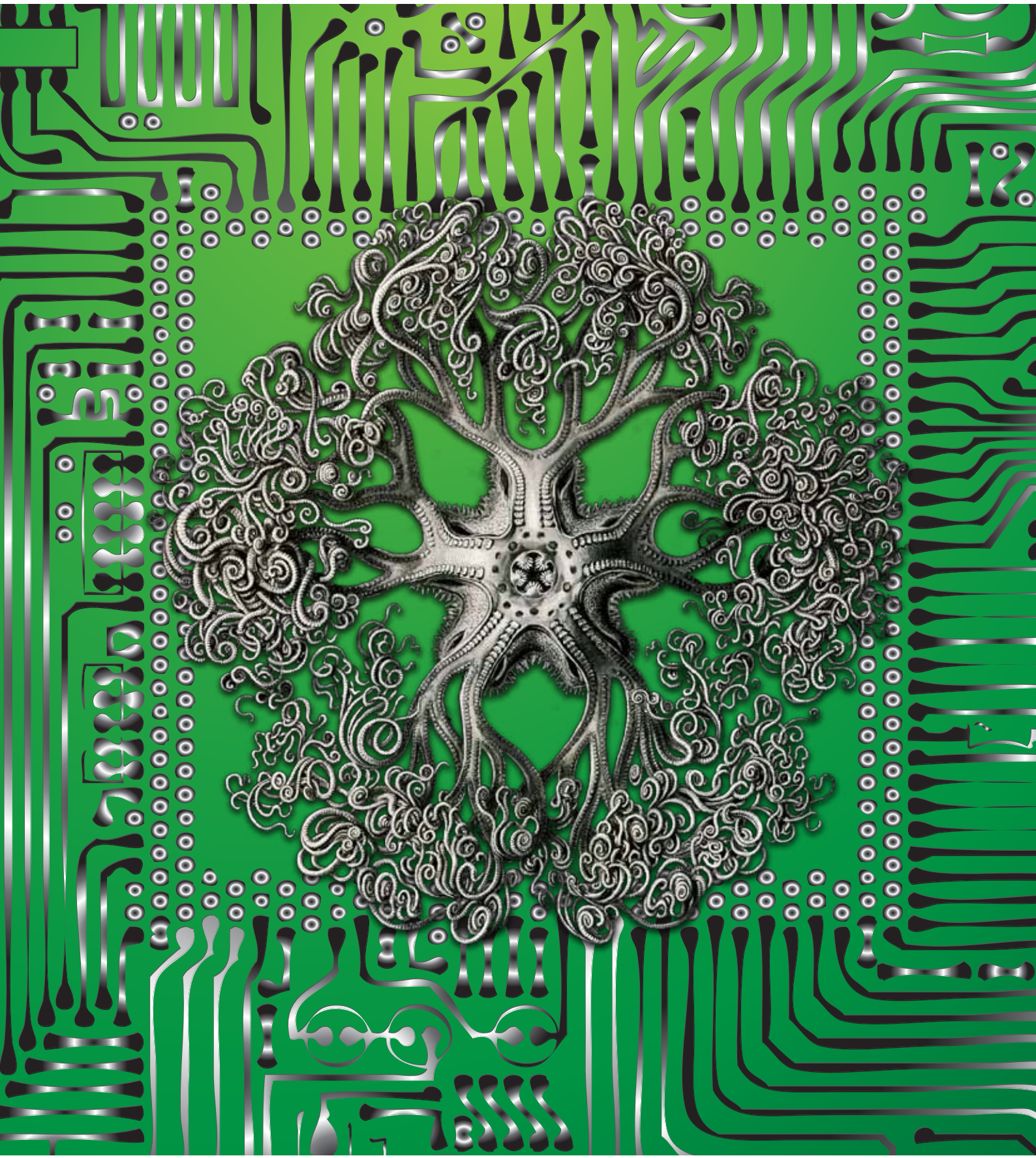
.toProB()

Toward Programmable Biology



York, UK - 20th July 2015

I(C)$_2$S

Published by:
ICOS Interdisciplinary Computing and Complex Biosystems Research Group
School of Computing Science
Newcastle University
http://programmable-biology.ico2s.org/

Cover design by Ana Vásquez Tapia

Toward Programmable Biology, Welcome

The discovery of the mechanisms underlying genetic regulation in the early 1960s sparked scientists with an interest about the general programmability of living matter. This abstract vision began turning into reality when modern day methods for genetic synthesis, amplification, and characterisation became available in the 1980s and 1990s. Based on these techniques and utilising an engineering mentality, the new millennium has seen a tidal wave of approaches aimed at modifying natural biological systems, gathered under the newly emerging discipline of Synthetic Biology.

Fifteen years later, the vision to repurpose living cells as substrates for general computation has manifested itself in genetic circuit designs that attempt to implement Boolean logic gates, digital memory, oscillators, and other circuits from electrical engineering. Yet, the various achievements in the realm of Synthetic Biology remain isolated and generally lack the modularity and scalability of their electronic counterparts.

This ECAL 2015 satellite workshop revisits cornerstone achievements in Synthetic Biology research that address general computability in biological substrates, in order to demarcate key in-vivo and in-silico challenges of this novel research area. Topics will deal with how paradigms borrowed from digital, electronic devices are best implemented in large-scale biological substrates, and whether unconventional computing paradigms, such as those developed in the research field of Artificial Life, might offer more promising routes toward full-fledged biological computation.

Harold Fellermann
Omer Markovitch
Curtis Madsen
Owen Gilfellon

Programme

9:00 - 10:00	Conference registration
10:00 - 10:20	**Harold Fellermann**
	Welcome & Introduction
10:20 - 10:50	**Richard Mayne & Andy Adamatzky**
	Collision-based computing implemented with calcium-containing vesicles in a myxomycete model 1
10:50 - 11:20	Coffee break
	Session chair: Curtis Madsen
11:20 - 12:05	**Martyn Amos** (keynote)
	Population-based microbial computing: a third wave of synthetic biology? 3
12:05 - 12:25	**Jonathan Naylor, Harold Fellermann & Natalio Krasnogor**
	Physical modelling exploring the self-organising behaviour of synthetic and natural biofilms 4
12:25 - 12:55	**Savas Konur, Marian Gheorghe & Omer Markovitch**
	Analysing synthetic bacteria colonies: an agent-based approach 6
13:00 - 14:15	Lunch break
	Session chair: Omer Markovitch
14:15 - 15:00	**Friedrich C. Simmel** (keynote)
	Programming molecular structures and processes using DNA 8
15:00 - 15:30	**Goksel Misirli, Jennifer Hallinan, Owen Gilfellon & Anil Wipat**
	A standard model repository for genetic design automation 9
15:30 - 16:00	Tea break
	Session chair: Owen Gilfellon
16:00 - 16:20	**Aurore Dupin & Friedrich C. Simmel**
	Biochemical circuits in cell-sized microcompartments 11
16:20 - 16:50	**Jerzy Kozyra, Chien-Yi Chang, Alessandro Ceccarelli, Harold Fellermann & Natalio Krasnogor**
	Programming synthetic scaffolds for DNA origami 12
16:50 - 17:20	**Matthaeus Schwarz-Schilling, Fabio Chizzolini, Andrea Mückl & Friedrich C. Simmel**
	Self-assembly of a functionalized RNA nanostructure 14

Collision-based computing implemented with calcium-containing vesicles in a myxomycete model

Richard Mayne & Andrew Adamatzky

Unconventional Computing Group, University of the West of England, Bristol, BS16 1QY, UK
Richard.Mayne@UWE.ac.uk

Abstract

Collision-based computing (CBC) is a form of unconventional computing in which travelling objects represent data and conditional routing of signals via object collisions determines the output state. We present here a brief introduction to our 'Vesicle Collision Model' (VCM) modification of Fredkin and Toffoli's seminal CBC Billiard Ball Model, in which we experimentally observe observe the rapid transport and subsequent collisions of Ca^{2+}-containing vesicles within the plasmodium of slime mould *Physarum polycephalum* and characterise these phenomena in the language of computation. This investigation reflects on the technical considerations to achieve full implementation of practical VCM circuits. The findings presented here represent the first intracellular realisation of CBC and hence are relevant to all fields of scientific inquiry where biology and computer science meet.

Introduction

Since their works of the late 1970's–early 1980's, Fredkin and Toffoli's treatises on physical models of computation (Fredkin and Toffoli (1982); Toffoli (1980)) have precipitated countless fascinating and profitable advances in the field of computer science, one of which is collision-based computing (CBC). Their billiard ball model (BBM) — in which computing is realised through the conditional routing of identical spheres ('signals') wherein collisions represent logical computation — is the archetypal model of CBC and the best example of a computationally universal, conservative computing paradigm. Inspired by Margolus' Soft Sphere Model (SSM) modification of the BBM (Margolus (2002)), in which non-rigid spheres deform and temporarily travel as a single entity following collision, we have developed an *in vivo* CBC model based on laboratory observations of calcium-containing vesicle collisions within our model organism, the plasmodium of myxomycete slime mould *Physarum polycephalum* (Mayne and Adamatzky (2015)). This investigation briefly delineates our 'Vesicle Collision Model' (VCM) of CBC and discusses the challenges and open problems for experimental implementation of practical *in vivo* collision-based circuits.

One of the major functions of intracellular vesicles is to distribute substances such as signalling molecules throughout the cell. To this end, they are coupled via motor proteins to the cellular cytoskeleton, an abundant, dynamical intracellular network of proteins with manifold functions, for rapid transit through the cell. Vesicle collisions that occur therein may therefore be said to occur on a three-dimensional nano-scale dynamical circuit.

Unconventional computation is an art of creative interpretation of nature: signalling molecules are an attractive model of biological 'information' because they represent signals whose delivery to a target cell/organelle is a quantifiable process due to their elicitation of reproducible, proportional responses. Calcium is an important intracelluar signalling molecule (Berridge (2004)) whose choice as 'molecule of interest' in this study was convenient due to its abundance within slime mould, but was essentially arbitrary.

Methods

Plasmodia of slime mould *P. polycephalum* were cultivated on agar blobs overlying glass coverslips (Fig. 1(upper)). The plasmodial 'tubes' growing between the blobs were injected with the fluorescent calcium indicators Fura-2 and Calcium Green-5N (Molecular Probes, USA) via a hollow glass needle using a CellTram Oil microinjection system (Eppendorf, Germany). Confocal laser scanning microscopy was performed with a Perkin Elmer Ultraview FRET-H ERS system and imaging data was post-processed with the Volocity Image Analysis software package (Improvision, USA).

Results

Calcium-containing vesicles were observed to collide regularly within the plasmodial cytoplasm. These events were found to result most commonly in vesicle reflection, as illustrated in Figure. 1(lower). Vesicles deform and travel as a single entity for a short period following collision in a manner similar to SSM-type collisions. The other collision types observed were categorised as fusion followed by adhesion, fusion followed by assimilation of one vesicle into another, annihilation of both vesicles followed by subsequent unloading of calcium into the cytosol and unknown/no discernible interaction.

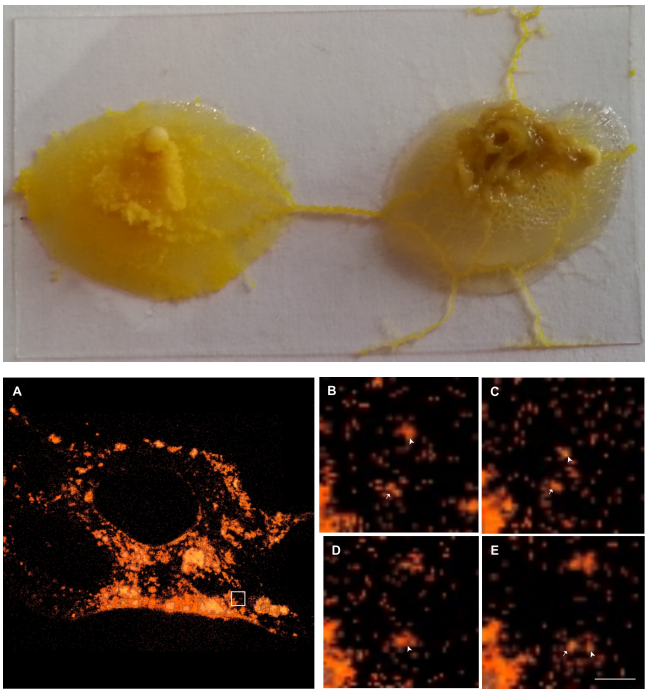


Figure 1: (Upper) Plasmodium of slime mould *P. polycephalum* propagating between two agar blobs overlying a glass coverslip. Note the plasmodial 'tube' that links both blobs that is used as a target for microinjection of fluorescent dyes. (Lower) Ratiometric confocal time-lapse footage of calcium-containing structures colliding within a plasmodial tube network: A, overview; B-E, two vesicles, indicated by arrows, collide and reflect. Time steps approx. 250ms intervals, scale bar 10μm.

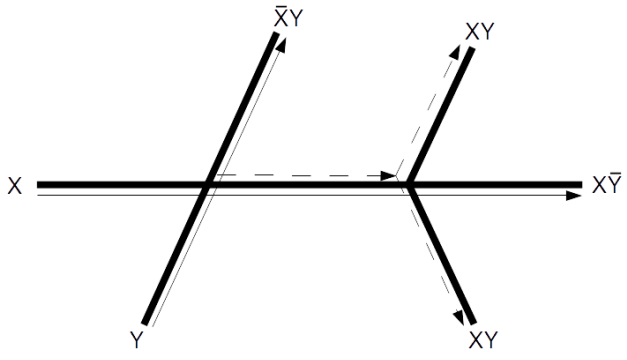


Figure 2: Design for a VCM interaction gate in an AND configuration based on known actin cytoskeleton conformations, where solid arrows represent vesicle trajectories if no other vesicle is present. When two vesicles collide, indicated by dashed arrows, they temporarily fuse and carry the incident vesicles along a different path in the network.

Discussion

Our key finding is that CBC can be realised through the collisions observed when confined to cytoskeletal networks in a manner similar to the precepts of the SSM, subject to certain modifications and restrictions. Figure 2, for example, shows a design for an interaction gate where vesicle trajectories are controlled by linkage to a hypothetical cytoskeletal network segment whose topology is based on well-characterised protein conformations — in this instance, actin filaments forming an 'X'-shaped junction to mimic filamin-induced cross-links, followed by a double branch based on Arp2/3 complex articulations. We find that through adaptation of this basic interaction gate, a range of logically reversible, computationally universal devices can be designed (Mayne and Adamatzky (2015)), although logical negation, disjunction and FAN-IN/OUT functions require the use of 'control signals', the isentropy associated with which disallows direct comparison with conservative CBC models.

Three key challenges must be overcome to practically implement VCM circuitry: control of circuit topology, synchronisation and collision type. We propose that all three of these aspects are programmable through manipulation of the cell's intrinsic control mechanisms. To further delineate, control over the cell's directional growth and cytoskeletal polymerisation, stimulation/repression of the biochemical oscillator (and hence, 'clocking' mechanism) that controls plasmodial cytoplasmic streaming and the synthesis of appropriate vesicles/vesicle surface modifications in conjunction with their loading onto specific cytoskeletal proteins are feasible routes to overcoming all of these criteria.

Appendix

The authors declare no conflict of interests and gratefully acknowledge funding from the EU Commission's Seventh Framework Program ('Phychip', GN 316366).

References

Berridge, M. (2004). Calcium signal transduction and cellular control mechanisms. *Biochimica et Biophysica Acta*, 1742(13):3–7. 8th European Symposium on Calcium.

Fredkin, E. and Toffoli, T. (1982). Conservative logic. *International Journal of Theoretical Physics*, 21:219–253.

Margolus, N. (2002). *Universal cellular automata based on the collisions of soft spheres*, chapter 5. Collision-based Computing. Springer.

Mayne, R. and Adamatzky, A. (2015). *In vivo* collision-based computing. *In Press*.

Toffoli, T. (1980). *Reversible computing*, volume 85 of *Lecture Notes in Computer Science*, pages 632–644. Springer Berlin Heidelberg.

Population-based microbial computing: a third wave of synthetic biology?

Martyn Amos

School of Computing, Mathematics & Digital Technology Manchester Metropolitan University, Manchester UK
M.Amos@mmu.ac.uk

Synthetic biology is an emerging research field, in which engineering principles are applied to natural, living systems. A major goal of synthetic biology is to harness the inherent "biological nanotechnology" of living cells for the purposes of computation, production or diagnosis. As the field evolves, it is gradually developing from a single-cell approach (akin to using standalone computers) to a distributed, population-based approach (akin to using networks of connected machines). We anticipate this eventually representing the "third wave" of synthetic biology (the first two waves being the emergence of modules and systems, respectively, with the second wave still yet to peak). In this talk we review the developments that are leading to this third wave, and describe some of the existing scientific and technological challenges.

Physical modelling exploring the self-organising behaviour of synthetic and natural biofilms

Jonathan Naylor, Harold Fellermann, Natalio Krasnogor
School of Computing Science,
University of Newcastle,
j.r.d.naylor@ncl.ac.uk

Introduction

To facilitate the understanding of biofilm self-organisation and inform the development of novel synthetic biofilms, computational models can be developed. Many different models exist for capturing a variety of biofilms, yet a standard modelling approach or tool has not yet been conceived (Dibdin, 1997; Lardon et al., 2011). The creation of such a generic tool is what we propose in this research, developing a multiscale modelling platform through which we will demonstrate the capacity for simulation to accurately capture and predict biofilm behaviour.

The complex interplay between microscale metabolic behaviour and macroscale spatial and functional organisation of biofilms leads to adaptive and ruthless survivors, in which altruistic behaviour can lead to the sacrifice of individual cells in favour of the survival of the colony as a whole (Kreft, 1999).

Understanding biofilms can offer insights into a range of situations, such as developing new anti-bacterials, counteracting dental plaque and in waste-water treatment (Fux et al., 2004). We can also endeavour on the development of synthetic biofilms, consisting of synthetic bacterial cohorts with targeted behaviour. Such artificial biofilms could form intelligent surfaces which respond to environmental changes (Rudge et al., 2012). Applications of these may include biosensors such as protective coatings against infection and intelligent toothpaste to fight dental plaque (Kumar and Rani, 2013).

Background

A thorough review of biofilm literature and existing models will offer insight into the appropriate level of abstraction for the simulation platform. Processes involved in the biofilm lifecycle vary in scale both temporally and physically; ranging from micro processes such as the diffusion of environmental chemicals, to macro processes such as the development of an *extracellular matrix (ECM)* which embodies the constituent cells (Flemming and Wingender, 2010).

A multiscale model must be developed such that it will accurately capture the key processes in the lifecycle of a typical biofilm. No rigid model would be able to handle a significant range of biofilm scenarios, thus a flexible software architecture and interface must be developed to allow model developers to substitute and add sub-models in a discrete fashion. Such an architecture could be readily extended or repurposed by future software development, advancing the potential of the tool.

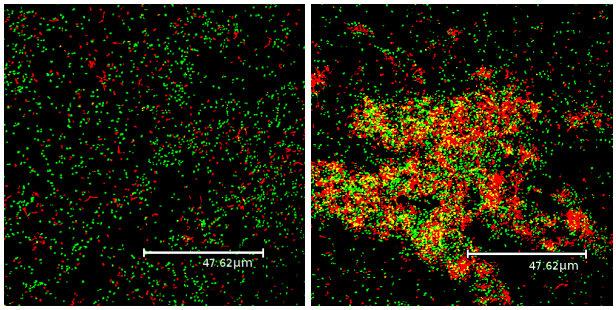


Figure 1: Microscopy images of floc formation in saliva
Waleed Mohammed, School of Denistry, Newcastle University

Developing an effective model of biofilms requires the simulation of about 10^5 physically interacting cells in some substrate. Each cell may consist of some internal metabolic behaviour and gene regulatory network, along with other behaviour such as interactions with their environment physically, biologically and chemically. To ensure a large range of individual cellular and environmental behaviour can be handled the platform should be *parallelised*, ensuring the feasability and scalability of the tool.

Both natural and synthetic biofilms will be modeled, aiding in the development of a standard tool for simulating biofilms. A natural context is dental plaque, consisting of mixed consortia of bacteria in the mouth which form flocs in the saliva and attach to the teeth and gingiva.

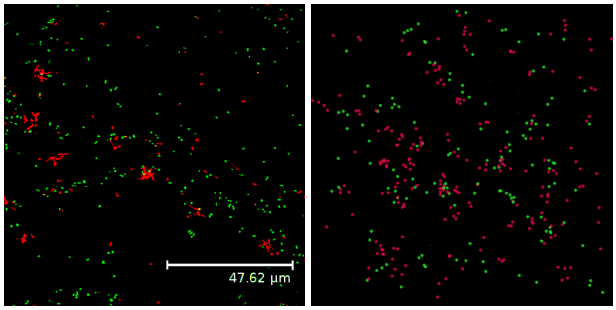
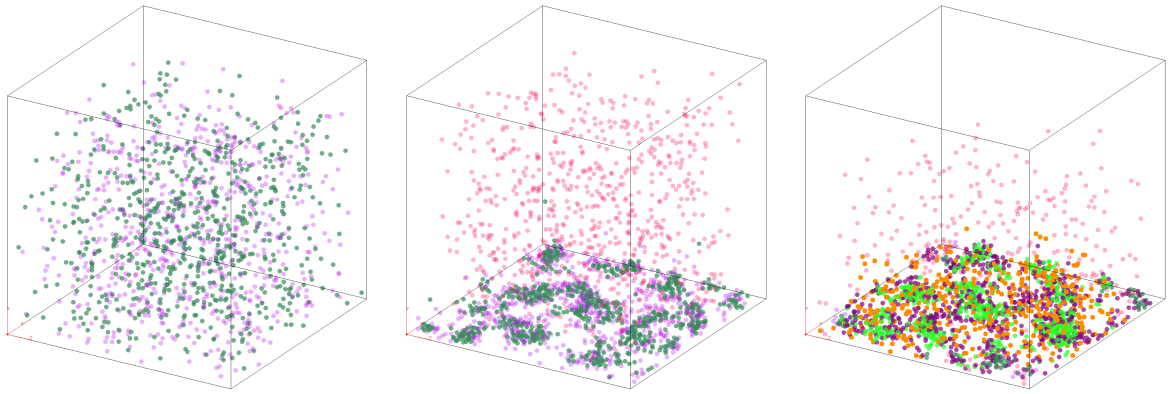


Figure 2: *Left:* Microscopy image of floc formation in plaque. *Right:* Simulation of floc formation



(a) **(b)** **(c)**

Figure 3: a) Initial planktonic synthetic *E. coli* which form defensive biofilm. **b)** Introduction of *Pseudomonas Aeruginosa* to grown defensive biofilm. **c)** Activation of the biosensor. *Purple* are *Sentinel cells* which detect the concentration of *AHL*, activating once this concentration is above some threshold. *Green* are *Bullet cells* which activate upon signal from *Sentinel cells*, in turn releasing chemicals which kill *P. Aeruginosa*. *Red* are *P. Aeruginosa*. Colour changes in *E. coli* strains indicates activation. *P. Aeruginosa* turns orange upon death.

A synthetic context is an artificial biofilm which aims to act as a biosensor, specifically a protective surface. Consisting of synthetic strains of E. coli which form biofilm to detect the presence of the pathogen *Pseudomonas Aeruginosa*. This is achieved by the sensing of the quorum-sensing molecule *AHL* used by *P. Aeruginosa*. Upon sensing a high concentration of *AHL*, the defensive biofilm activates, releaseing chemicals to interrupt the pathogen's coordination in forming biofilm and ultimately kill the invading pathogen.

Preliminary Results

An iterative software and literature review feeding back on development of the tool has lead to rapid development of the simulation platform, dubbed Bioform.

Constructed on a modified version of the Cortex3D-parallel software (Zubler, 2009), Bioform allows for the specification of user-defined physically, biologically and chemically interacting cells embedded in a 3D medium. This is achieved via a modelling interface which separates the concerns of the model specification and the simulation platform itself. Using this interface a variety of modules to represent cellular behaviour have been implemented; a *geometry* module to capture the shape and physical movement of the cell and an *internal* module to capture intracellular processes such as metabolism and gene regulation. Other modules can be readily designed and inserted capturing cellular features including but not limited to surface proteins, pili, secretion of *extracellular polymeric substances (EPS)* and sensor-secretors involved in quorum-sensing.

A 3D environment provides a physical medium for the cells to move though, a surface for the biofilm to develop on, and the diffusion of modeller-specified chemicals. Mixed consortia of bacteria can be simulated and viewed in a real-time visualisation.

Development of a base simulation platform and multiple models has lead to the formation of a generic engine, and a library of models and submodels with which it can interface.

Callibration of the tool involves modelling multiple biofilm scenarios, ensuring that the platform is not overfit to any specific model. Multiple prototype models are being developed during the creaton of the platform to inform the software architecture. Self-organising properties of biofilms can be observed in the simulations, both in the physical structure of the grown colonys and in the differentiation of cells.

These prototypes have been developed with the aid of collaborators; Association with the *School of Denistry, Newcastle University* insofar offers data on the processes involved in cell-to-cell adhesion leading to the simulation of plaque aggregation in the saliva. Affiliation with the *ROADBLOCK* project provides a variety of data on the development of a protective coating against the pathogen *P. aeruginosa*. As Bioform and these collaborations evolve, more precise models of these systems will be developed.

References

Dibdin, G. H. (1997). Mathematical modeling of biofilms. *Advances in Dental Research*, 11(1):127132.

Flemming, H.-C. and Wingender, J. (2010). The biofilm matrix. *Nature Reviews Microbiology*.

Fux, C. A., Wilson, S., and Stoodley, P. (2004). Detachment characteristics and oxacillin resistance of staphyloccocus aureus biofilm emboli in an in vitro catheter infection model. *Journal of Bacteriology*, 186(14):44864491.

Kreft, J.-U. (1999). Conflicts of interest in biofilms. *Biofilms*, 1(4):265276.

Kumar, H. and Rani, R. (2013). Development of biosensors for the detection of biological warfare agents: its issues and challenges. *Science Progress*, 96(3):294308.

Lardon, L. A., Merkey, B. V., Martins, S., Dtsch, A., Picioreanu, C., Kreft, J.-U., and Smets, B. F. (2011). idynomics: next-generation individual-based modelling of biofilms. *Environmental Microbiology*, 13(9):2416–2434.

Rudge, T. J., Steiner, P. J., Phillips, A., and Haseloff, J. (2012). Computational modeling of synthetic microbial biofilms. *ACS Synthetic Biology*, 1(8):345352.

Zubler, F. (2009). A framework for modeling the growth and development of neurons and networks. *Frontiers in Computational Neuroscience*, 3.

Analysing Synthetic Bacteria Colonies: An Agent-based Approach

Savas Konur[1], Mariam Kiran[1], Marian Gheorghe[1] and Omer Markovitch[2]

[1]School of Electrical Engineering and Computer Science, University of Bradford, UK
[2]School of Computing Science, Newcastle University, Newcastle, UK

Abstract

We discuss the agent-based modelling and simulation approach to analyse large bacterial colonies formed by many individual synthetic cells. We discuss the advantages and current limitations of the approach when modelling biological models.

Introduction

Synthetic biology is a subfield of biology, aiming at engineering new biological systems. It allows biologists to utilise biochemical mechanisms that implement specified behaviours. These biochemical systems (e.g. bacterial cells) inherently employ robust but complex mechanisms (e.g. genetic networks). Synthetic cells designed to perform simple computational tasks benefit from massive parallelism via biological reproduction at no extra development cost to the designer, resulting in the biological equivalent of a supercomputing cluster, where a large number of individually unremarkable compute nodes are pooled together to provide tremendous computational power Sanassy et al. (2014).

Although synthetic biology brings in novel engineering approach to build new biological systems, the interactions within genetic networks of synthetic biology systems are in general very complex and intricate. The computational methods, e.g. simulation techniques (in particular, stochastic simulation algorithms), provide significant advantages when designing and analysing synthetically constructed systems. This might be practical when the behaviour of single cell is being analysed. However, when synthetic bacterial colonies are in question, the usefulness of these methods becomes very limited, or even intractable. This is simply because bacterial cells comprise large scale reaction networks, and this escalates when bacterial colonies, formed by many individual cells, are considered.

Agent-based modelling is an alternative approach to model such large systems. In this approach, each agent (person, robot, bacterium, vehicle, etc.) is considered as an autonomous entity and the behaviour of a set of agents (i.e. *multi-agent system*) is defined by the actions and interactions between the agents within the system. This provides significant advantages over traditional methods when the collective and overall behaviour is simulated.

We propose the agent-based modelling and simulation approach to analyse large bacterial colonies formed by many individual cells. We discuss the advantages and current limitation of the approach.

Agent-based Modelling and Simulation

The agent-based modelling approach is particularly useful to model complex heterogeneous populations. It focuses on defining the behaviour of agents and observing the resulting emergent behaviours as agents interact with each other. An agent in this case can be thought of as an individual with their own memory and are capable of perceiving and interacting with the surrounding environment. The aggregate behaviour of the system emerges from the interaction of many independent agents with much more limited cognitive and computational capabilities.

The approach has been applied to model and study features of economic systems, including the distribution of firm sizes, income and wealth, and various aspects of the spatial structure of human activities Holcombe et al. (2012). It has also found further applications in systems biology, such as studying the intracellular NF–κB signalling pathway and the behaviour of bacteria in oxygen starved environments Maleki-Dizaji et al. (2009).

There are several agent-based tools and frameworks. FLAME (Flexible Large scale Agent-based Modelling Environment) Coakley et al. (2012) is a general purpose agent based framework, built on top of a state based model with transformation functions associated to the transitions of the model. FLAME has become very popular and widely used for numerous applications. FLAME allows modellers to simulate multi-agent models on parallel hardware architectures, including both high performance computers (HPCs) and modern graphics processing units (GPUs) Kiran et al. (2010).

FLAME GPU is a GPU version of the FLAME framework Kiran et al. (2010). It helps with the difficult task of managing inter-agent communications, births and deaths. The

framework abstracts this complexity and lets users define models in a simple way.

Both FLAME and FLAME GPU can be used to model and analyse bacterial colonies. The advantages and limitations of each approach is discussed below.

Case Study

We have studied different simulation methods to analyse synthetic bacterial populations. One of the example systems used in our analysis was the synthetic *pulse generator* Basu et al. (2004). A pulse generator is composed of two different bacterial strains: the *sender* and *pulsing* cells. The sender cells produce a signalling molecule `3OC12HSL` (`AHL`), which is transmitted to the nearest pulsing cells. The absorbed signalling molecules trigger the expression of the green fluorescent protein (`GFP`) in the pulsing cells. The excess amount of the signalling molecules is transmitted to the neighbour pulsing cells that are further away from the source sender cells. This creates a *propogation* of a wave of `GFP` from one end of the lattice to the other end. The sender cells are placed at one side of the lattice, and the pulsing cells are placed at the remaining part of the lattice.

In Blakes et al. (2014), we have used a multi-compartmental stochastic simulator to analyse the dynamics of the pulse generator system. Because of the complexity of the resulting system, we could only analyse small populations, e.g. up to 1000 cells.

In Bakir et al. (2014) and Burkitt et al. (2015), we have modelled and analysed the same system with FLAME and FLAME GPU. Whereas FLAME could analyse populations up to 10,000 cells, FLAME GPU could analyse much larger populations, e.g. 100,000 and above. FLAME GPU also provided massive increase in performance, at least by 3 orders of magnitudes, thanks to its HPC features.

Discussion

Based on the experiments conducted, we have concluded that stochastic simulation algorithms are inefficient if large synthetic populations are analysed. The advantage of this approach is that the exact kinetic behaviour of the entire system is known once simulations are run.

Agent-based approach provides significant improvements over conventional approaches, as shown in the case of FLAME and FLAME GPU. The GPU implementation of FLAME significantly reduces the time to perform a single simulation iteration. As the number of cells increases, the performance of FLAME GPU appears to be constant. This is due to the way the GPU is designed.

FLAME GPU, however, has its own limitations. For example, the FLAME model dynamically allocates memory as the simulation runs, using agent specific dynamic collections, whereas the GPU version requires all memory to be pre-allocated before the simulation starts. Also, the GPU version places some constraints in the way rules and other data structures are constructed. In addition, in GPU each agent can only create a single message, which limits the amount of possible communications between agents.

One major drawback of both FLAME and FLAME GPU is that they do not capture the stochastic behaviour. This is an important limitation if the exact kinetic behaviour of the system is in question.

We conclude that agent-based modelling and simulation is an efficient approach to analyse large bacterial colonies formed by many individual synthetic cells. However, the limitations reported in this abstract are the main challenges that need to be tackled in the next generation synthetic biology platforms.

Acknowledgements

SK and MG acknowledge the support provided for the synthetic biology research by EPSRC ROADBLOCK (project number: EP/I031812/1). OM is supported by EPSRC ROADBLOCK (project number: EP/I031642/1).

References

Bakir, M. E., Konur, S., Gheorghe, M., Niculescu, I., and Ipate, F. (2014). High performance simulations of kernel P systems. In *Proceedings of HPCC '14*, pages 409–412.

Basu, S., Mehreja, R., Thiberge, S., Chen, M.-T., and Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *PNAS*, 101(17):6355–6360.

Blakes, J., Twycross, J., Konur, S., Romero-Campero, F., Krasnogor, N., and Gheorghe, M. (2014). Infobiotics workbench: A P systems based tool for systems and synthetic biology. In *Applications of Membrane Computing in Systems and Synthetic Biology*, volume 7, pages 1–41. Springer.

Burkitt, M., Kiran, M., Konur, S., and Gheorghe, Marianand Ipate, F. (2015). Agent-based high-performance simulation of biological systems on the GPU. In *HPCC '15. Submitted*.

Coakley, S., Gheorghe, M., Holcombe, M., Chin, S., Worth, D., and Greenough, C. (2012). Exploitation of high performance computing in the flame agent-based simulation framework. In *Proceedings of HPCC '12*, pages 538–545.

Holcombe, M., Chin, S., Cincotti, S., and et.al. (2012). Large-scale modelling of economic systems. *Complex Systems*, 22(2):175–191.

Kiran, M., Richmond, P., Holcombe, M., Chin, L. S., Worth, D., and Greenough, C. (2010). FLAME: Simulating large populations of agents on parallel hardware architectures. In *Proceedings of AAMAS '10*, pages 1633–1636.

Maleki-Dizaji, S., Holcombe, M., Rolfe, M., Fisher, P., Green, J., Poole, R., and Graham, A. (2009). A systematic approach to understanding escherichia coli responses to oxygen: from microarray raw data to pathways and published abstracts. *Online Journal of Bioinformatics*, 10(1):51–59.

Sanassy, D., Fellermann, H., Krasnogor, N., Konur, S., Mierlă, L., Gheorghe, M., Ladroue, C., and Kalvala, S. (2014). Modelling and stochastic simulation of synthetic biological Boolean gates. In *Proceedings of HPCC '14*, pages 404–408.

Programming molecular structures and processes using DNA

Friedrich C. Simmel

Systems Biophysics and Bionanotechnology – E14,
Physics Department and ZNN, Technische Universität
München, Am Coulombwall 4a, 85748 Garching, Germany
simmel@tum.de

Maybe the most important feature of biopolymers such as DNA, RNA and proteins is that they can be regarded as information-encoding strings of molecular "letters". As these letters can be connected into strings in an arbitrary order, a vast combinatorial number of different "messages" can be generated. Depending on the context, these messages may mean something – for instance, DNA may code for a protein, a functional RNA molecule, or for a DNA nanostructure. Molecular programmers would now like to learn how to efficiently generate sequences that produce a desired (synthetic) biological structure or process – i.e., program biology using molecular code. In this talk, we will discuss a few of our own efforts along these lines, ranging from DNA nanostructures over in vitro biochemical circuits to in vivo synthetic biology.

A standard model repository for genetic design automation

Goksel Misirli, Jennifer Hallinan, Owen Gilfellon, Anil Wipat

Interdisciplinary Computing and Complex BioSystems, School of Computing Science, Newcastle University, Newcastle, UK
goksel.misirli@ncl.ac.uk, jennifer.hallinan@mq.edu.au, o.gilfellon@newcastle.ac.uk, anil.wipat@ncl.ac.uk

One ambition of synthetic biology is the large-scale engineering of biological systems. However, as the complexity and size of designs increases, the manual design of genetic circuits becomes more challenging. Computational tools often use libraries of mathematical models of biological parts in order to aid the user in building complex and predictable designs. To support automated, model-driven design it is desirable that in silico models are modular, composable and in standard formats. Here, we present an approach for composable and modular models for synthetic biology, termed standard virtual parts (SVPs), to support software tools in model-driven process. A repository of SVPs has been established to facilitate computational genetic circuit design.

Standard Virtual Parts (SVPs)

SVPs are reusable, composable, models of genetic parts (Cooling *et al.*, 2010). SVPs represent genetic features and their associated biochemical reactions, providing a one-to-one mapping between biological parts and their dynamic models.

SVPs are defined with standard inputs and outputs which are annotated semantically with machine-level information (Figure 1). Therefore, the process of model composition can be automated, facilitating the efficient construction of simulatable models for large-scale biological systems.

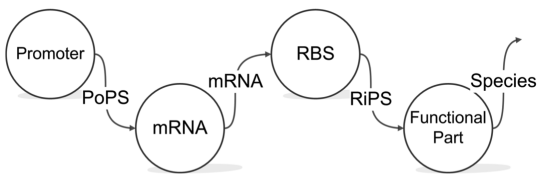


Figure 1: Interfacing SVPs. Each SVP is defined with standard inputs and outputs which are annotated to facilitate computational model composition.

A Repository of SVPs

Data are integrated (Misirli *et al.*, 2013) and mined to create models of biological parts and their constraints (Figure 2). SVPs allow the behaviour of genetic constructs built with these parts to be simulated. SVPs are annotated with information from the integrated data including types and nucleotide sequences of biological parts. These annotations facilitate the automated derivation of DNA sequences directly from dynamic models (Misirli *et al.*, 2011).

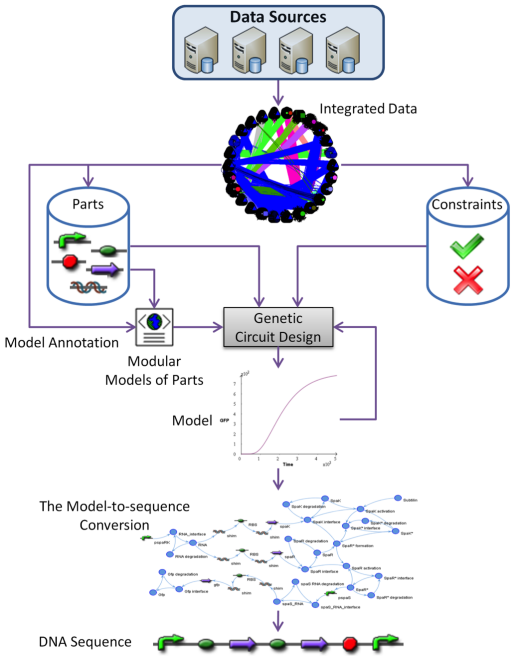


Figure 2: Integrated data are mined to derive information about biological parts and how they function together. Modular models of these parts are then represented with SVPs in order to facilitate automated derivation of DNA sequences using model-driven design approaches.

The Virtual Parts Repository (VPR)[1] includes around

[1] http://virtualparts.org

3000 SVPs, and is accessible as a human-browsable website and via a REST-based Web service (Misirli *et al.*, 2014). Currently SVPs are available in the Systems Biology Markup Language (SBML) (Hucka *et al.*, 2003) and the Synthetic Biology Open Language (SBOL) (Galdzicki *et al.*, 2014) formats. Promoter, FunctionalPart, RBS, Shim, Terminator, and Operator are types of SVPs. These SVPs can be searched for according to their functional roles such as biological processes, molecular functions and orthology memberships.

An API that enables programmatic access to the Web service is also available. Using the API, SVPs can be retrieved and joined together to create simulatable models.

Evolutionary Computation

We use an abstract SVP programming language, SVP-Write to represent potential genetic circuits constructed from SVPs. We also developed a tool that implements evolutionary algorithms in order to automate the search of the space of possible solutions for genetic circuit designs. The tool starts with an initial specification and a desired behaviour, and manipulates the topology and composition of the circuits, and compiles them into simulatable models (Hallinan *et al.*, 2014).

Conclusion

The VPR and the SVP approach presented here facilitates the model-driven design of biological systems. These models are reusable and can be assembled computationally to create systems models, acting as genetic blueprints that encode a desired biological system. SVPs can be retrieved and searched for from the publicly available VPR. Moreover, data are presented using existing standards such as SBOL and SBML for computational access. As a result, model-driven design and automation using SVPs provides an efficient approach facilitating predictable computational design of large-scale and complex biological systems.

References

Cooling, M.T. *et al.* (2010). Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics*, **26**(7), 925–931.

Galdzicki, M. *et al.* (2014). The synthetic biology open language (sbol) provides a community standard for communicating designs in synthetic biology. *Nature Biotechnology*, **32**(6), 545–550.

Hallinan, J.S. *et al.* (2014). Tuning receiver characteristics in bacterial quorum communication: An evolutionary approach using standard virtual biological parts. In *2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, pages 1–8, Honolulu, HI, USA.

Hucka, M. *et al.* (2003). The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**(4), 524–531.

Misirli, G. *et al.* (2011). Model annotation for synthetic biology: automating model to nucleotide sequence conversion. *Bioinformatics*, **27**, 973–979.

Misirli, G. *et al.* (2013). Bacillondex: An integrated data resource for systems and synthetic biology. *Journal of Integrative Bioinformatics*, **10**(2), 224.

Misirli, G. *et al.* (2014). Composable Modular Models for Synthetic Biology. *ACM Journal on Emerging Technologies in Computing Systems*.

Biochemical circuits in cell-sized microcompartments

Aurore Dupin[1] and Friedrich C. Simmel[1]
[1]Systems Biophysics and Bionanotechnology – E14, Physics Department and ZNN, Technische Universität München, Am Coulombwall 4a, 85748 Garching, Germany
contact : aurore.dupin@tum.de

In a cell, at every point of time a large number of biochemical processes occur that require specific, sometimes incompatible conditions. In the case of eukaryotic cells, many processes are separated from each other through compartmentalization into specialized organelles. The cell nucleus and these organelles permanently exchange nutrients with the cytosol while keeping their metabolism within physical boundaries. For instance, this allows reactions to be conducted at a particular pH, or in the presence or absence of specific enzymes, etc...

In the context of synthetic biology, compartmentalization would be extremely beneficial when attempting to build an artificial cell from the bottom up. It would allow for more complex reaction networks, replacing the need for completely orthogonal chemistry with the localization of the different steps of a synthetic metabolism.

Until now, in vitro biochemical circuits have been relatively limited in size and number of components in order to keep their kinetics and behavior predictable. The possibility to spatially separate circuit components would allow to scale up the circuits, and to potentially access more complex biochemistry.

Our present project aims at implementing such spatial separation within cell-scale systems using a class of biochemical reaction networks termed "genelet circuits".[1] Genelets are simplified in vitro gene networks that are based exclusively on in vitro transcription reactions. In the past, they have been used to generate multistability, oscillations, or other interesting dynamical behaviors. We previously studied the influence of compartmentalization on the dynamics of genelet circuits using water-in-oil droplets with impermeable surfactant membranes.[2]

In order to mimic communication between artificial organelles, we here study the behavior of genelet circuits encapsulated within droplets surrounded by a lipid monolayer. Using a technique previously developed by the Bayley lab[3], droplets are brought into contact to form a bilayer of lipids through which small molecules can migrate from one compartment to another. Apolar species can transfer directly through the bilayer, while ions and small charged or polar molecules are transferred via transmembrane pores such as the antibiotic α-hemolysin.

With those droplets, we build networks where the components are isolated, so that the circuit can only be activated when the compartments are brought together. We currently explore various modes of communication and activation between such droplets. For instance, the translocation of small ions could be used to feed a genelet circuit. A first droplet would contain a gene template and a transcription polymerase while a second droplet containing magnesium ions would serve as a reservoir to feed the transcription process. Similarly, larger molecules such as rNTPs could be isolated from the genelet to trigger the transcription and activate the fore-mentioned circuits.

On the long run, we envision that our networks will utilize larger molecules such as polymers as communicating agents. Translocated DNA sequences could act as genelet activators, while the transfer of RNA molecules would mimic the communication between nucleus and cytosol in eukaryotic cells. However, such transport of biopolymers presents a variety of technical challenges that will have to be solved.

We will present a series of designs and implementations of spatially-resolved circuits, and experimental approaches to achieve the translocation of diverse molecules. In the future, potentially artificial membrane pores made from DNA nanostructures of larger diameter could also be used as communication channels.

[1] E. Franco, E. Friedrichs, J. Kim, R. Jungmann, R. Murray, E. Winfree, F.C. Simmel, Timing molecular motion and production with a synthetic transcriptional clock, Proceedings of the National Academy of Sciences, *108*(40), 2011

[2] M. Weitz, J. Kin, K. Kapsner, E. Winfree, E. Franco, F.C. Simmel, Diversity in the dynamical behavior of a compartmentalized programmable biochemical oscillator, Nature Chemistry, *6*, 2014

[3] T. Wauer, H. Gerlach, S. Mantri, J. Hill, H. Bayley, K.t. Sapra, Construction and Manipulation of Functional Three-Dimensional Droplet Networks, ACS Nano, *8*(1), 2014

Programming synthetic scaffolds for DNA origami

Jerzy Kozyra[1], Chien-Yi Chang[2], Alessandro Ceccarelli[2], Harold Fellermann[1], and Natalio Krasnogor[1]

[1]School of Computing Science, Newcastle University
[2]The Centre for Bacterial Cell Biology, Newcastle University
j.w.kozyra@ncl.ac.uk

Abstract

Scaffolded DNA origami is one of the most successful methods allowing precise matter arrangement and manipulation on a nano-scale. The nanostructures are realised through rational programming of short oligos which fold a long viral DNA called the 'scaffold' strand. Here we propose a generalisation of the origami programmability which will now also include manufacturing protocol for purely synthetic scaffolds. We developed a strategy for *de novo* generation of long artificial DNA sequences which are uniquely addressable and biologically inert. Next, we show how they can be used as either DNA scaffolds or transcribed RNA scaffolds for origami systems. This new method not only allows building nanostructures which are programmed for a specific functionality but is also far more feasible for novel applications in Synthetic Biology.

Introduction

DNA origami is a powerful one-pot self-assembly technique proposed by Rothemund (2006) which enables robust construction of custom-shaped objects with a sub-nanometre precision. It allowed construction of many DNA nanodevices with controllable behaviour including nanorobots (Douglas et al., 2012). Natural biocompatibility of DNA nanostructures makes them an attractive candidate for cellular studies. DNA origami are stable in lysed cells (Mei et al., 2011) and can easily interface with biomolecules (Sacc et al., 2010). Also, DNA origami robots have been shown to emulate logic gates which remain functional in a living animal (Amir et al., 2014).

One of the factors restricting the complexity, applicability and scalability of the DNA origami is the source of the scaffold which is of viral origin. Not much research has been focused on addressing this issue, mainly because of the increasing ease of obtaining and manipulating the single-stranded DNA of M13 bacteriophage or variations thereof. Currently the sequence design and its optimisation is limited to either cyclic permutations of the existing scaffolds or alterations of folding paths producing alternative scaffold-staples layouts.

Here, I explain how we can tackle this problems by exploiting a concept from combinatorial mathematics, namely a de-Bruijn sequence, to generate synthetic and programmable DNA origami scaffolds.

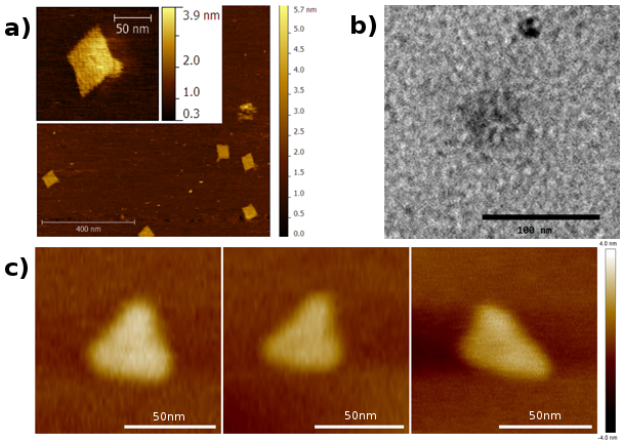


Figure 1: Imaging: a) AFM of pUC19-based square DNA origami; b) TEM of de-Bruijn square DNA origami; c) AFM of de-Bruijn triangle RNA-DNA hybrid origami.

Synthetic de-Bruijn scaffold

De-Bruijn is a cyclic sequence over a certain alphabet A which contains every possible subsequence of length n exactly once. Efficient algorithms for construction of such sequences exist and rely on finding Eulerian cycles in graphs representing all possible overlaps between sequences (i.e. de-Bruijn graphs). That in essence allows the generation of addressable scaffolds which can be subjected to our filtering algorithms. The removal of user-specified sequences provide a route to prevent any undesired biological activity in the fundamental components of DNA origami.

De-Bruijn sequences were already used in the context of DNA origami (Wagenbauer et al., 2014) as quality control probes, but never before as universal scaffolds. Here, we use this method to (commercially) synthesise two sequences: 2.4kb long DNA scaffold and 1.1kb RNA scaffold which we fold into square and triangle nanostructures, respectively (Figure 1).

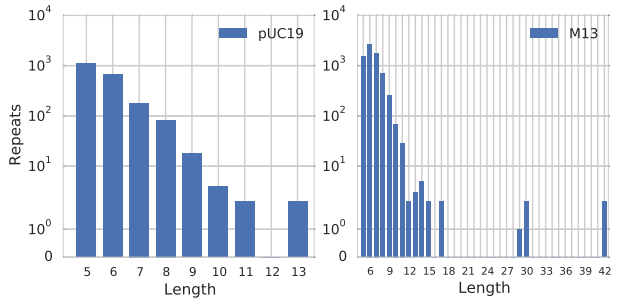


Figure 2: Repetitions in scaffolds: 2.6kb-long pUC19 vector (left) and 7.2kb-long M13mp18 bacteriophage genome (right). Length of the repeated k-mer (x-axis) ploted against total number of repeated k-mers in the scaffold (y-axis).

Results

The shortest addressable staple domain in DNA origami is typically 8-nt and 7-nt long for 2-dimensional and 3-dimensional structures, respectively. Our statistical analysis shows many repetitions in the existing scaffolds which are longer than staple domains (Figure 2). Moreover, as the scaffold length increases the number of sequence repetitions grows exponentially. Interestingly, M13, which is the most popular scaffold, has the longest repeated subsequences, spanning over 42-nt. In comparison, the two de-Bruijn scaffolds that were synthesised for this study do not contain repeats longer than 5-nt.

We also investigate the thermodynamics of staples in different designs and scaffolds configurations. As expected, lack of long repeats in de-Bruijn sequences has a desired effect on the Boltzmann (probability) distribution highly favouring specific hybridisation of staples. In the short viral sequence, the probability measure indicated (in the worst case) a chance as low as 30% for a staple to bind the correct section of the scaffold (Figure 3). In contrast, this measure rarely declines below 80% when de-Bruijn scaffold is used.

Conclusions

The repetition of sequences in all natural scaffolds suggests that they are not uniquely addressable by staples. This is especially problematic for large DNA origami because the number of energetically-preferable targets for a staple grows exponentially with the scaffold length. Thus, the use of natural scaffolds is not truly scalable. Moreover, self-assembly of DNA origami relies on three assumptions: (i) cooperative effect of staples, (ii) strand-displacement reactions, and (iii) large enough distances between 'competing' scaffold domains in the final nanostructure. Those effects are necessary for the correct folding of structures, but are currently hard to control in pragmatic manner.

Our study eliminates above problems by challenging the fundamental assumption in the field of DNA origami,

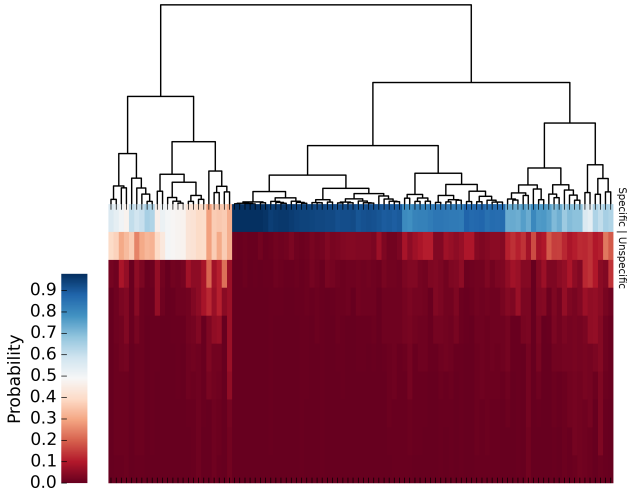


Figure 3: Probability distribution: 10 most energetically-stable targets in the pUC19 scaffold. Each column represent one short (8-nt) staple domain in the DNA origami design.

namely that a viral sequence is to be used as the scaffold. Advantages provided by our methodology could find novel applications in Synthetic Biology and are especially suited for extreme-precision patterning as well as self-assembly in biotic environments.

Acknowledgements

This work was supported by EPSRC grant No. EP/J004111/1.

References

Amir, Y., Ben-Ishay, E., Levner, D., Ittah, S., Abu-Horowitz, A., and Bachelet, I. (2014). Universal computing by dna origami robots in a living animal. *Nat Nano*, 9(5):353–357.

Douglas, S. M., Bachelet, I., and Church, G. M. (2012). A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335(6070):831–834.

Mei, Q., Wei, X., Su, F., Liu, Y., Youngbull, C., Johnson, R., Lindsay, S., Yan, H., and Meldrum, D. (2011). Stability of dna origami nanoarrays in cell lysate. *Nano Lett.*, 11(4):1477–1482.

Rothemund, P. W. (2006). Folding dna to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302.

Sacc, B., Meyer, R., Erkelenz, M., Kiko, K., Arndt, A., Schroeder, H., Rabe, K. S., and Niemeyer, C. M. (2010). Orthogonal protein decoration of dna origami. *Angewandte Chemie International Edition*, 49(49):9378–9383.

Wagenbauer, K. F., Wachauf, C. H., and Dietz, H. (2014). Quantifying quality in dna self-assembly. *Nat Commun*, 5:–.

Self-assembly of a functionalized RNA nanostructure

Matthaeus Schwarz-Schilling[1], Fabio Chizzolini[2], Andrea Mückl[1] and Friedrich C. Simmel[1]

[1]Systems Biophysics and Bionanotechnology – E14, Physics Department and ZNN, Technische Universität München, Am Coulombwall 4a, 85748 Garching, Germany
[2]CIBIO, University of Trento, via delle Regole 101, 38123 Mattarello, Italy
contact: matthaeus@mytum.de

Living organisms are regarded as complex systems that exist far off from equilibrium. The complexity arises from the high degree of interconnectivity between the components. In addition, the non-equilibrium nature of biological systems can introduce non-apparent dynamical behavior. This has to be considered in Synthetic Biology, the study of building or redesigning biological like systems. Hence a quantitative prediction of the interactions within and at the interface of the system is of crucial importance.[1]

For example, the insight into DNA-DNA interactions has led to a zoo of three-dimensional nanostructures using only base-pairing interactions.[2] However, due to non-physiological folding conditions and single-stranded regions of DNA the compatibility of the nanostructures with biological systems is limited.

RNA shares the aspect of sequence programmability with DNA, while being suited for *in vivo* applications due to isothermal folding of structures.[3] The RNA structures can be introduced into cells, by encoding them on plasmids and in contrast to the genetic code, RNA is constantly degraded by RNase. For RNA structure design, there is library of well-characterized structure motifs, such as protein or-fluorophore binding aptamers, that can be used as functional subunits.[4]

In order to study the dynamical process of the self-assembly of a nanostructure *in vitro* and *in vivo,* we designed a RNA-scaffold that constitutes of three different parts. The first part is a Malachite Green (MG) binding aptamer, which enhances the fluorescence of its ligand. The second part is composed of two peptide-binding RNA aptamers PP7 and Biv-Tar, The third part constitutes of a streptavidin binding RNA aptamer. The three branches are connected via a three-way junction (3WJ) motif from the packaging RNA of the bacteriophage phi29.

Each arm of the RNA structure serves as a reporter for a different step in the self-organization of the RNA-Protein nanostructure. From the fluorescence of the Malachite Green aptamer complex, one can deduce the concentration of the RNA. The PP7-peptide is attached to a CFP and the Biv-Tat peptide to YFP. The positions of the aptamers arrange both proteins within their Förster radius of the FRET pair. The emergence of the FRET signal reports the assembly of the two proteins. The streptavidin binding RNA aptamer is used to locate the RNA nanostructure at streptavidin coated surfaces.

As a proof-of-principle, the assembly of the nanostructure from the purified components was shown in a bulk measurement. Subsequently the plasmid DNA coding for the nanostructure has been placed in three different environments: *E.coli* cell extract in bulk, *E.coli* cell extract in a micrometer-sized emulsion droplet with streptavidin coated walls and *E.coli.* The *in vitro* transcription and translation reduce the complexity of the chassis and may serve as a breadboard for the optimization of the system before placing it in bacteria.

As a next step, the optimization of the self-assembly process in the *E.coli* extract is envisioned. One measure is the replacement of the Biv-Tar RNA aptamer by the peptide binding RNA aptamer MS2. This RNA aptamer has a higher dissociation constant, which is also closer to the one of PP7. Aiming towards stoichiometry of the components will facilitate the dynamics and preserve resources. In addition, degradation tags will be attached to the proteins, in order to avoid the accumulation of the proteins.

At last, for the self-assembly of the nanostructure *in vivo* a second design will be introduced, that replaces the protein FRET-pair with two enzymes, such as LuxS and LsrK kinase. Here the RNA scaffold will catalyze the phosphorylation of AI-2 by allowing substrate channeling.

This system is used to study the self-assembly of a RNA-protein nanostructure. The fluorescence spectroscopy of the Malachite Green aptamer, the CFP, the YFP and the FRET signal allow for a quantitative analysis of each different step. We hope to gain insight into the dynamics and interactions of the RNA nanostructure in cell-free systems and bacteria.

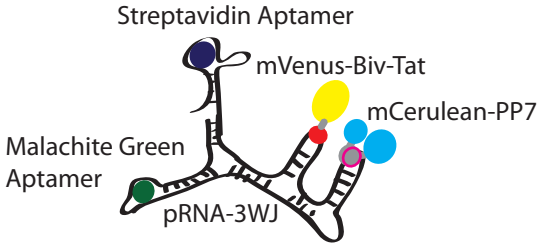


Figure 1 Sketch of the fully assembled RNA-scaffolded nanostructure.

[1]S. Rollié, M. Mangold, K. Sundmacher, Chemical Engineering Science, **69** (2012), 1-29
[2]V. Linko, H. Dietz, Current Opinion in Biotechnology, **24** (2013), 555-561
[3]C.J. Delebeque, A.B. Linder, P.A. Silver, F.A. Aldaye, Science, **333** (2011), 470-474
[4]R.L. Strack, S.R. Jaffrey, Current Opinion in Chemical Biology, **17** (2013), 651-655

Participants

Martyn Amos
School of Computing, Mathematics & Digital Technology
Manchester Metropolitan University
Manchester, UK
m.amos@mmu.ac.uk

Aurore Dupin
Systems Biophysics and Bionanotechnology
Physics Department and ZNN
Technische Universität München
München, Germany
aurore.dupin@tum.de

Harold Fellermann
Interdisciplinary Computing and Complex Biosystems
School of Computing Science
Newcastle University
Newcastle, UK
harold.fellermann@ncl.ac.uk

Owen Gilfellon
Interdisciplinary Computing and Complex Biosystems
School of Computing Science
Newcastle University
Newcastle, UK
o.gilfellon@ncl.ac.uk

Jerzy Kozyra
Interdisciplinary Computing and Complex Biosystems
School of Computing Science
Newcastle University
Newcastle, UK
j.w.kozyra@ncl.ac.uk

Curtis Madsen
Interdisciplinary Computing and Complex Biosystems
School of Computing Science
Newcastle University
Newcastle, UK
curtis.madsen@ncl.ac.uk

Omer Markovitch
Interdisciplinary Computing and Complex Biosystems
School of Computing Science
Newcastle University
Newcastle, UK
omermar@gmail.com

Richard Mayne
International Center of Unconventional Computing
Faculty of the Environment and Technology
University of the West of England
Bristol, UK
richard.mayne@uwe.ac.uk

Goksel Misirli
Interdisciplinary Computing and Complex Biosystems
School of Computing Science
Newcastle University
Newcastle, UK
goksel.misirli@ncl.ac.uk

Jonathan Naylor
Interdisciplinary Computing and Complex Biosystems
School of Computing Science
Newcastle University
Newcastle, UK
j.r.d.naylor@newcastle.ac.uk

Matthaeus Schwarz-Schilling
Systems Biophysics and Bionanotechnology
Physics Department and ZNN
Technische Universität München
München, Germany
matthaeus@mytum.de

Friedrich C. Simmel
Systems Biophysics and Bionanotechnology
Physics Department and ZNN
Technische Universität München
München, Germany
simmel@tum.de