

Introduction

Felipe Aguilera Millacura

January 19, 2021



THE UNIVERSITY
of EDINBURGH

School of Biological Sciences
Institute of Quantitative Biology,
Biochemistry and Biotechnology
University of Edinburgh
Edinburgh, EH9 3JQ
United Kingdom

Contents

1	General Overview	2
---	------------------	---

Chapter 1

General Overview

Since the beginning of time life has made his way through multiple adverse conditions. Organisms have evolved from single cells interacting directly with their environment, to specialised multi cellular entities that respond specifically to different stimuli. Even though cells can specialise their functions by differentiating into diverse cell types, aggregating into tissues or organs, or by generating a variety of organisms, they all respond to the same paradigm. DNA is used to save and protect the hereditary information, which is later transcribed into functional orders in the form of RNA, being most of the time later translated into protein products that execute those final orders. This process was broadly studied on the twentieth century in order to explain how life works. DNA was discovered to be the source of information that carries and saves all instructions for securing life maintenance over generations (?). Although RNA and protein synthesis can occur almost simultaneously in the cell (?), proteins need to acquire their secondary, tertiary or quaternary structure to become functional, with even sometimes needing post-translational modifications to work.

One of the first mentions of Synthetic Biology can be tracked way back to ?, who sought to synthesise life "by directing the physical forces which are its cause".

Cell-free expression systems are a powerful tool in applications where cell based expression systems are too variable, slow, difficult to store or prohibitive due to legislation on the release of genetically modified organisms. Cell-free systems (CFS) are not without their limitations; since their very nature is to be a defined, precisely controlled system that lacks self-replicating functionality, they depend on a limited quantity of supplies, such as ATP, amino acids and nucleotides ?. Consequently, in-vitro synthesis of protein is limited by two dimensions: the overall maximum protein amount that can be synthesized with the supplied components, and the time during which the CFS stays active before background processes and chemical deterioration have used up supplies or inactivated the system ??? One of the most promising applications of CFS is the possibility to design cell-free biosensors, which do not present a living prospect subject to the current GMO legal regulations and offer an alternative to standard analytical techniques, such as ICP-MS and ICP-OES. However, cell-free biosensors that rely on protein expression also present the common limitations of CFS, such as a long response time and a lack of precursor regeneration. Sensors that

could function without the need for protein expression would drastically reduce the requirements for such a system. Paige et al., (2011) designed synthetic fluorescent RNA aptamers that mimic fluorescent proteins without the need for translation. The secondary structure of these aptamers presents a loop in which a fluorophore is trapped, conferring the same intramolecular immobilization that confers fluorescence in fluorescent proteins (Paige et al. 2011; Strack et al. 2013). These RNA aptamers also suffer from several limitations as incorrect folding is observed under non optimal thermal or ionic conditions (Autour et al. 2016). Improved aptamers with a stronger secondary structure, such as Broccoli and iSpinach, show increased stability under non-optimal conditions (Filonov et al. 2014, Autour et al. 2016). Both aptamers are based on the 95 base core sequence of Spinach2 and can be flanked by a tRNA scaffold, increasing the stability of their secondary structures (Filonov et al. 2014, Autour et al. 2016, Strack et al. 2013). However, these were designed for in vivo use (e.g. analysing RNA quantities) and have never been tested as outputs for a cell-free sensor. Such a sensor could require only transcription, not translation, hence could be much faster and simpler. On the other hand, the simplicity of CFS can become a disadvantage when complex processes need to be performed. For instance, the high dependence on *E. coli* as a model organism generates a problem, since genes or genetic pathways from non-model organisms or obtained by metagenomic studies, may fail to be expressed in standard organisms. Moore et al., (2016 and 2017) have attempted to solve this issue by using non-model bacteria, such as *Bacillus* or *Streptomyces*, for the generation of new cell-free TX/TL systems. This approach can be applied to other organisms, such as extremophiles, in order to use their unique capacities. For example, *Cupriavidus metallidurans* possesses the capability to tolerate, remove and degrade diverse environmental pollutants (Millacura et al., 2017). A cell-free system based on *Cupriavidus metallidurans* extract might show increased tolerance for heavy metals and might be superior for heavy metal sensors. Another challenge for CFS is to analyze multiple variables simultaneously. As they rely on the normal cell processing/synthesis machinery (interaction between transcription factors, polymerases, ribosomes, and other diverse macromolecules), they suffer the same issues as a living organism. Furthermore, when complex interactions are carried out, problems may arise due to kinetic mismatches, lack of oscillation or Boolean/Reversible designs (Guz et al., 2016). The generation of cell-free systems that respond to variables using totally synthetic processing machinery seems to be one of the most promising approaches. There have been some attempts to generate logic gates without mimicking normal cell machinery (Bordoy et al., 2016, Chatterje et al., 2016, Kim et al., 2014 and Zhang et al., 2016), however, due to their dependence on translation and/or the need for manually addition of foreign oligonucleotides (Kim et al., 2014) they still show slow response. Additionally, some of them rely on recombination processes that make their implementation even more complicated than under in vivo conditions (Zhang et al., 2016). In electronics, a single circuit accepts one or more binary inputs to generate one or more binary outputs. There have been many attempts to replicate such circuits using in vivo genetic networks. A typical biological logic gate consists of an output macromolecule that is produced only if the corresponding pattern of inputs is present, inputs that are commonly associated

with the presence of transcriptional regulators, transcription factors, polymerases or other macromolecules ? . Here we propose an alternative approach, BioLogic, which decomposes a large circuit into a collection of small subcircuits, solved in parallel. Rather than having a single type of cell (or genetic material) doing the computation, we have separate versions (subcircuits) each reacting to a different combination of inputs and generating the desired response by combination of each final output. For instance, if each input bit is considered in two forms, ZERO and ONE, each of which is essential to certain output agents, any arbitrary pattern of outputs for any pattern of inputs could be generated, making all kinds of binary operation possible. A further proposal is to achieve this goal by using just a transcriptional genetic network, without need for translation. Each agent in this case, rather than a cell, is a plasmid or double stranded DNA that does not possess functional transcriptional networks (Figure 1). As cell-free systems can assemble DNA parts in vitro at very low cost by using cell extracts instead of enzyme cocktails (Casini et al., 2015), and Golden Gate Assembly is an excellent tool to assemble multiple DNA fragments in a defined linear order, by using type IIS restriction enzymes (Engler et al., 2008), these techniques may be combined to generate a new and inexpensive in vitro system, which through DNA assembly and processing (hereafter DNALogic) would be able to sidestep common problems associated with the use of live cell based systems. As a "triggering" step is vital for the DNALogic processing machinery, the novel CRISPR/Cpf1 technology (or similar) will be used to generate the respective recognition and cutting steps. CRISPR/Cpf1 generates 5' overhanging cuts, producing 8 nt sticky ends and cleaving at the 14th base from the PAM site TTN (Lei et al., 2017 and Li et al., 2016). Due to its similarities to Type IIS restriction enzymes, Cpf1 can also be used to generate a new kind of Golden Gate Assembly, which would not rely on the presence or absence of cutting sites, but on gRNA. The gRNA expression system can be combined with the recognition step, working as a transduction system between the detection and processing steps (Figure 1a). This CRISPR triggered sequence recognition, followed by cutting and ligation steps, would result not only in a fully functional transcriptional unit, but also in a different double stranded genetic sequence, accomplishing the parameters for synthetic memory generation explained and demonstrated in vivo by Siuti et al. (2013). This creates DNA memory that can not only be later amplified and processed by sequencing and/or PCR (later response), but that can also generate RNA expression of a fluorescent aptamer or a further gRNA if there is formation of a proper transcription unit (immediate response).

Figure 1. Proposed DNALogic system.

A) crRNA expression is regulated by inducible promoters B) Ribonucleoprotein formation is followed by recognition and cutting of the target DNA C) Ligation of the overhangs produced by the Ribonucleoprotein generates a transcriptional unit D) Expression of the transcriptional unit generates a functional fluorescent RNA aptamer.

Aims: This work has as its general aim the generation of a modular cell-free system capable of not only sensing variables present in the surrounding environment, such as heavy metals, but that is also able of analyzing them by using a totally synthetic in vitro logic gate. Specific aims are: 1. Testing fluorescent RNA aptamers as output signals for cell-free sensors. 2. Developing Cpf1-based signal processing technology for use in cell-free systems. 3. Creation

of an adaptable DNALogic gate, which allows memory and more complex analysis in cell free systems. 4. Testing *Cupriavidus metallidurans* extracts as basis for metal-sensing cell free systems using endogenous and exogenous metal sensing promoters. 5. Integrating these to develop a modular cell-free sensing system with versatile in-vitro signal processing.