This article appeared in Spanish in the journal of the Spanish society of biochemistry and molecular biology:
Isalan, M., Schaerli, Y., Biología sintética en busca de andamios, *SEBBM* 2012, 171, 16-20
http://www.sebbm.com/dossier171_Isalan_Schaerli.htm

Synthetic biology in search of a scaffold

Mark Isalan and Yolanda Schaerli

EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG) and UPF, 08003 Barcelona, Spain.
Email: isalan@crg.es

Abstract
Synthetic biology has recently become a fashionable scientific term, attracting interest from high-profile journals, funding agencies and even the wider media. Despite this, the pace of progress has been rather slow because we lack the tools to engineer synthetic biological systems in any generic manner. Each project is currently developed painstakingly, on a case-by-case basis. Taking a step back, we argue that the forerunners of synthetic biology were the fields of recombinant molecular biology and protein engineering. The latter routinely generates new proteins with desired functions, via a mixture of rational and combinatorial engineering. Here we discuss how the successful concepts from this field could be applied to synthetic biology. The first key issue that needs to be addressed is how to make the conceptual equivalent of a 'scaffold' for network engineering. In protein engineering, versatile scaffolds such as antibodies or zinc fingers tolerate extensive targeted mutations and thus provide efficient sources of new functions. Second, we lack foolproof ways to predict the outcome of combining multiple positive and negative interactions, for example to integrate multiple activators and repressors on the same gene promoter. If these issues are overcome, then perhaps we can truly start to engineer gene networks reliably.

Introduction

Synthetic biology is an emerging research field that applies engineering approaches to biological systems. A browse through scientific journals these days quickly uncovers headlines such as "2011 - Year of Synthetic Biology?" [1]. Such editorials invariably cite Craig Venter's work on chemically synthesising and transplanting a Mycoplasma genome in a recipient cell [2]. The work, which made news around the world along the lines of 'Creation of synthetic life', has both captured and worried the public imagination. This has undoubtedly contributed to creating a peak of interest in synthetic biology (**Fig. 1**).

However, it is well-known that synthetic biology faces serious technical difficulties. For example, an article entitled "Five hard truths for synthetic biology" [3] recently outlined several problems in the field, including that biological 'parts' [4] do not necessarily fit together like lego and that it requires many person-years to make a pathway employing a dozen steps [5]. Rather than dwelling on the negative, in this essay we will consider ways to overcome these issues, discussing what might

need to be done, both theoretically and practically, to increase the reliability of gene network engineering strategies.

Recycling the same parts

When looking at high-profile synthetic biology papers one sees that the same few biological components are reused over and over again. For example, the negative feedback loop (TetR) [6], the toggle switch (LacI, λ cI, TetR) [7], the repressilator (TetR, LacI, λ cI) [8], combinatorial networks (TetR, LacI, λ cI) [9], the Atkinson oscillator (LacI, GlnA, GlnG, NRI) [10], the bistable attractor-selection switch (TetR, LacI) [11] and the band-detect sender-receiver network (LacI, λ cI, LuxR, LuxI) [12], all use related components. Regarding the last example, the majority of the existing synthetic cell-cell communication circuits, have been engineered by using elements from quorum sensing systems [12,13,14,15,16]. Many further examples could be cited (reviewed in [17]), but the general observation is that most successful designs are recycling the same parts.

Why are the same parts used so often? Part of the answer lies in the fact that they are well characterised, and have relatively modular functions (e.g. adding lac operator sites to a new promoter can make it susceptible to repression, even though even this still can need 'tweaking' [18]). Other modular components, such as T7 or SP6 RNA polymerases, can also have their activities transplanted from one construct to the next, by adding relatively short target sequences. Consequently, these lend themselves to flexible synthetic re-engineering [19,20]. To expand this repertoire of components, MIT recently set up a Registry of Standard Biological Parts (http://partsregistry.org/) based on the concept of a 'biobrick' [4]. This system aims to standardise cloning, and is mainly used by the International Genetically Engineered Machine (iGEM) competition [21], where undergraduate students design and build synthetic biology systems. Despite being a useful resource of functional DNA sequences, the relative lack of papers derived from this repository attests to the fact that biological components often do not behave as predicted when simply slotted together. The devil is often in the detail of small context-dependent effects.

Protein engineering - the precursor of synthetic biology

The idea of synthetic biology has its roots in molecular cloning and recombinant DNA technologies, where genetic components such as transcription promoters and coding regions are now routinely combined to make protein expression constructs or other new plasmids [22]. The idea of making synthetic gene networks is seemingly just one level of complexity higher, simple harnessing the appropriate recombinant contructs to make networks.

However, because of the unpredictable effects mentioned above it is not always straightforward to combine parts or mutations together to generate new functions. One field of biological engineering which is now relatively mature, and where new functional constucts are routinely made, is protein engineering. New proteins are engineered, often using structural information and an element of rational design [23,24], but also through screening or selecting from large randomised combinatorial libraries. Methods of directed evolution that link the phenotype to the genotype, and add an element of Darwinian selection, are particularly powerful [25]. In selection systems such as phage display [26,27], millions of randomised antibody fragments can be displayed on the surface of phages. The one-in-a-million variant with a desired binding activity can bind its target, survive washing steps, and infect new host cells. Thus survival is linked to the new desired function.

Scaffolds for network engineering

One issue preventing a straightforward application of selection to network engineering is the lack of an obvious generic scaffold upon which to randomise and link together the network components. Scaffold has a specific meaning here - the key property being that the scaffold must generally withstand extensive mutations without losing its overall functionality. In protein engineering, a typical scaffold such as a zinc finger can be simultaneously mutated at nearly every residue (except for key folding residues) and it will still fold correctly [28]. Thus, introduced library randomisations have a greater chance of generating new functions than if the scaffold were 'fragile' with respect to mutation [29]. Interestingly, there exists formal theoretical work underpinning this effect; evolvability is greater in robust systems such as protein motif scaffolds [30].

The major question is therefore how to define a robust network scaffold for synthetic biology purposes. The vector architecture is not so much of a problem; standard plasmid cloning with restriction enzyme sites to insert randomised library cassettes should be suitable [9]. The question is where to introduce the combinatorial diversity. For instance, when engineering synthetic transcription networks, it would be wasteful to randomise each transcription factor residue (TF); most mutants would be non-functional or similarly-functional, and the library size would quickly be too big to screen. Rather, it would be better to make rational mutations around the TF DNA binding site to modulate affinity. Thus, targeted mutations would provide functional diversity in a relatively small, easy-to-handle library. Modulating parameters such as half-life and binding site copy-number might also be varied as part of a robust, yet randomised, network scaffold. Many practical options are possible (**Fig. 2**).

A second issue is how to add a Darwinian selection pressure to the randomised network scaffold. Combinatorially-randomised gene networks have been built [9], but these have had to be screened one-by-one rather than applying a selection pressure to select the rare variant(s) with the desired properties. There are several possibilities for turning a network scaffold into a workable selection system. One solution is to link a survival gene, such as an antibiotic resistance gene or metabolic mutant complementation gene to the network [11]. Perhaps even more elegant is the use of dual selection and counterselection markers, whose expression either enable cell survival or induce cell death under particular conditions [31,32,33].

We and others are implementing these markers in order to select networks with desired properties. For example, the selection pressure for a switch, in response to a particular input, would be positive selection for expression under "ON" conditions and negative selection under "OFF." One could imagine many rounds of conditional selection to obtain more complex input-output behaviours. Some selection pressures would be relatively straightforward, such as for generating switches [32]. However, for making complicated patterns such as the regular spots or stripes found in Turing or Gierer-Meinhardt patterns [34,35], the selection system would have to be rather more complicated. If one could design a robust library scaffold, based on some positive and negative feedback motifs [36], one could perhaps imagine an automated system screening for the desired behaviour (**Fig. 3**). However, this remains a truly challenging goal for the synthetic biology community.

Predicting the effects of integrating multiple inputs

One major issue that synthetic biologists come across is how to encode multiple inputs on the same node of a network. For instance, when looking at the logic of developmental gene networks, they often have more than one positive or negative interaction controlling each gene, including feedbacks. Although we can understand some of the logic of these, by working backwards and reverse engineering [37], the forward engineering approach of synthetic biology is much harder to implement. What happens when two synthetic activators and two repressors control a gene? This is often hard to predict and effects such as competition or non-competition can have drastically different outcomes [38]. To an extent, if we use combinatorial strategies with the correct selection pressures, some of these issues should resolve themselves, even if we do not necessarily understand the resulting networks *a priori*. However, really understanding network logic from an engineering perspective would be a great advance.

Distributed networks

Even if we were to understand how to build complicated regulatory regions, the issue of a limited number of well-characterised components would still lead to running out of unique components when building larger networks. A recent study avoided both the problems of limited availability of building blocks and of integrating multiple inputs in a rather elegant way. By distributing very simple logic gates combinatorially, in heterogenous populations of cells, each cell carried out a simple function. However, the whole consortia carried out far more complicated distributed computational tasks [39]. A similar idea was to utilise multiple simple-function colonies to communicate with each other to make more complex logic gates [40]. Perhaps this concept represents an alternative network scaffold and is truly the future of synthetic biology.

In this brief review, we have examined the roots of synthetic biology in protein engineering, discussed ways to implement combinatorial selection concepts, and have touched upon the remaining engineering problems we face. We are far from having a standard robust engineering method for biological systems, but perhaps part of the fun is simply in trying to build in order to understand.

Acknowledgments

The group of MI is funded by the European Research Council, FP7-ERC-201249-ZINC-HUBS, the MEC-EMBL agreement and Ministerio de Ciencia e Innovación Grant MICINN BFU2010-17953. YS is an owner of a SNSF fellowship.

References

1. Schaefer U (2011) 2011 - Year of Synthetic Biology? BiotecVisions: A1; DOI: 10.1002/biot.201000456.
2. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329: 52-56.
3. Kwok R (2010) Five hard truths for synthetic biology. Nature 463: 288-290.
4. Knight TF (2003) Idempotent Vector Design for Standard Assembly of BioBricks. Tech rep, MIT Synthetic Biology Working Group Technical Reports.
5. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440: 940-943.
6. Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. Nature 405: 590-593.
7. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403: 339-342.
8. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403: 335-338.
9. Guet CC, Elowitz MB, Hsing W, Leibler S (2002) Combinatorial synthesis of genetic networks. Science 296: 1466-1470.
10. Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. Cell 113: 597-607.
11. Kashiwagi A, Urabe I, Kaneko K, Yomo T (2006) Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. PLoS One 1: e49.
12. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. Nature 434: 1130-1134.
13. Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. Proc Natl Acad Sci U S A 101: 6355-6360.
14. Brenner K, Karig DK, Weiss R, Arnold FH (2007) Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. Proc Natl Acad Sci U S A 104: 17300-17304.
15. You L, Cox RS 3rd, Weiss R, Arnold FH (2004) Programmed population control by cell-cell communication and regulated killing. Nature 428: 868-871.
16. Bulter T, Lee SG, Wong WW, Fung E, Connor MR, et al. (2004) Design of artificial cell-cell communication using gene and metabolic networks. Proc Natl Acad Sci U S A 101: 2299-2304.
17. Michalodimitrakis K, Isalan M (2009) Engineering prokaryotic gene circuits. FEMS Microbiol Rev 33: 27-37.
18. Weiss R, Basu S (2002) The Device Physics of Cellular Logic Gates. First Workshop on Non-Silicon Computing Feb 2002: 1-8.
19. Noireaux V, Bar-Ziv R, Libchaber A (2003) Principles of cell-free genetic circuit assembly. Proc Natl Acad Sci U S A 100: 12672-12677.
20. Isalan M, Lemerle C, Serrano L (2005) Engineering gene networks to emulate *Drosophila* embryonic pattern formation. *PLoS Biology* ***In press***.
21. Goodman C (2008) Engineering ingenuity at iGEM. Nat Chem Biol 4: 13.

22. Sambrook J, Fritsch, E. F., Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press
23. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302: 1364-1368.
24. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, et al. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science 332: 816-821.
25. Griffiths AD, Tawfik DS (2003) Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. Embo J 22: 24-35.
26. Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228: 1315-1317.
27. Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR (1994) Making antibodies by phage display technology. Annu Rev Immunol 12: 433-455.
28. Michael SF, Kilfoil VJ, Schmidt MH, Amann BT, Berg JM (1992) Metal binding and folding properties of a minimalist Cys2His2 zinc finger peptide. Proc Natl Acad Sci U S A 89: 4796-4800.
29. Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459: 668-673.
30. Wagner A (2005) Robustness, evolvability, and neutrality. FEBS Lett 579: 1772-1778.
31. Nomura Y, Yokobayashi Y (2007) Dual selection of a genetic switch by a single selection marker. Biosystems 90: 115-120.
32. Muranaka N, Sharma V, Nomura Y, Yokobayashi Y (2009) An efficient platform for genetic selection and screening of gene switches in Escherichia coli. Nucleic Acids Res 37: e39.
33. Tashiro Y, Fukutomi H, Terakubo K, Saito K, Umeno D (2011) A nucleoside kinase as a dual selector for genetic switches and circuits. Nucleic Acids Res 39: e12.
34. Turing AM (1952) The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London B 237: 37-72.
35. Gierer A, Meinhardt H (1972) A theory of biological pattern formation. Kybernetik 12: 30-39.
36. Isalan M (2009) Gene networks and liar paradoxes. Bioessays 31: 1110-1115.
37. Jaeger J, Surkova S, Blagov M, Janssens H, Kosman D, et al. (2004) Dynamic control of positional information in the early Drosophila embryo. Nature 430: 368-371.
38. Munteanu A, Constante M, Isalan M, Sole RV (2010) Avoiding transcription factor competition at promoter level increases the chances of obtaining oscillation. BMC Syst Biol 4: 66.
39. Regot S, Macia J, Conde N, Furukawa K, Kjellen J, et al. (2011) Distributed biological computation with multicellular engineered networks. Nature 469: 207-211.
40. Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. Nature 469: 212-215.
41. Meinhardt H, Gierer A (2000) Pattern formation by local self-activation and lateral inhibition. Bioessays 22: 753-760.

Figures

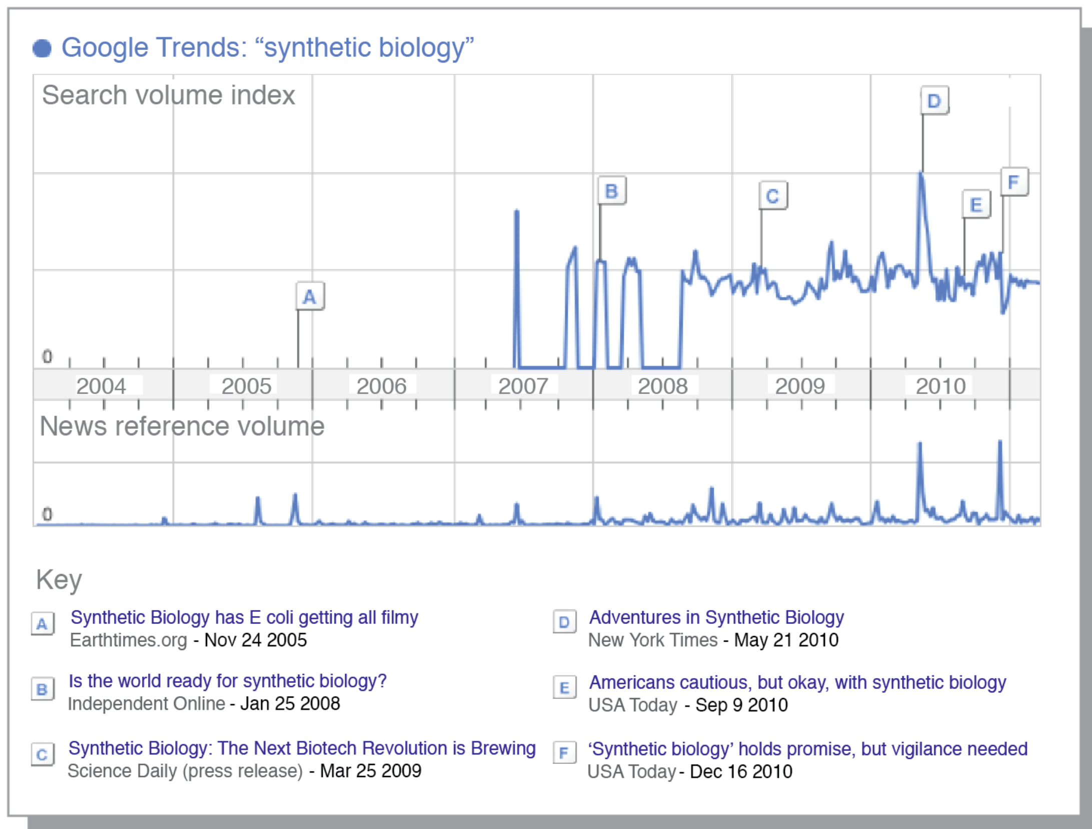

Figure 1. Google Trends view of "synthetic biology". Search engine volume (above) and news reference volume (below) show sporadic, but increasing, public interest over time. Until the latter half of 2008, the search volumes do not regularly exceed the detection threshold of the Google Trends algorithm, after which time they stay relatively steady. The largest peak (D) corresponds to the day after the publication of a highly-publicised paper in *Science* [2] about the chemical synthesis and transplantation of a Mycoplasma bacteria genome.
Source: Google Trends (http://www.google.com/trends).

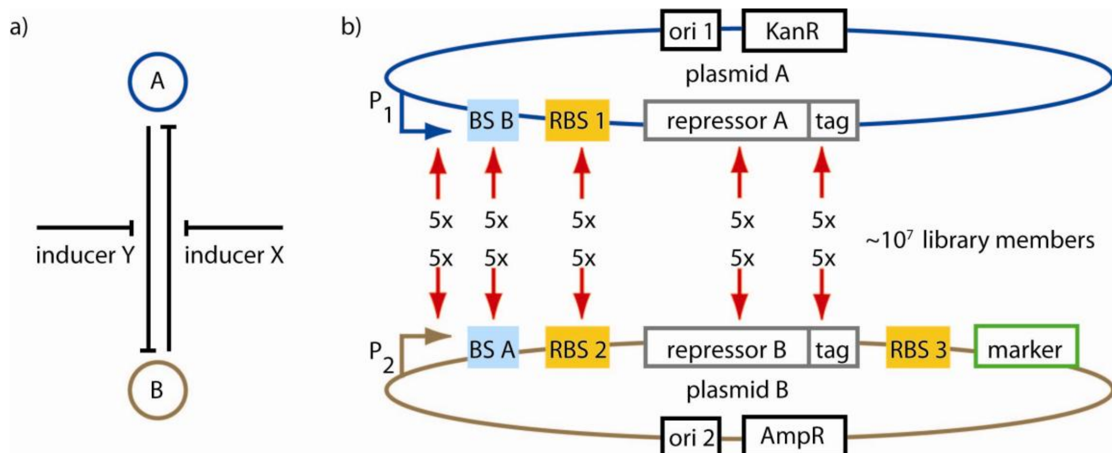

Figure 2. Scaffolds for network engineering. a) Scheme of a genetic toggle switch constructed from a mutually inhibitory two-node network. Inducers are used to flip between the two stable states [7]. b) The toggle switch could be encoded on two plasmids, each carrying an origin of replication (ori), an antibiotic resistance gene, a promoter (P), a repressor with a degradation tag, a repressor binding site (BS) and ribosomal binding sites (RBS). This scaffold could be randomised at the positions indicated by the red arrows. Five variations at each position would lead to a combinatorial library with almost 10^7 members. A marker could be used to find the networks with the desired behaviour in consecutive selections in the "ON" and "OFF" states [31,32,33].

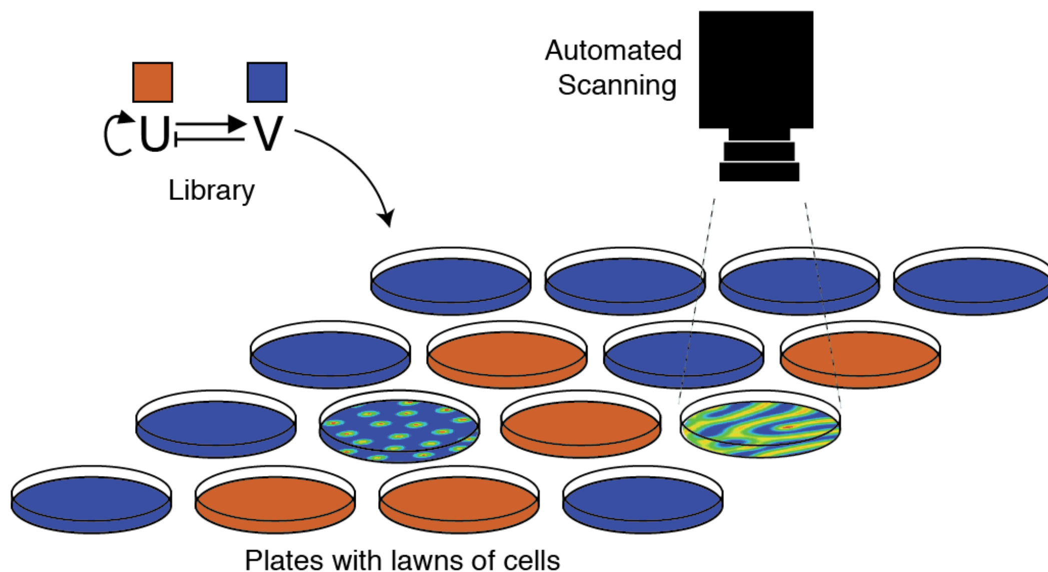

Figure 3. Automated selection of spatial patterning systems from a network scaffold. The schematic shows a thought-experiment on how one might screen a combinatorial library scaffold for a Gierer-Meinhardt system [41]. Thousands of randomised candidates might have to be tested to find the correct behaviour. The library would comprise variants of an activator (U; red) and an inhibitor (V; blue) which would communicate local, non-linear activation and long range inhibition signals to other cells. Randomisations would need to encode parameter variations such as activation and inhibition strengths, component half-lives, secretion rates and diffusion rates. By plating library members on dishes or multiwell plates (here, one per plate) thousands of randomised parameter sets might be screened for potential patterning behaviour.