automata [1], thereby guaranteeing the possibility of parameterizing the degree of continuity introduced in the model. The result is a lattice of hybrid automata models, that increasingly remove discreteness in favor of continuity. The approach allows also a dynamic reconfiguration of such degree, in accordance to properties of the current state of the system. This allows the description in a formal setting of different hybrid simulation strategies, opening the way for their use in the context of process algebra modelling.

We will discuss some biological examples, mentioning the problems involved in identifying the 'right' degree of discreteness (the best compromise between accuracy and efficiency), some mathematical results that guarantee correctess of the approximation (in a limit sense) [3], and the potential of hybrid models for managing uncertainty and for describing multi-scale systems.

References

1 L. Bortolussi and A. Policriti, 2009. Hybrid Semantics of Stochastic Programs with Dynamic Reconfiguration. Proceedings of Second International Workshop on Computational Models for Cell Processes - EINDHOVEN. 3rd November 2009. EPTCS. Vol.6., pp.63-76.
2 L. Bortolussi and A. Policriti, 2010. Hybrid Dynamics of Stochastic Programs. Theoretical Computere Science. Vol.411/20. pp.2052-2077.
3 L. Bortolussi, 2010. Limit behavior of the hybrid approximation of Stochastic Process Algebras. Lecture Notes in Computer Science. Vol.6148/2010. pp.367-381.

3.6 Modelling HIV infection: A computational overview

Andrea Bracciali (University of Stirling, UK)

We survey a modelling experiment regarding HIV infection, which relies on computationally inspired modelling and analysis techniques.The results presented are a case of methodological cross comparison (stochastic and deterministic) and a novel implementation of model checking in therapy validation. On the methodology side, this work represents a paradigm of integrating formal methods and mathematical models as a general framework to study HIV multiple strains during disease progression and the associated therapies. Results ranges from traditional ODE-based deterministic analysis, such as sensitivity analysis, to Gillespie-based stochastic simulations, to a novel application of stochastic model checking to quantitatively measure and compare the efficacy of (idealised models of) HIV-related therapies.

References

1 A. Sorathiya, P. Lió and A. Bracciali 'An integrated modelling approach for R5-X4 mutation and HAART therapy assessment'. Swarm Intelligence. Springer. 4(4), pages 319–340, 2010. (DOI: 10.1007/s11721-010-0046-4).
2 A. Sorathiya, P. Lió and A. Bracciali 'Formal reasoning on qualitative models of coinfection of HIV and Tuberculosis and HAART therapy'. BMC Bioinformatics. 11(1): Asia Pacific Bioinformatics Conference. 2010.
3 M. Aldinucci, A. Bracciali, P. Lió, A. Sorathiya, and M. Torquati, 'StochKit-FF: Effcient Systems Biology on Multicore Architectures'. Euro-Par 2010 Parallel Processing Workshops HeteroPar, HPCC, HiBB, CoreGrid, UCHPC, HPCF, PROPER, CCPI, and VHPC Workshops. Ischia, Italy, 2010. Lecture Notes in Computer Science 6586. Springer.

3.7 Metabolomics Systems Biology

Rainer Breitling (University of Glasgow, UK)

Metabolomics, the comprehensive study of small molecules in a biological system, has a privileged position in Systems Biology; many of the most prominent success stories of the field relate to the analysis of metabolic systems, and metabolic changes are a much closer reflection of physiologically relevant differences in cellular function than, e.g., changes in global gene expression profiles. This presentation discussed a variety of methodological challenges faced

in large-scale metabolomic analysis, focussing on the collection of computational tools that we have developed for the processing, exploration and interpretation of the large data sets created in metabolomic experiments. It concludes by presenting a case study of metabolic modelling, which integrates detailed kinetic information on individual enzymatic reactions and the often noisy and uncertain data provided by metabolomics.

3.8 Single cell signaling and protein expression noise give rise to digital response on the population level

Maciej Dobrzyński (University College - Dublin, IE)

Genetically identical cells can respond differently to extracellular stimuli and may therefore follow different fates within cellular population. The appearance of multiple distinct gene expression states is typically attributed to feedback regulation in nonlinear signaling networks which leads to bistability. Here we present a different mechanism, which also results in heterogenous response between cells in a population but does not rely on feedback topology.

Our measurements of extracellular signal-regulated kinase (ERK) response to epidermal growth factor (EGF) in single HEK 293 cells suggest that the population displays digital behavior where cells assume either ON or OFF states. Based on experiments augmented with stochastic models and computational analysis, we argue that this phenomenon results from the interplay of variability in the amount of network components and the nonlinearity of the network response. Our results may be applicable to a wider class of problems in which single cells need to discretize noisy extracellular cues to achieve a robust response.

3.9 Tav4SB: analysis of kinetic models of biological systems

Anna Gambin (University of Warsaw, PL)

Taverna Workbench (Hull et al., 2006) eases integration of tools for life science research in experiments described as workflows. We provide new services that extend the functionality of Taverna Workbench for systems biology. These services allow to perform numerical ODE simulations or model-checking of SBML (Hucka et al., 2003) models, and they allow to visualize the results of model analysis. As an example of usage we constructed exemplary workflows. Our services are executed in the newly created grid environment, which integrates heterogeneous software such as Mathematica (Wolfram Research, Inc., 2008, Mathematica Edition: Version 7.0), PRISM (Hinton et al., 2006) and SBML ODE Solver. The enduser goal of the Taverna services for Systems Biology (Tav4SB) project is to abstract details of the technology infrastructure 'in the cloud' which supports provided services. User guide, examples and all resources are available from http://bioputer.mimuw.edu.pl/tav4sb Web page.

3.10 Stochastic model of the plant circadian clock - A Bio-PEPA case study

Maria Luisa Guerriero (University of Edinburgh, UK)

Main reference Akman, O.E.; Guerriero, M.L.; Loewe, L. and Troein, C., "Complementary approaches to understanding the plant circadian clock," Proc. Third Workshop From Biology To Concurrency and back (FBTC), 2010, pp. 1–19, EPTCS 19.
URL http://dx.doi.org/10.4204/EPTCS.19.1

Circadian clocks are oscillatory genetic networks that help organisms adapt to the 24-hour day/night cycle. The clock of the green alga Ostreococcus tauri is the simplest plant clock discovered so far. Its many advantages as an experimental system facilitate the testing of computational predictions. We present a model of the Ostreococcus clock in the stochastic process algebra Bio-PEPA and exploit its mapping to different analysis techniques, such as ordinary differential equations, stochastic simulation algorithms and model-checking. The small number of molecules reported for this system tests the limits of the continuous approximation underlying differential equations. We investigate the difference between continuous-deterministic and discrete-stochastic approaches. Stochastic simulation and model-checking allow us to formulate new hypotheses on the system behaviour, such as the presence of self-sustained oscillations in single cells under constant light conditions. We investigate how to model the timing of dawn and dusk in the context of model-checking, which we use to compute how the probability distributions of key biochemical species change over time. These show that the relative variation in expression level is smallest at the time of peak expression, making peak time an optimal experimental phase marker. This is joint work with Ozgur Akman, Laurence Loewe and Carl Troein, published in [1].

References

1 Akman, O.E.; Guerriero, M.L.; Loewe, L. and Troein, C., "Complementary approaches to understanding the plant circadian clock," Proc. Third Workshop From Biology To Concurrency and back (FBTC), 2010, pp. 1–19, EPTCS 19.

3.11 Analysis of the regulatory motif dynamic of signaling networks using Petri net-based dynamic graphs

Simon Hardy (Mount Sinai Medical School – New York, US)

Numerous signaling pathways have been discovered in the mammalian cell and they form a highly connected network. It is generally accepted that this signaling apparatus not only transmits information from sources to targets, but also processes and transforms signals using biological control devices known as regulatory motifs. The simpler motifs are loops and bifans, but multiple motifs can be aggregated together to form more complex processing systems. As a result, cellular modules can exhibit high-level behaviors like bistability, oscillation, pulse generation and noise filtering. We have developed the dynamic graph using Petri net theory to analyze the emergent behavior of interacting regulatory motifs and we will present some examples.

3.12 Generalized Hybrid Petri Nets

Mostafa Herajy (BTU Cottbus, DE)

Recently hybrid modelling and simulation of biochemical systems have attracted increasing interest. This is motivated by the need of simulating systems which integrate different sub-cellular models, and the fact that bio networks themselves are inherently stochastic, however stochastic simulation is time expensive.

Compared to other methods of biological modelling, Petri nets are characterized by their intuitive visual representation and executability of biological models.

Generalized Hybrid Petri Nets (GHPN) are recently introduced into Snoopy to combine both continuous Petri nets and generalized stochastic Petri nets. The modelling power of this class of Petri nets combines stochastic, discretely timed and continuous reactions into one model, which permits representing biological switches, in which continuous elements are turned on/off based on discrete elements. Moreover the defined model can be simulated using static or dynamic partitioning. The implementation of GHPN is freely available as part of Snoopy. In this talk, we provide a short presentation of the GHPN's elements as well as how it can be used to represent and simulate stiff biochemical networks.

3.13 Equivalence and Discretisation in Bio-PEPA

Jane Hillston (University of Edinburgh, UK)

Bio-PEPA is a process algebra for modelling biological systems. An important aspect of Bio-PEPA is the ability it provides to discretise concentrations resulting in a smaller, more manageable state space. The discretisation is based on a step size which determines the size of each discrete level and also the maximum number of levels.

This talk considers equivalence relations for Bio-PEPA, particularly an equivalence addressing the relationship between two discretisations of the same Bio-PEPA model that differ only in the step size and hence the maximum number of levels. We use the idea of bisimulation from concurrency and process algebra. We present a novel behavioural semantic equivalence, compression bisimulation, that equates two discretisations of the same model.

3.14 Investigating the Switch-like Response in Yeast Pheromone Signaling

Matthias Jeschke (Heidelberg, DE)

Coming originally from the field of stochastic simulation, the models used in previous work were artificially constructed to ease the performance evaluation of simulation algorithms in terms of execution speed and accuracy. But results from those studies tell only half the story.

That is why I have decided to focus on the application site of my work, i.e. the development and analysis of real-world models, e.g., the pheromone signaling pathway found in yeast cells. Still at the beginning, the talk will provide a brief overview of this pathway and provide some details about selected components, especially the scaffold protein Ste5 and its crucial yet not entirely unraveled role in tuning the pathway response.

3.15 Reaction Rules with Constraints

Mathias John (IRI – Villeneuve d'Ascq, FR)

Chemical reaction rules are the natural formalism for the stochastic modeling of biochemical systems. However, they only provide limited expressiveness. In this talk, I will present the modeling language React(C) that allows to impose constraints on the occurrence of reactions. I will roughly describe how rules with constraints allow to describe complex protein-protein interaction and to consider spatial aspects. I will underline the expressiveness of React(C) by scatching an encoding from the stochastic pi-calculus to React(C). I will also report on the price of this expressiveness in terms of complexity of simulation algorithms.

3.16 Understanding gene function by analysis of large scale molecular interaction network model behaviour

Andrzej M. Kierzek (University of Surrey, UK)

Understanding of how gene expression determines phenotype of the living cell under particular environmental conditions is one of the paramount goals of basic science. Prediction of the effects of gene activity perturbation on the behaviour of the living system is prerequisite for personalised molecular medicine and rational engineering in synthetic biology. Decades of the research in molecular biology resulted in the vast corpus of scientific articles on individual interactions between the molecular components in the living cells. The literature knowledge can be used to reconstruct molecular interaction networks and perform computational analysis of the cellular behaviour, where the gene function is analysed in the context of the global molecular machinery of the cell. Due to the time scale separation between gene regulation and metabolism it is useful to study linear models of the flux distribution in the metabolic networks. The genome scale metabolic reaction networks can be explored by linear programming to identify genes essential for different metabolic activities. The gene-metabolic phenotype relationship determined in this way can be further exploited to analyse transcriptomics data. Our recent publication (Bonde et al. PLoS Computational Biology, in press) shows applications to understanding of metabolic vulnerabilities of Mycobacterium tuberculosis. Despite the success of the genome scale metabolic network analysis, these models allow investigation of only about 1000 genes in the cell. Majority of the genes participate in processes other than metabolism that cannot be modelled in quasi-steady state framework. Recently, it has been demonstrated that useful approximation of large scale systems dynamics can be obtained by sampling token game trajectories in the Petri-net

representation of the molecular interaction networks. I will show preliminary results on how this approach can be integrated with flux balance analysis of metabolism towards modelling of genome scale molecular interaction networks capable of predicting function of majority of the genes in the cell.

3.17 Proving stabilization of large-scale biological systems

Elzbieta Krepska (Vrije Universiteit – Amsterdam, NL)

Main reference B. Cook, J. Fisher, E. Krepska, N. Piterman, "Proving stabilization of biological systems," Proc. 12th Verification, Model Checking and Abstract Interpretation Conference (VMCAI'11), LNCS vol. 6538, pp. 134–149, 2011.
URL http://dx.doi.org/10.1007/978-3-642-18275-4_11

We describe an efficient procedure for proving stabilization of biological systems modeled as qualitative networks or genetic regulatory networks. For scalability, our procedure uses modular proof techniques, where state-space exploration is applied only locally to small pieces of the system rather than the entire system as a whole. Our procedure exploits the observation that, in practice, the form of modular proofs can be restricted to a very limited set. For completeness, our technique falls back on a non-compositional counterexample search. Using our new procedure, we have solved a number of challenging published examples, including: a 3-D model of the mammalian epidermis; a model of metabolic networks operating in type-2 diabetes; a model of fate determination of vulval precursor cells in the C.elegans worm; and a model of pair-rule regulation during segmentation in the Drosophila embryo. Our results show many orders of magnitude speedup in cases where previous stabilization proving techniques were known to succeed, and new results in cases where tools had previously failed.

3.18 Modeling the community effect in development

Céline Kuttler (Université de Lille I, FR)

Main reference Yasushi Saka, Cedric Lhoussaine, Céline Kuttler, Ekkehard Ullner and Marco Thiel, "Theoretical basis of the community effect in development," BMC Systems Biology 2011, 5:54
URL http://dx.doi.org/10.1186/1752-0509-5-54

Background: Genetically identical cells often show significant variation in gene expression profile and behaviour even in the same physiological condition. Notably, embryonic cells destined to the same tissue maintain a uniform transcriptional regulatory state and form a homogeneous cell group. One mechanism to keep the homogeneity within embryonic tissues is the so-called community effect in animal development. The community effect is an interaction among a group of many nearby precursor cells, and is necessary for them to maintain tissue-specific gene expression and differentiate in a coordinated manner. Although it has been shown that the cell-cell communication by a diffusible factor plays a crucial role, it is not immediately obvious why a community effect needs many cells.

Results: In this work, we propose a model of the community effect in development, which consists in a linear gene cascade and cell-cell communication. We examined the properties of the model theoretically using a combination of stochastic and deterministic

modelling methods. We have derived the analytical formula for the threshold size of a cell population that is necessary for a community effect, which is in good agreement with stochastic simulation results.

Conclusion: Our theoretical analysis indicates that a simple model with a linear gene cascade and cell-cell communication is sufficient to reproduce the community effect in development. The model explains why a community needs many cells. It suggests that the community's long-term behaviour is independent of the initial induction level, although the initiation of a community effect requires a sufficient amount of inducing signal. The mechanism of the community effect revealed by our theoretical analysis is analogous to that of quorum sensing in bacteria. The community effect may underlie the size control in animal development and also the genesis of autosomal dominant diseases including tumorigenesis.

3.19 Colored Petri nets for modeling and analyzing biological systems

Fei Liu (BTU Cottbus, DE)

Petri nets are especially suitable for representing and modeling the concurrent, asynchronous, and dynamic behavior of biological systems. However, standard Petri nets do not scale, so they are restrained to modeling small systems. Therefore, we introduce the application of colored Petri nets in Systems Biology, which provide a possibility to model complex biological systems in a compact and scalable way.

In this talk, we will first give the motivation to use colored Petri nets for modeling and analyzing biological systems. Then we will describe some scenarios where colored Petri nets can contribute to Systems Biology, e.g. systems with repetitive components. Finally, we will introduce our colored Petri net modeling tool, Snoopy, by briefly describing its functionalities and features, especially for biological purpose.

References

1 F. Liu and M. Heiner. Colored petri nets to model and simulate biological systems. In Int. Workshop on Biological Processes and Petri Nets (BioPPN), satellite event of Petri Nets 2010, June 2010.
2 F. Liu and M. Heiner. Computation of enabled transition instances for colored petri nets. In Proc. 17th German Workshop on Algorithms and Tools for Petri Nets (AWPN 2010), volume 643 of CEUR Workshop Proceedings, pages 5–65. CEUR-WS.org, October 2010.
3 C. Rohr, W. Marwan, and M. Heiner. Snoopy - a unifying petri net framework to investigate biomolecular networks. Bioinformatics, 26(7):974–975, 2010.

3.20 Automatic Reconstruction of Extended Petri Nets from Experimental Data

Wolfgang Marwan (Universität Magdeburg – IBIO, DE)

Network inference methods reconstruct mathematical models of molecular or genetic networks directly from experimental data sets. A previously reported mathematical method [1] deliveres all possible alternative minimal networks that are consistent with a given discrete time series data set. In combination with answer set programming the approach is computationally highly efficient [2]. The original method is exact as it does not involve any heuristic interaction with the reconstruction process, however it produces only simple place/transition Petri nets. We refined the previously published algorithm to consider catalysis and inhibition of the reactions that occur in the underlying network. The results of the reconstruction algorithm are encoded in the form of an extended Petri net involving control arcs. This allows the consideration of processes involving mass flow and/or regulatory interactions. As a non-trivial test case, the phosphate regulatory network of enterobacteria was reconstructed using in silico-generated time-series data sets on wild-type and in silico mutants. The new exact algorithm reconstructs extended Petri nets from time series data sets by finding all alternative minimal networks that are consistent with the data. It suggested biochemically meaningful alternative molecular mechanisms for certain reactions in the network. The algorithm is useful to combine data from wild-type and mutant cells and may potentially integrate physiological, biochemical, pharmacological, and genetic data in the form of a single model.

References

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

2 Ostrowski, M., Schaub, T., Durzinsky, M., Marwan, W., Wagler, A.: Automatic network reconstruction using ASP. Theory and Practice of Logic Programming 2011, in press.

3.21 Multi-Level Rule Schemas

Carsten Maus (Universität Rostock, DE)

Observations of biological systems behaviors are performed at different levels of organization, i.e. proteins, compartments, individual cells, and cell populations. In order to capture all relevant dynamics and to structure a system accordingly, one might need to combine different levels when modeling a biological system. Formal languages explicitly supporting multi-level modeling facilitate the description of such models. In this talk, I will present the idea of a formal rule- based language for modeling biological systems at multiple levels. Our approach allows to structure models in a nested hierarchical manner and to let the different levels influence each other via upward and downward causation, while its rule-based modeling metaphor close to the notation of chemical reactions makes it easy to understand the syntax and keeps models small and readable.

3.22 Elementary, or not?

Ivan Mura (Microsoft Research – University Trento, IT)

This presentation discusses about the exactness of the Stochastic Simulation Algorithm proposed by Gillespie for the simulation of biochemical systems. It focuses on the accuracy issues that modelers may be confronted to when playing with the level of abstraction. In particular, we show that the core of Gillespie's formulation, i.e. the negative exponential distribution of event reaction times, sets a precise amount of molecular noise in the models, which not necessarily corresponds to the one existing in the modeled system. The propagation of this noise may affect stochastic model results in hardly predictable ways.

3.23 Utilizing Stochastic Model Checking to Determine the Robustness of Genetic Circuits

Chris J. Myers (University of Utah – Salt Lake City, US)

When designing and analyzing genetic circuits, researchers are often interested in the probability of the system reaching a given state within a certain amount of time. Usually, this involves simulating the system to produce some time series data and analyzing this data to discern the state probabilities. However, as the complexity of models of genetic circuits grow, it becomes more difficult for researchers to reason about the different states by looking only at time series simulation results of the models. To address this problem, this talk describes how to use stochastic model checking, a method for determining the likelihood that certain events occur in a system, with continuous stochastic logic (CSL) properties to obtain similar results. This goal is accomplished by the introduction of a methodology for converting a genetic circuit model (GCM) into a labeled Petri net (LPN). The state space of the LPN is then computed, and the resulting continuous-time Markov chain (CTMC) is analyzed using transient Markov chain analysis to determine the likelihood that the circuit satisfies a given CSL property in a finite amount of time. This talk illustrates a use of this methodology to determine the likelihood of failure in a genetic toggle switch and other circuits, and it compares these results to stochastic simulation-based analysis of the same circuits. Our results show that this method results in a substantial speedup as compared with conventional simulation-based approaches.

3.24 What does it cost to be flexible? A constraint based approach to modelling a micro-organism in a changeable environment

Brett Olivier (Free University – Amsterdam, NL)

Constraint based modelling is a widely used methodology used to analyse and study biological networks on both a small and whole organism (genome) scale. Typically these models are underdetermined and constraint based methods (e.g. linear, quadratic optimization) are used to optimise specific model properties. This is assumed to occur under a defined set of constraints (e.g. stoichiometric, metabolic) and bounds (e.g. thermodynamic, experimental and environmental) on the values that the solution fluxes can obtain.

Perhaps the most well known (and widely used) analysis method is Flux Balance Analysis (FBA) where for a model a target flux is maximised (typically a flux to biomass) where the other input/output fluxes have been bound to simulate a single set of defined environmental conditions. However, in the wild, such an organism may experience continuous changes in state that arise from sources either externally (e.g. a change in nutrient supply) or internally such as a mutation (deletion) in a particular gene which leads to a concomitant loss of (or large change in) cellular function.

In this presentation I will be discussing how we are attempting to extend established constraint based approaches to include micro-organisms living in a changeable environment. The question of what an organism can do in order to become more (or less flexible) in such an environment has necessitated the development of new theory, models, software tools and even a proposed standard for model exchange.

3.25 Rule-based modeling and application to biomolecular networks

Alessandro Romanel (ENS – Paris, FR)

Modelers of molecular signaling networks must cope with the combinatorial explosion of protein states generated by post-translational modifications and complex formation. Rule-based models provide a powerful alternative to approaches that require an explicit enumeration of all possible molecular species of a system. Such models consist of formal rules stipulating the (partial) contexts for specific protein-protein interactions to occur. These contexts specify molecular patterns that are usually less detailed than molecular species. Yet, the execution of rule-based dynamics requires stochastic simulation, which can be very costly. We briefly introduce some recent results on a formal abstract interpretation-based method to convert a rule-based model into a reduced system of differential equations and highlight actual research directions.

3.26 Abstractions in Spatial Simulation in Cell Biology

Adelinde M. Uhrmacher (Universität Rostock, DE)

In modelling and simulation of cell biological processes, spatial homogeneity in the distribution of components is a common but not always valid assumption. Thus, more and more spatial simulation algorithms are developed. To keep the calculation costs at bay some trade accuracy for execution speed, others combine different algorithms. The talk discusses two approaches most recently developed at our laboratory and the abstractions they are based upon.

3.27 Combining Bayesian and Frequentist parameter inference methods in systems biology

Natal van Riel (TU Eindhoven, NL)

Parameter estimation in systems biology models is in general an ill-posed inverse problem. Many different combinations of parameter values yield a model that can describe the data equally well. Non-identifiable model parameters hamper the development of predictive models. Strategies for parameterization that consider parameter identifiability are necessary. A strategy for model identification is proposed, based upon a combination of Monte Carlo Multiple Minimization, Profile Likelihood analysis and Monte Carlo Markov Chains. The approach can diagnose potential pitfalls regarding model parameterization. This is important information to have prior to a full Bayesian analysis. The analyses results mainly focus on model predictions rather than parameter values. The approach is applied to parameterize and analyze a model of the JAK-STAT signaling pathway.

3.28 Cell lineage dynamics analysis with individual regulatory networks

Bartek Wilczyński (University of Warsaw, PL)

Understanding the dynamics of proliferating cells in tissues is one of the key challenges for the modeling community. Currently, most widely used models assume large number of cells and use differential equations for modeling population sizes of different cell subtypes within tissue. Recently, a very intersecting work of Lander et al. [1] showed what types of negative feedback are preferred for the stability of the regeneration of an epithelium consisting of cells at 3 different stages of specification. Their results are supported by experiments indicating that changing concentration of certain gene products affects proliferation and specification rates in cell cultures.

The question we are asking is whether the behavior predicted and observed results in study can be reproduced with a model working on the level of gene regulation. We know that the gene regulation is the process underlying cell specification, but current models of

gene regulatory networks usually focus on the autonomous system within one cell and are not applicable to cell population analysis. However, while typical Boolean network models are more useful in stable state analysis, stochastic modeling frameworks, such as Stochastic Logical Networks [2], allow for simulation of multiple regulatory networks with adjustable transition rates.

In this work, we propose a regulatory network capable of representing the cell lineage system described by Lander et al. [1] with all its key components: linear progression through 3 different stages with cell death and simple signaling. We show how this system can be simulated in the SLN framework leading to a Continuous Time Markov Chain (CTMC) model specified by very few parameters and able to reproduce typical trajectories of the original dynamical model.

References

1 A.D. Lander, K.K. Gokoffski, FY Wan, Q. Nie, and A.L. Calof. Cell lineages and the logic of proliferative control. *PLoS Biol*, 7(1):e1000015, 2009.

2 B. Wilczyński and J. Tiuryn. Regulatory network reconstruction using stochastic logical networks. In *Computational Methods in Systems Biology*, pages 142–154. Springer, 2006.

3.29 Inference and Stability of Stochastic Models in Systems Biology

Verena Wolf (Universität des Saarlandes, DE)

Main reference A. Andreychenko, L. Mikeev, D. Spieler, and V. Wolf, "Parameter Identification for Markov Models of Biochemical Reactions," Proc. 23rd International Conference on Computer Aided Verification (CAV'11), LNCS vol. 6806, pp. 83–98, 2011.
URL http://dx.doi.org/10.1007/978-3-642-22110-1_8

In this talk I discuss two common problems of stochastic models in Systems Biology. The first problem is the inference of stochastic reaction rate constants from noisy observations at certain arbitrarily spaced observation time intervals. I will present a numerical method for the estimation of the rate constants based on the maximum likelihood method. The second problem is the stability analysis of stochastic models and in particular the analysis of multistable models. I will explain why deterministic approximations are not useful for studying stability properties and suggest two methods to approximate the equilibrium distribution of the stochastic model. The first method is based on drift arguments while the second one relies on a partial fluid approximation.

4 Working Groups

4.1 Team 1: DICTYPAT

Mary Ann Blätke, Qian Gao, David Gilbert, Simon Hardy, Monika Heiner, Andrzej Kierzek, Fei Liu, Wolfgang Marwan

This talk project is an exercise in modelling multiple spatial scales via the dynamics of cooperative biological entities, using as a test case pattern formation in Dictyostelium discoideum.

We developed a generic approach to support the systematic modelling of multiscale biological systems by the use of colour in Petri nets, which promises to be particularly helpful in investigating spatial aspects of the behaviour of biochemical systems, such as communication and behaviour at the intra and intercellular levels.

We represent space using a regular grid and developed a generic network pattern expressed as a coloured Petri net which can be easily configured for one, two or three dimensional grid scenarios of varying grid size. The pattern clearly separates intra-cellular and inter-cellular behaviour to allow for refinements of the pattern component which corresponds to specific cell behaviour.

We have validated our approach by performing some initial computational experiments. However, further experiments are required due to their computational expense which is beyond what can be done within the time limits of this workshop.

4.2 Team 2: A bistable gene switch

Andrea Bracciali, Mostafa Herajy, Pietro Lio, Chris J. Myers, Brett Olivier, Natal van Riel

Our talk considered the model a bistable gene switch proposed by Mehra et al. [1]. Utilizing our combined expertise, we attempted to model this system using a variety of methods including ODE's using Sundials CVode, Matlab, and PySCes, stochastic methods using iBioSim, generalized hybrid Petri nets using Snoopy, and stochastic process algebra using BioPEPA. Our initial attempts were hampered by the lack of a critical parameter, the basal rate of production for the activator. ODE simulation was used though to determine the effect of varying this parameter, and a reasonable value was determined. ODE simulation was further utilized to show the effect of the external concentration control signal as well as the dynamical behavior of the switch (noting, for instance that when the control signal is removed, the system switches back to the initial state). ODE analysis was utilized to analyze the effect of the growth rate on stability of the switch which was also found to be very critical. In addition to reproduction of the paper's results, we developed alternative models using Snoopy, Bio-PEPA, and iBioSim. Using iBioSim, we were able to show how automatic abstraction can reduce stochastic simulation time substantially (i.e., from 6 minutes to 0.6 seconds for 100 SSA runs). Using stochastic simulation of this abstracted model, we are able to show the probability of switching over time for different initial conditions and parameter values. We also considered an alternative model in which the activator does not need to

bind to the repressor, and we demonstrated that this model has a much sharper switching behavior. Finally, we constructed a population model for visualization purposes, showing how a multitude of individuals behave coherently with different assumptions on on the concentration of the control signal. Overall, this effort not only reproduced previous results, but it also produced new analyses leading to new insights into this system. We hope to further this analysis in the future.

References

1 Mehra S, Charaniya S, Takano E, Hu W-S (2008) A Bistable Gene Switch for Antibiotic Biosynthesis: The Butyrolactone Regulon in Streptomyces coelicolor. PLoS ONE 3(7): e2724.

4.3 Team 4: Simulating coupled gene expression oscillations of cell populations during somitogenesis using delays

Antje Beyer, Kirill Batmanov, Matthias Jeschke, Carsten Maus

Somitogenesis is the process of formation of somites, the segmented blocks of mesoderm which will eventually differentiate into other tissues.

We first reproduced the results reported in [1] using the open-source mathematics software SAGE and an additional package for solving delay differential equations (DDEs). To test whether intrinsic noise plays a significant role in maintaining sustained oscillations and the synchronization between cells, we simulated the model stochastically using two different delay representations: (i) a certain number of consecutive exponentially distributed events and (ii) fixed intervals. A stochastic simulation essentially exhibits the same behaviour as the system of DDEs, but additionally shows synchronization between coupled cells in an extended three-cell model.

A new rule-based multi-level language [2] was applied to integrate the non-elementary rate laws into a coupled multi-cellular model. First results showed a weak performance of the prototype simulator in case of larger cell populations. To study the behavior of many cells with local coupling, we implemented the delay stochastic simulation algorithm described in [3]. Simulations show that despite local differences in oscillations the coupling is able to synchronize the cells globally in 1- and 2-dimensional configurations.

A hybrid model was built by integrating the stochastic simulation algorithm into a physical model of cell movement based on previous work [4] using the Molecular Dynamics framework [5]. We were able to show how cell divisions cause a movement of the PSM. As cells dropped out of PSM to become somites, they were arrested at different gene levels resulting in a distinct pattern formation.

References

1 Lewis, J. *Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator.* Cur. Biol. 13, pp. 1398–1408, 2003.
2 Maus, C. et al. *Towards Rule-Based Multi-Level Modeling.* Poster, ICSB'10, 2010.
3 Schlicht R, Winkler G. *A delay stochastic process with applications in molecular biology.*. J. Math. Biol., 57(5), pp. 613–48, 2008.
4 Beyer, A. et al. *A Dynamic Physical Model of Cell Migration, Differentiation and Apoptosis in Caenorhabditis elegans.* Proc. of the 11th ICSB, Springer, 2011 (under review).

5 Alder, B.J., Wainwright, T.E. *Studies in Molecular Dynamics. I. General Method.* J. Chem. Phys., 31(2), p. 459, 1959.

4.4 Team 5: Counter-intuitive stochastic behavior of simple gene circuits with negative feedback

Ivan Mura, Davide Chiarugi, Maria Luisa Guerriero, Chiara Bodei, Luca Bortolussi, Alessandro Romanel, Jane Hillston, Alberto Policriti

Results: Gene expression is a fundamentally stochastic process with randomness in transcription and translation. Usually, stochastic modelling techniques used in gene expression are based on the Gillespie approach, where the role of parameters and the level of modelling abstraction are essential. Even in a simple model of gene regulation different combinations of parameter sets and abstraction levels can lead to completely different dynamics. Moreover, these combinations can have a crucial impact on the computational feasibility of the Gillespie approach. We have considered a simple model of gene regulation where a transcriptional repressor negatively regulates its own expression and we have investigated to what extent hybrid approaches and QSSA can be useful. In particular, the characteristic dynamics emerging from the randomness in transcription and translation can be captured also by hybrid approaches and we have shown that for different biological realistic parameter sets, different combinations of hybrid approaches and QSSA cannot only preserve the stochastic dynamics, but also speed up simulation times.

Conclusions: We have shown that both mRNA and protein degradation play a role in noise control and that, in general, there can be multiple control points in feedback loops.

In particular, no single parameter's variation is responsible for feedback loop intensity but a (linear? non-linear?) combination of them.

References

1 Tatiana T. Marquez-Lago & Jörg Stelling: Counter-intuitive stochastic behavior of simple gene circuits with negative feedback. Biophys. J. (2010) 98: 1742–1750.

Participants

- Kirill Batmanov
Université de Lille I, FR
- Giampaolo Bella
Universita di Catania, IT
- Antje Beyer
University of Cambridge, UK
- Mary Ann Blätke
Univ. Magdeburg – IBIO, DE
- Chiara Bodei
University of Pisa, IT
- Luca Bortolussi
Universitá di Trieste, IT
- Andrea Bracciali
University of Stirling, UK
- Rainer Breitling
University of Glasgow, UK
- Davide Chiarugi
University of Siena, IT
- Maciej Dobrzyński
University College – Dublin, IE
- Anna Gambin
University of Warsaw, PL
- David Gilbert
Brunel University, UK
- Maria Luisa Guerriero
University of Edinburgh, UK
- Simon Hardy
Mount Sinai Medical School – New York, US
- Monika Heiner
BTU Cottbus, DE
- Mostafa Herajy
BTU Cottbus, DE
- Jane Hillston
University of Edinburgh, UK
- Matthias Jeschke
Heidelberg, DE
- Mathias John
IRI - Villeneuve d'Ascq, FR
- Andrzej M. Kierzek
University of Surrey, UK
- Elzbieta Krepska
Vrije Univ. – Amsterdam, NL
- Céline Kuttler
Université de Lille I, FR
- Pietro Lió
University of Cambridge, UK
- Fei Liu
BTU Cottbus, DE
- Wolfgang Marwan
Univ. Magdeburg – IBIO, DE
- Carsten Maus
Universität Rostock, DE
- Ivan Mura
Microsoft Research – University Trento, IT
- Chris J. Myers
University of Utah - Salt Lake City, US
- Brett Olivier
Free Univ. – Amsterdam, NL
- Alberto Policriti
Universitá di Udine, IT
- Alessandro Romanel
ENS – Paris, FR
- Adelinde M. Uhrmacher
Universität Rostock, DE
- Natal van Riel
TU Eindhoven, NL
- Bartek Wilczyński
University of Warsaw, PL
- Verena Wolf
Universität des Saarlandes, DE