Final Publishable Summary Report

Directed Evolution in vivo enabled through genetic circuits in a Synthetic Biology approach

FP7-MC-IEF 628019

Background

Given our limited ability to rationally design proteins to our needs, their supply for biotechnological solutions leading to a sustainable Bioeconomy is relying on proteins found in nature and methods for deriving improved versions. Proteins evolved their structure to fulfil certain functions in very specific niches, which does not resemble optimal properties in the new context of biotechnological applications. Hence, improving proteins in their activity, affinity and stability is a very desirable, but underdeveloped key ability.

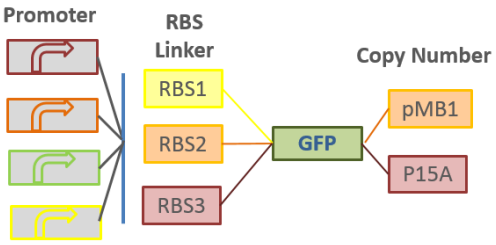
The current physical methods for improving proteins through directed evolution require intense human intervention and are limited in their ability to sample sequence space beyond single and double mutations. This project will establish a novel system, in which the whole process of mutation and selection takes place *in vivo* inside *E.coli* and can be controlled through a genetic network designed in a Synthetic Biology approach. It will enable exploring bigger sequence space during directed evolution, while controlling mutagenesis rate and selection stringency.

We have demonstrated a ground-breaking novel method to target mutagenesis *in vivo*. Work has also progressed to increasing the degree of diversification within the *in vivo* mutation system. In addition gene circuitry has been developed to take the project towards its ultimate goal of a self-evolving self-selectable system. The resulting directed evolution system will be readily adaptable to new proteins of interest and is anticipated to have a big impact in the development of new protein specificity and functionality for industrial purposes.

As part of the development of the synthetic toolbox required to achieve the aims of this project, we have developed a new DNA assembly standard and method. This greatly facilitates our ability to rapidly build and prototype new biological systems. This has been used extensively throughout this project and has provided a good demonstration of the enhanced capabilities that can be achieved through the implementation of standards in a biological design and optimisation workflow.

Main Results

To achieve the goals of the project, it has been necessary to implement a design-build-test workflow to achieve the robust development of mutation and circuit design in line with a rational synthetic biology workflow. To this end we have developed and implemented Biopart Assembly Standard for Idempotent Cloning (BASIC). This enables modular DNA assembly for rapid prototyping of new biological systems. The system can rationally programme the expression of individual proteins (Fig. 1). This is essential for the development of high performance gene circuitry and has been extensively implemented in the development of the gene circuits responsible for controlling the genetic mutator system.



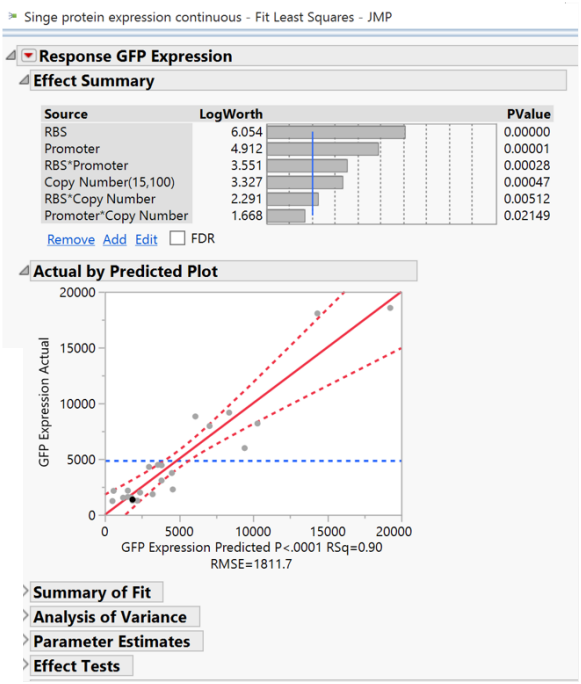


Figure 1. Modular assembly enables specific assembly of different strength promoters with different strength RBS sequences to drive expression of a protein, copy number can also be controlled. Fitting of data from the above circuit demonstrates a strong correlation of output that can be used as a model to optimise expression of circuitry

Mutator system

We have demonstrated the viability of targeting mutations to a specific gene inside a living cell without significantly impacting the fitness of the host. The overall outline is demonstrated in Fig. 2: a mutator is targeted to a specific gene and this leads to deamination of cytosine to uracil. Normally this is repaired, but if this is blocked then these mutations accumulate in the gene. This is demonstrated in Fig. 2b, where the mutator is targeted to a GFP gene, accumulation of mutations leads to loss of fluorescence due to mutations in the gene. If the mutator protein is not targeted the loss of fluorescence is associated with off-target mutations. These effects can be observed in the DNA sequence (not shown). The distribution of mutations that accumulate over 24h can be seen in Fig. 2c, where it is apparent that we can accumulate multiple mutations with a mean frequency of 1-2 mutations per gene. This is ideal for the purposes of directed evolution: if the mutation rate were too high then we would lose all function, and if too low no evolution would occur.

Work in this project has further demonstrated methods to broaden the mutational diversity being sampled and to develop *in vivo* gene circuitry to create a self-evolving self-selecting system.

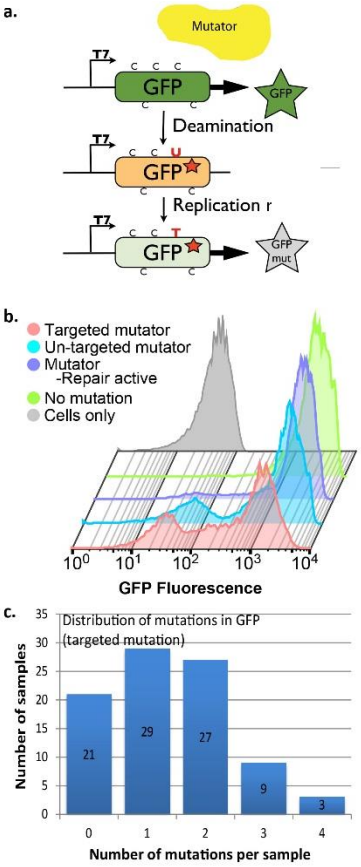


Figure 2. Function of *in vivo* mutator. **a**. overview of scheme. **b**. mutations directed at GFP lead to loss of fluorescence. **c**. number of mutations accumulated per gene over 24h.