Report from Dagstuhl Seminar 18082

Formal Methods for the Synthesis of Biomolecular Circuits

Edited by

Yaakov Benenson[1], Neil Dalchau[2], Heinz Koeppl[3], and Oded Maler[4]

1 **ETH Zürich – Basel, CH,** kobi.benenson@bsse.ethz.ch
2 **Microsoft Research UK – Cambridge, GB,** ndalchau@microsoft.com
3 **TU Darmstadt, DE,** heinz.koeppl@bcs.tu-darmstadt.de
4 **VERIMAG – Grenoble, FR,** oded.maler@imag.fr

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 18082 "Formal Methods for the Synthesis of Biomolecular Circuits". Synthetic biology aims for the rational bottom-up engineering of new biological functionalities. Recent years have witnessed an increase in the degree of "rationality" in the design of synthetic biomolecular circuits. With it, fewer design-build-test cycles were necessary to achieve a desired circuit performance. Most of these success stories reported the realization of logic circuits, typically operating via regulation of gene expression and/or direct manipulation of DNA sequences with recombinases, executing combinatorial and sometimes sequential logic. This was often achieved with the help of two ingredients, a library of previously well-characterized parts and some computational modeling. Hence, although circuits in synthetic biology are still by far less understood and characterized than electronic circuits, the opportunity for the formal synthesis of circuit designs with respect to a behavioral specification starts to emerge in synthetic biology.

Seminar February 18–23, 2018 – https://www.dagstuhl.de/18082
2012 ACM Subject Classification Theory of computation → Verification by model checking, Theory of computation → Verification by model checking, Theory of computation → Concurrency, Computing methodologies → Machine learning, Mathematics of computing → Probability and statistics, Software and its engineering → Specification languages
Keywords and phrases Synthetic biology, Electronic design automation, Program synthesis and verification
Digital Object Identifier 10.4230/DagRep.8.2.88

1 Executive Summary

Heinz Koeppl

The seminar brought together experts in formal methods for the verification and synthesis of hardware and software with wet-lab and dry-lab synthetic biologists to (1) achieve a common understanding of the current state of design methodology in synthetic biology; (2) to identify the limitations of current approaches and (3) to investigate dedicated solutions to the synthesis problem in synthetic biology. Some of these methods are based on leveraging experience and methods from electronic design automation (EDA) and from program synthesis and verification. In addition, ideas for entirely new methodologies specifically tailored for synthetic biology are likely to emerge. For example, features that are not apparent in

electronic circuits such as heterogeneity and variability between the cells and between the circuits embedded in different cells, were addressed.

Apart from talk by participants, the seminar also featured break out session that were well received by the participants. In particular, we had sessions on "Modeling context-dependency of synthetic circuits" on "Metrology in Synthetic Biology" and on "Formal Specification for Biological Circuit Synthesis".

2 Table of Contents

Executive Summary
Heinz Koeppl . . . 88

Overview of Talks

How AI can help synthetic biology
Aron Adler . . . 92

Bayesian approaches to engineering synthetic biological systems
Chris Barnes . . . 92

Foundational Metrology for Engineering Biomolecular Circuits
Jacob Beal . . . 92

Syntax-Guided Optimal Synthesis for Chemical Reaction Networks
Milan Ceska . . . 93

Programming DNA circuits
Neil Dalchau . . . 93

Automated Reasoning in Stem Cell Biology
Sara-Jane Dunn . . . 94

A compiler of computable real-valued functions in abstract biochemical reaction networks
François Fages . . . 94

A critical view on Synthetic Biology (by an outsider!)
Eric Fanchon . . . 94

Exploring and verifying complex genetic circuit designs and design spaces using deep-sequencing
Thomas Gorochowski . . . 95

Three synthetic biology design challenges we face, and how we are approaching them
Nathan Hillson . . . 95

CloneFlow – Computer-aided Planing of DNA Assembly Reactions and Experimental Workflows
Johannes Kabisch . . . 95

Towards flexible, and data-driven construction and analysis of dynamical models for synthetic biology
Gareth Molyneux . . . 96

An executable software model of tetracycline-aptamer-mediated translation in yeast
Radu Muschevici . . . 96

Design of Asynchronous Genetic Circuits
Chris Myers . . . 97

Model-based multiobjective optimization framework for automated design in Synthetic Biology
Irene Otero-Muras . . . 97

Visual Representations for CRN Synthesis
James Scott-Brown 98

Synthesis of Biomolecular Circuits with Z3
Boyan Yordanov 98

SMT-based Set Synthesis for Biological Models
Paolo Zuliani 98

Working groups 99

Participants 100

3 Overview of Talks

3.1 How AI can help synthetic biology

Aron Adler (BBN Technologies – Cambridge, US)

A wide variety of Artificial Intelligence (AI) techniques, from expert systems to machine learning to robotics, are needed in the field of synthetic biology. This talk will look back at progress in recent years and highlight places where AI has already helped and the ongoing opportunities for applying the lessons from decades of AI research to problems in synthetic biology.

3.2 Bayesian approaches to engineering synthetic biological systems

Chris Barnes (University College London, London, UK)

An aim of synthetic biology is to apply engineering tools and principles to the design and construction of novel biological systems. There is huge potential for clinical applications but for such advanced therapeutics to be implemented, we must first be able to design and build systems that can function reliably in complex and changing environments. Using the example of engineering a two-species bacterial community, I will describe how Bayesian statistics and dynamical modelling can be used at different points within the design-build-test cycle.

3.3 Foundational Metrology for Engineering Biomolecular Circuits

Jacob Beal (BBN Technologies – Cambridge, US)

People have been running experiments to characterize biological circuit components ever since these notions were first conceived. Such data, however, is very rarely actually able to used effectively for circuit design by any besides its creators and often not even by them. I will argue that this is largely due to inadequacies in how circuits and components are typically measured. This discussion will include process vs. experimental controls, statistical analysis of biomolecular circuits, reliable unit calibration, and precision requirements. Motivating examples and evidence will be drawn from my experience with precision prediction of biological circuit behavior, development of high-performance biological computing devices, and large scale interlaboratory studies with iGEM and the NIST Synthetic Biology Standards Consortium.

3.4 Syntax-Guided Optimal Synthesis for Chemical Reaction Networks

Milan Ceska (Brno University of Technology, CZ)

We study the problem of optimal syntax-guided synthesis of stochastic Chemical Reaction Networks (CRNs) that plays a fundamental role in design automation of molecular devices and in the construction of predictive biochemical models. We propose a sketching language for CRNs that concisely captures syntactic constraints on the network topology and allows its under-specification. Given a sketch, a correctness specification, and a cost function defined over the CRN syntax, our goal is to find a CRN that simultaneously meets the constraints, satisfies the specification and minimizes the cost function. To ensure computational feasibility of the synthesis process, we employ the Linear Noise Approximation allowing us to encode the synthesis problem as a satisfiability modulo theories problem over a set of parametric Ordinary Differential Equations (ODEs). We design and implement a novel algorithm for the optimal synthesis of CRNs that employs almost complete refutation procedure for SMT over reals and ODEs, and exploits a meta-sketching abstraction controlling the search strategy. Through relevant case studies we demonstrate that our approach significantly improves the capability of existing methods for synthesis of biochemical systems and paves the way towards their automated and provably-correct design.

3.5 Programming DNA circuits

Neil Dalchau (Microsoft Research UK – Cambridge, GB)

Biological organisms use complex molecular networks to navigate their environment and regulate their internal state. The development of synthetic systems with similar capabilities could lead to applications such as smart therapeutics or fabrication methods based on self-organization. To achieve this, molecular control circuits need to be engineered to perform integrated sensing, computation and actuation. In this talk, I will describe an approach based on DNA hybridization and strand displacement to implement the computational core of such control circuits. We use domain-specific programming languages to specify the sequence-level circuit design, which compile to chemical reaction networks, a well-established formalism for describing and simulating chemistry. Furthermore, we have integrated parameter inference techniques into this design platform, which facilitates design-build-test-learn cycles via model-based characterization and circuit prediction. A first example will introduce how we designed and constructed a DNA implementation of the approximate majority algorithm, which seeks to establish consensus in a population of agents (molecules). A second example will illustrate how DNA circuits can be considerably accelerated by tethering DNA hairpin molecules to a fixed template, overcoming molecular diffusion as a rate-limiting step.

3.6 Automated Reasoning in Stem Cell Biology

Sara-Jane Dunn (Microsoft Research UK – Cambridge, GB)

A fundamental goal in developmental biology is to understand the logic of cellular decision-making at the molecular level. However, the complexity of biological processes presents a major challenge when trying to delineate fate decisions, which are typically influenced by a multiplicity of extrinsic and intrinsic regulators displaying non-intuitive interactions. Against this backdrop, automated reasoning is a powerful methodology that can allow researchers to navigate biological complexity and derive explanations of behaviour that are provably consistent with experimental evidence. In this talk, I will discuss the synthesis and analysis of dynamic networks of biological components that govern decision-making in embryonic stem cells. I will demonstrate how this approach can be used to derive a predictive explanation of cellular behaviour, generating counterintuitive and informing future experiments.

3.7 A compiler of computable real-valued functions in abstract biochemical reaction networks

François Fages (INRIA Saclay – Île-de-France, FR)

I shall present a compiler of real-valued functions in continuous CRNs based on a recent proof of their Turing completeness. More specifically, our compiler, implemented in BIOCHAM v4, takes as input a polynomial differential equation system and produces as output a finite reaction network which implements it with at most binary reactions and a fixed set of molecular species with positive concentration values. Then I will discuss the issues of error control, robustness measure a posteriori, robustness control a priori, and concrete implementations with real enzymes in DNA-free vesicles in partnership with Franck Molina's lab Sys2diag.

3.8 A critical view on Synthetic Biology (by an outsider!)

Eric Fanchon (TINC-IMAG Lab – La Tronche, FR)

This talk is mainly about sharing some thoughts on synthetic biology and asking questions...

I will first discuss some conceptual/philosophical issues which influence the way the biological engineering problems are approached: the genome as a book or a program; the cell as a machine.

I will then mention some practical issues: the lack of data needed to build dynamical models, the dense web of molecular interactions which makes the composition of modules hard to realize, the large variability of biological systems. Regarding Formal Methods the biggest issue seems to be related to the interaction of the engineered system with the host cell. In view of the many unexpected interactions that might occur, how to prove safety?

3.9 Exploring and verifying complex genetic circuit designs and design spaces using deep-sequencing

Thomas Gorochowski (University of Bristol, GB)

As the size and complexity of genetic circuits grows (as well as our ambitions) their assembly and functional verification becomes increasingly difficult due to the sheer number of parts involved. I will show how we have been using deep-sequencing and one-pot pooled assembly methods to construct subsets of design spaces to make exploration of large design spaces quicker and easier. I will also introduce our efforts to move away from fluorescent reports as a means of probing internal circuit states and propose RNA-sequencing as a viable alternative that enables the entire transcriptional state of all components to be observed at a point in time. I will end by giving my personal perspective on how these threads might come together with new design methods to better understand the underlying principles for effective genetic circuit synthesis.

3.10 Three synthetic biology design challenges we face, and how we are approaching them

Nathan Hillson (JBEI – Emeryville, US)

After a brief introduction to the Agile BioFoundry and to the Joint Genome Institute (for context), three synthetic biology design challenges that we face, and how we are currently approaching them, will be presented. These challenges include: 1) how to reliably/predictably control protein levels, in context of statistical modelling design of experiments; 2) the design of (and Learning from) operon structure variants; and 3) informatic integration/continuity across layers of a hierarchical Build process.

3.11 CloneFlow – Computer-aided Planing of DNA Assembly Reactions and Experimental Workflows

Johannes Kabisch (TU Darmstadt, DE)

Progress in synthetic biology can be facilitated by rapid prototyping approaches to test a wide variety of, e.g., genetic circuits. In the last decades, many methods for assemblies were used to build DNA constructs (e.g., Gibson assembly, Golden Gate assembly). Among these methods ligase cycling reaction (LCR) fulfills many prerequisites for an easy application and automation like the usage of unmodified DNA parts and that the assembly order is only determined by single-stranded oligonucleotides building a bridge (so called bridging oligonucleotide, BO) between adjacent parts. In order to enable an easy access for scientist to apply LCR we are developing the web-application CloneFlow. CloneFlow offers to create LCR

workflows based on the input of DNA sequences in a multi-fasta format and experimental parameters. Different output formats are available including csv-files enabling easy downstream processing e.g. for automation with liquid-handling systems. In our case we feed such files into a nanoliter dispenser. As a unique feature, all oligonucleotides are secondary structure optimized, i. e. our service finds the oligos which have low free energy and thus a small contribution to secondary structures. Our work presents a synergistic collaboration between experimental expertise and computational know-how.

3.12 Towards flexible, and data-driven construction and analysis of dynamical models for synthetic biology

Gareth Molyneux (Oxford University, Oxford, UK)

We propose a technique that integrates Bayesian learning and model verification to quantify the likelihood that the underlying data-generating biological system satisfies a given dynamical property of interest. We extend an approach developed for diverse models and systems and adapt it to biological circuits, and specifically to continuous-time Markov Chain models of Chemical Reaction Networks. We argue that the approach is flexible and adaptable to time varying data, and that it extracts more information from gathered data than standard statistical techniques. From the perspective of synthesis, it can be used to build parts of the model of a genetic circuit, as well as to generate optimal experiments for model learning.

3.13 An executable software model of tetracycline-aptamer-mediated translation in yeast

Radu Muschevici (TU Darmstadt, DE)

Gene expression in yeast cells can be controlled by mediating translation through insertion of tetracyclin aptamers into the 5' UTR of the mRNA molecule. This mechanism has been shown to work well in practice, but its exact mode of action and the regulatory parameters needed to fine-tune the rate of protein synthesis are not fully understood. We formalize the underlying bio-chemical processes in a concurrent, executable modelling language. Using object orientation and asynchronous communication yields a model that captures bio-chemical concepts in a natural manner at the level of domain experts. It can be made more fine-grained incrementally as needed. Our model is fully executable and, after calibration, can precisely simulate in vivo experiments. It can be used to predict and gain a deeper understanding of the effect of different assumptions about the working mechanisms during tetracycline-aptamer-mediated translation, such as the aptamer structure, the quantity of inserted aptamers and their position, tetracycline concentration, etc. Hence, the model can be used to design targeted experiments that test new theories. Simulations can be executed in a manner of minutes and it is easy to obtain different kinds of visualizations.

3.14 Design of Asynchronous Genetic Circuits

Chris Myers (University of Utah, US)

Researchers are now able to engineer synthetic genetic circuits for a range of applications in the environmental, medical, and energy domains. Crucial to the success of these efforts is the development of methods and tools for genetic design automation (GDA). While inspiration can be drawn from experiences with electronic design automation (EDA), design with a genetic material poses several challenges. In particular, genetic circuits are composed of very noisy components making their behavior more asynchronous, analog, and stochastic in nature. This talk presents our research in the development of the GDA tool, iBioSim, which leverages our past experiences in asynchronous circuit synthesis and formal verification to address these challenges. The iBioSim tool enables the synthetic biologist to construct models in a familiar graphical form, analyze them using a variety of methods that leverage efficient abstractions, visualize their analysis results using an intuitive interface, and ultimately synthesize a genetic implementation from a library of genetic parts. Each step of this design process utilizes standard data representation formats enabling the ready exchange of results.

3.15 Model-based multiobjective optimization framework for automated design in Synthetic Biology

Irene Otero-Muras (Spanish National Council for Scientific Research – Vigo, ES)

One of the challenges of Synthetic Biology is building genetic circuits with higher complexity, not only in terms of the number of regulatory regions involved, but also in the kind of tasks that these circuits can accomplish. I present a framework for automated design of biocircuits starting from libraries of standard parts that takes the following aspects into account: i) optimality, ii) several design criteria, iii) high computational efficiency. The method is based on optimization, and any target behaviour (adaptation, change fold detection, oscillations, pattern formation, bistability...) can be encoded in the set of objective functions. In this way can design, for example, synthetic circuits with optimal performance with respect to a given criterion, while keeping the protein cost at minimum, or oscillators with optimal tunability without compromising the stability of the limit cycle (i.e. with optimal robustness with respect to molecular noise). Our strategy exploits hybrid Mixed Integer Nonlinear Optimization Programming solvers, allowing to search simultaneously topology and parameter spaces. At present, the method relies on a deterministic description of the gene regulation dynamics. In a future work I plan to incorporate stochastic simulations exploiting a recently developed method for efficient simulation of stochastic gene regulatory networks, based on a Partial Integro-Differential Model approximation of the Chemical Master Equation.

3.16 Visual Representations for CRN Synthesis

James Scott-Brown (University of Oxford, GB)

In this talk I introduce TimeRails, a diagrammtic representation for formal specifications expressed in Signal Temporal Logic. This represents specifications using the visual metaphor of rectangles and rails, along which other rails or rectangles can slide. I describe how this is integrated into a tool for the automated synthesis of synthesis and verification of Chemical Reaction Networks (CRNSynth). This provides a graphical interface for the user to express constraints on the network structure of a parametric CRN and specifications describing its behaviour, which are translated into a set of ordinary differential equations and constraints. Together these form a satisfaction problem modulo the theory of ODEs that can be passed to an SMT-ODE solver such as iSAT-ODE or dReach.

3.17 Synthesis of Biomolecular Circuits with Z3

Boyan Yordanov (Microsoft Research UK – Cambridge, UK)

Z3 is an automated theorem prover that integrates specialized solvers for domains relevant to program analysis, testing and verification. At Microsoft, Z3 powers a variety of tools that reason about program states and transformations to improve the quality and security of software and services. Z3 also provides a powerful tool for reasoning about biological programs and biomolecular circuits, enabling the development of analysis and synthesis methods. In this talk, I will introduce our approach to encoding biological queries as Satisfiability Modulo Theories (SMT) problems that are solved using Z3. I will illustrate the approach with results from two projects involving (i) synthesizing biological interaction networks from prior knowledge and experimental data, and (ii) synthesizing chemical reaction networks (CRNs) from input/output specifications of desired computations.

3.18 SMT-based Set Synthesis for Biological Models

Paolo Zuliani (University of Newcastle, GB)

for which a given system model satisfies a desired behaviour. In this talk we present BioPSy, a tool that performs guaranteed parameter set synthesis for ordinary differential equation (ODE) biological models expressed in the Systems Biology Markup Language given a desired behaviour expressed by time-series data. Three key features of BioPSy are: 1) BioPSy computes parameter intervals, not just single values; 2) for the identified intervals the model is formally guaranteed to satisfy the desired behaviour; and 3) BioPSy can handle virtually any Lipschitz-continuous ODEs, including nonlinear ones. BioPSy is able to achieve guaranteed synthesis by utilising Satisfiability Modulo Theory (SMT) solvers to determine

acceptable parameter intervals. Furthermore, BioPSy can formally check parameter estimates generated by other (non-formal) methods. We have successfully applied our tool to several biological models including a prostate cancer therapy model, a human starvation model, and a cell cycle model.

4 Working groups

After initial presentations from Monday to Wednesday, we identified in a disussion session which topics are most relevant and pressing to advance the field of synthesis methods in synthetic biology. We finally voted on proposed topics and decided in the following three: (1) "Modeling Context-Dependency of Synthetic Circuits", (2) "Metrology in Synthetic Biology" and (3) "Formal Specification for Biological Circuit Synthesis". We then split up in working groups and discussed the respective topics in details. In order to consolidate the findings every group was asked to prepare a short presentation in the plenum on Friday morning. In the following we briefly summarize the findings.

For topic (1), the group converged after an initial discussion on the problem of weak termination as it was identified as a major source of context-dependency. In particular, the group investigated how terminators can structurally be optimized and whether biophysical folding models and machine learning can be combined to generate new terminator libraries that can be used in design automation.

For topic (2), a smaller working group considered the concrete problem of converting fluorescence data from flow cytometers into absolute copy number of fluorescent proteins by using commercially available reference beads. They used a reference dataset provided by participant Jake Beal.

For topic (3), the working group identified different classes of specifications for circuits, e.g., static versus dynamic properties, Boolean versus real-valued or whether a cost model is specified or not. Accordingly, three main dimensions for a specification were determined: From functional to temporal, from qualitative to quantitative, from robustness not specified to robustness specified. Furthermore a synthesis challenge was conceived and programming was performed throughout the whole evening. One had to design a two-input reaction system that oscillates between a NOR gate and OR gate in a pre-specified time period. The informal challenge was won by Francois Fages using his tool BIOCHAMP. A reaction system with 13 reactions was found semi-automatically that realized this function.

Participants

- Aaron Adler
BBN Technologies – Cambridge, US
- Ben Barak
Tel Aviv, IL
- Chris Barnes
University College London, GB
- Jacob Beal
BBN Technologies – Cambridge, US
- Yaakov Benenson
ETH Zürich – Basel, CH
- Milan Ceska
Brno University of Technology, CZ
- Neil Dalchau
Microsoft Research UK – Cambridge, GB
- Sara-Jane Dunn
Microsoft Research UK – Cambridge, GB
- François Fages
INRIA Saclay – Île-de-France, FR
- Eric Fanchon
TINC-IMAG Lab – La Tronche, FR
- Thomas Gorochowski
University of Bristol, GB
- Maleen Hanst
TU Darmstadt, DE
- Nathan Hillson
jibei – Emeryville, US
- Johannes Kabisch
TU Darmstadt, DE
- Heinz Koeppl
TU Darmstadt, DE
- Jan Madsen
Technical University of Denmark – Lyngby, DK
- Oded Maler
VERIMAG – Grenoble, FR
- Gareth Molyneux
University of Oxford, GB
- Radu Muschevici
TU Darmstadt, DE
- Chris J. Myers
University of Utah, US
- Irene Otero-Muras
CSIC – Vigo, ES
- James Scott-Brown
University of Oxford, GB
- Boyan Yordanov
Microsoft Research UK – Cambridge, GB
- Paolo Zuliani
University of Newcastle, GB