

## Introduction to synthetic biology

Synthetic biology aims at the rational design and construction of biological parts, devices, and systems in order to engineer organisms to perform new tasks Lu:2009ez,Andrianantoandro:2006bi. A part is a basic unit, like a promoter or a ribosome binding site that when combined with other parts will make a functional unit, a device Heinemann:2006ht. A device processes inputs, performs functions and produces outputs Andrianantoandro:2006bi. A system comprises of a collection of devices.

Emphasis is put on the use of engineering principles such as modularity, standardisation, use of predictive models and the separation of design and construction Agapakis:2009bt, Heinemann:2006ht. A hierarchy similar to computer science is used, with cells, pathways and biochemical reactions acting as computers, modules and gates respectively Andrianantoandro:2006bi.

Numerous applications of synthetic biology have emerged, from altering existing metabolisms to producing synthetic drugs Holtz:2010bm or creating new synthetic life forms Agapakis:2009bt. Despite the successes there is still a lack of predictive power due to the stochasticity and lack of complete knowledge of the cellular environment Andrianantoandro:2006bi.

Synthetic biology is now entering an age where simple synthetic circuits have been built, such as toggle switches Gardner:2000vha, Kramer:2004kq, Isaacs:2003ht, Ham:2008hh, Deans:2007cya, Friedland:2009ce, oscillators Stricker:2008jqa, Fung:2005jd, Tiggles:2009jx and pulse generators Basu:2004gn, but larger circuits have proven more difficult XXX. The leap from building low-level circuits to assembling them into complex networks has yet to be made successfully Lu:2009ez, and predictable circuit behaviour remains challenging XXX. Efforts to do so are plagued by intra-circuit crosstalk and incompatibility, as well as cellular noise, which can render synthetic networks non-functional in vivo XXX.

## System design in synthetic biology

Creating synthetic devices that are robust to changing cellular contexts will be key to the success of synthetic biology. Unknown initial conditions and parameter values as well as the variability of the cellular environment, extracellular noise and crosstalk makes the majority of synthetic genetic devices non-functional Chen:2009ea. Designing devices robust to this environment will lead to reliable behaviour of the systems. When faced with a set of competing designs for a given genetic circuit, one is likely to choose the simplest possible model that can achieve the desired behaviour. However, simple systems are often the least robust. Feedback loops are well known key regulatory motifs Brandman:2005ci. Negative feedback loops are essential for homeostasis and buffering Thomas:1995id thus increasing robustness to extrinsic noise sources and positive feedback loops can generate multistationarity in a system Thomas:1995id. Incorporating this kind of additional feedback interactions can make a design more robust and reliable. Maximising production is an important goal for a metabolic engineering project if it is to produce an economically viable substance Holtz:2010bm. Network topologies and parameter values of different toggle switch designs are explored here in order to identify the design that maximises robustness and distance between steady states. This ensures the reliable production of the product with the greatest distance between the on and off states of the switch. In the future, by selecting the system components accordingly, the parameter values can be adjusted in vivo. For example, the parameter value corresponding to the translation initiation rate can be chosen by selecting the appropriate RBS sequence which given a nucleotide sequence will produce the desired rate Holtz:2010bm, a method developed by Salis:2009gk Salis:2009gk. Another method to tweak the parameter values in vivo is to select the promoter to have the strength corresponding to the levels of gene expression and repression desired. Activity of each promoter can be measured and standardised Kelly:2009bj making this process possible. For a system requiring more than one promoter, these can be efficiently selected from a promoter library using a genetic algorithm created by Wu:2011bq Wu:2011bq. These standardised interchangeable components with known sequence and activity are what synthetic biology classes as BioBricks Kelly:2009bj,Canton:2008fv. These can be selected and used to construct a desired system and replicate the parameter values found in the scan presented here.

The first computational approach for the tuning of robust synthetic networks was that of Batt:2007jl Batt:2007jl where they examined the problem of finding a subset of the parameter set for which a given property was satisfied for all the parameters. Chen:2009ea Chen:2009ea used the fuzzy dynamic game method to solve the minimax regulation design problem of synthetic genetic networks. In that method the worst case effect of all disturbances is minimised for a given network. An evolutionary algorithm has also been used to solve the robust design problem by evolving the parameters of the system in order to make it more robust to

cellular disturbances by Chen:2011hj Chen:2011hj. The added value of the methodology presented here is that the network structure in addition to the network parameters are adjusted to select a network that can robustly create the desired behaviour.

#### Introduction to Biochemical Modelling Graphical representation of biochemical systems

It is common to represent coupled biochemical reactions graphically. In a graph, as shown in Figure fig:Toggle switch example, nodes represent the species and the edges represent an interaction between the species it connects, in which a transcription factor directly affects the transcription of a gene alon:2007b. An arrow at the end of an arc represents activation, i.e. that when the transcription factor binds to the promoter the rate of transcription of the gene increases. A flat line perpendicular to the arc at the end of an arc represents repression, i.e. that when the transcription factor binds to the promoter the rate of transcription of the gene decreases alon:2007b.

#### Deterministic and Stochastic modelling

Modelling attempts to describe the elements and dynamics of the biochemical system of interest. It is a tool used for integrating knowledge and experimental data as well as for making predictions about the behaviour of the system wilkinson:2006. When modelling a biochemical system it is generally assumed that the rates of a reaction are directly proportional to the concentration of the reactants, raised to the power of their stoichiometry wilkinson:2006. This is known as mass-action kinetics and is used in this work to model the various systems. There are two main ways of modelling a system, deterministically and stochastically. Deterministic modelling utilises ordinary differential equations (ODE) and models the concentrations of the species (proteins or other molecules) by time-dependent variables deJong:2002ft. Rate equations are used to model gene regulation where the rate of production of a species is a function of the concentrations of the other species deJong:2002ft. When modelling deterministically the model is viewed as a system which, with sufficient knowledge of the system, its behaviour is entirely predictable. Nevertheless we are still a long way away from having complete knowledge of a system of interesting size wilkinson:2006. Deterministic modelling also assumes a homogenous mixture where species concentrations vary continuously and deterministically, assumptions that often are not met in vivo. A cell is spatially and temporally separated, due to small molecule numbers and fluctuations in the timing of processes deJong:2002ft.

In stochastic modelling, species are measured in discrete amounts rather than concentrations and a joint probability distribution is used to express the probability that at time  $t$  the cell contains a number of molecules of each species deJong:2002ft. It takes uncertainty into account and does not assume a homogenous mix. It is thus often more appropriate for modelling cellular systems, although more computationally intensive. In stochastic systems the Gillespie algorithm is widely used to simulate the time-evolution of the state of the system wilkinson:2006. The algorithm, developed by Gillespie:1977ww Gillespie:1977ww can be summarised in four steps: enumerate

N umber of molecules in the system initialised

T wo random numbers generated, one to determine which reaction will occur next and one to determine the time step

T ime step increased and molecule counts updated according to Step 2

R epeat from Step 2 until total simulation time reached