

## Harnessing nature's toolbox: regulatory elements for synthetic biology

Patrick M. Boyle and Pamela A. Silver

*J. R. Soc. Interface* 2009 **6**, S535-S546 first published online 4 March 2009  
doi: 10.1098/rsif.2008.0521.focus

---

### References

[This article cites 99 articles, 24 of which can be accessed free](#)

[http://rsif.royalsocietypublishing.org/content/6/Suppl\\_4/S535.full.html#ref-list-1](http://rsif.royalsocietypublishing.org/content/6/Suppl_4/S535.full.html#ref-list-1)

[Article cited in:](#)

[http://rsif.royalsocietypublishing.org/content/6/Suppl\\_4/S535.full.html#related-urls](http://rsif.royalsocietypublishing.org/content/6/Suppl_4/S535.full.html#related-urls)

### Rapid response

[Respond to this article](#)

[http://rsif.royalsocietypublishing.org/letters/submit/royinterface;6/Suppl\\_4/S535](http://rsif.royalsocietypublishing.org/letters/submit/royinterface;6/Suppl_4/S535)

### Subject collections

Articles on similar topics can be found in the following collections

[synthetic biology](#) (31 articles)

[bioengineering](#) (50 articles)

[systems biology](#) (67 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

---

To subscribe to *J. R. Soc. Interface* go to: <http://rsif.royalsocietypublishing.org/subscriptions>

---

## REVIEW

# Harnessing nature's toolbox: regulatory elements for synthetic biology

**Patrick M. Boyle and Pamela A. Silver\****Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, WAB 536, Boston, MA 02115, USA*

Synthetic biologists seek to engineer complex biological systems composed of modular elements. Achieving higher complexity in engineered biological organisms will require manipulating numerous systems of biological regulation: transcription; RNA interactions; protein signalling; and metabolic fluxes, among others. Exploiting the natural modularity at each level of biological regulation will promote the development of standardized tools for designing biological systems.

**Keywords:** biological regulation; natural modularity; standardized tools

## 1. INTRODUCTION

By analogy to other engineering disciplines, synthetic biologists aim to construct complex 'devices' as assemblies of well-defined modular parts. The synthetic biology approach is predicated on the idea that modular biological elements exist, which can be repurposed or modified for the construction of new devices (Drubin *et al.* 2007).

The unique capabilities of biological organisms make them an attractive target for engineering. Living cells are self-replicating, self-repairing chemical factories that can, in principle, be reconfigured by altering their DNA blueprint. The development of DNA synthesis technology has allowed researchers to create any DNA sequence they desire, even entire genomes (Gibson *et al.* 2008). The ability to synthesize DNA promises to make the entire repertoire of known biological diversity available to synthetic biologists.

Despite this power, rationally designed biological devices rarely function entirely as predicted, and are often less robust than natural systems. With the price of DNA synthesis falling and worldwide interest in synthetic biology rising, the progress of synthetic biology has not been limited by funding or ambition. Instead, progress made in the last decade of synthetic biology research has revealed the unique difficulties of engineering living systems. In particular, endogenous regulatory systems often interfere with the function of synthetic biological devices (Arkin & Fletcher 2006).

Improving our ability to engineer biology will require an improved ability to predictably regulate biological

systems. Natural biological systems are regulated at many levels: transcription; RNA processing; translation; protein–protein interactions; and protein–substrate interactions all exert control on cellular processes. The spatial and temporal organization of these myriad control systems is vital to their proper function. The fact that RNA interference, an integral control mechanism in eukaryotes, was only discovered in the late 1990s demonstrates that we are still discovering fundamental building blocks of biological systems (Fire *et al.* 1998).

The identification of biological modules with regulatory functions will allow the construction of more complex devices. For decades, promoters have been used to control the expression of recombinant genes (Reznikoff *et al.* 1969; Casadaban 1975). Similarly, modular protein elements, such as zinc fingers, allow the design of novel protein interactions (Drubin *et al.* 2007). Harnessing biological modularity provides insight into engineering biological regulation. In this review, we will explore recent advances in the field of synthetic biology, with emphasis on the development of biological modules for the regulation of synthetic devices.

## 2. TRANSCRIPTIONAL SIGNAL PROCESSING

It has long been understood that transcription factors target specific DNA elements to control gene expression (Jacob & Monod 1961). Promoters are modular DNA elements that can be used to drive the transcription of a gene, a property that has often been exploited for biological research (Casadaban 1975). As a well-understood mechanism of biological regulation, transcriptional control has been a feature of many synthetic devices.

\*Author for correspondence (pamela\_silver@hms.harvard.edu).

One contribution to a Theme Supplement 'Synthetic biology: history, challenges and prospects'.

Table 1. Basic two-input logic gates. (Each gate produces a single true or false output based on the state of two inputs (Irving 1961). F, 'false'; T, 'true'; NAND, 'not AND'; NOR, 'not OR'.)

| gate | input |   | output |
|------|-------|---|--------|
|      | A     | B |        |
| AND  | F     | F | F      |
|      | T     | F | F      |
|      | F     | T | F      |
| OR   | T     | T | T      |
|      | F     | F | F      |
|      | T     | F | T      |
| NAND | F     | T | T      |
|      | T     | T | T      |
|      | F     | F | T      |
| NOR  | T     | T | F      |
|      | F     | F | T      |
|      | T     | F | F      |
|      | F     | T | F      |
|      | T     | T | F      |

Transcriptional control can be exploited for signal integration, i.e. processing multiple inputs and producing well-defined outputs. The operations that signal integration elements perform can be represented as logic gates. Boolean logic gates integrate two or more input signals and output one of two values (Irving 1961). For example, a two-input AND gate returns an output of 'true' when both inputs are true, and returns 'false' when either or both of the inputs are false (table 1). Transcriptional devices can serve as logic gates by producing outputs based on the state of input promoters; an AND gate would produce a certain output only if both input promoters were induced.

Ideally, signal integration elements and promoter inputs would be functionally separate in a synthetic device. For example, replacing an ara promoter with a lac promoter should modify the specificity of the device from arabinose to lactose without affecting how the rest of the device functions. Towards this end, a 'modular' AND gate was developed, allowing the modification of the promoter inputs and the gate output, leaving the core logic module intact (figure 1a; Anderson *et al.* 2007). Two promoters receive input to the device: one drives the expression of a T7 polymerase containing two amber mutations (T7ptag), while the other directs expression of the amber-suppressing tRNA supD. When both promoters are activated, supD expression allows the proper translation of T7ptag, which in turn transcribes a reporter gene with a T7 promoter. The original input promoters were induced by salicylate and arabinose for the expression of supD and T7ptag, respectively, with green fluorescent protein (GFP) as the output reporter.

This supD/T7ptag AND gate was intended to be modular, in that the device functions as an AND gate regardless of the input promoters or the output protein. To confirm modular function, the device inputs were replaced with magnesium-repressed and AI-1-inducible

(LuxR quorum sensor) promoters, and the output was replaced with the invasin gene that triggers the invasion of mammalian cells. Although the initial reconfiguration of the device did not perform as an AND gate, tuning expression by varying the ribosome-binding sites of each promoter restored functionality. The new device allows *Escherichia coli* cells to invade mammalian cells in the case that exogenous magnesium is absent and the AI-1 signal is present. Importantly, the signal integration portion of the device was unmodified, and only minimal tuning was required to restore logic gate function after vastly altering the device inputs and outputs.

Producing high-fidelity outputs is as important as proper signal integration. Ideally, a synthetic device will remain in a persistent output state until the input state is changed. Genetic toggle switches have been demonstrated, incorporating transcriptional feedback loops to ensure that the device remains in one of two stable states (figure 1b; Gardner *et al.* 2000). Devices that feature oscillatory outputs based on transcriptional feedback have also been constructed (Elowitz & Leibler 2000; Stricker *et al.* 2008). 'Memory devices' feature a persistent output in response to a transient input; they are activated by a specific input and remain active after the input is removed (Ajo-Franklin *et al.* 2007).

A memory device constructed in the yeast *Saccharomyces cerevisiae* transmits a fluorescent response for many generations following a transient stimulus (figure 1c; Ajo-Franklin *et al.* 2007). The device features two fluorescently labelled genes: a red fluorescent protein (RFP)-tagged sensor gene, which, when expressed, activates the transcription of a yellow fluorescent protein (YFP)-tagged auto-feedback gene. The auto-feedback gene possesses the same transcriptional activator module as the sensor gene, thus the auto-feedback gene activates its own expression following the transient stimulus. Experimentation and quantitative models of the device suggested that tight regulation of transcription was required to maintain memory. Leaky promoters would trigger activation of the device in the absence of inducer, and the concentration of the auto-feedback protein following induction needs to remain above the threshold for maintaining memory following cell division. The final device exhibits the required bistability, maintaining the YFP signal even after the inducer was removed.

### 3. RNA SIGNAL PROCESSING

RNA is a versatile molecule with a variety of roles in cellular functions. The ability of RNA to transmit genetic information as well as conduct enzymatic catalysis has led to the hypothesis that early life used RNA in lieu of DNA and proteins (Bartel & Unrau 1999). Regulatory RNAs can exert control by antisense binding with mRNA, by conformational changes that disrupt transcription or translation or by ribozyme activity (Isaacs *et al.* 2006; Saito & Inoue 2008). In the context of synthetic biology, many devices have been constructed that take advantage of RNA-mediated regulation.

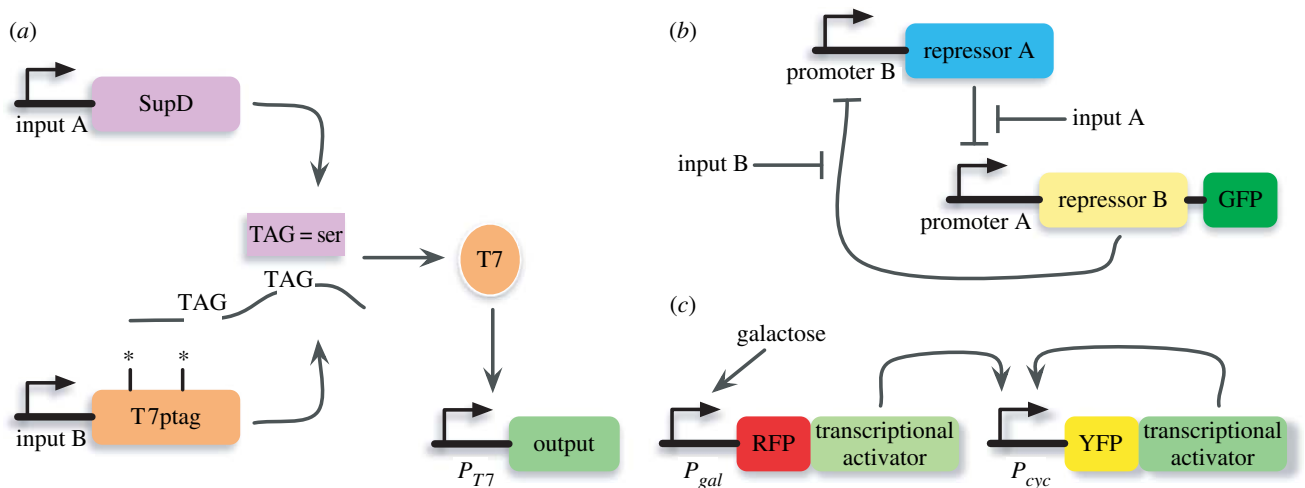


Figure 1. Synthetic devices based on transcriptional logic. (a) A modular AND gate. Two input promoters control the expression of the SupD tRNA suppressor and a T7 polymerase with two TAG stop codons, denoted by asterisks. Activation of both input promoters allows SupD to suppress the early stop codons by inserting serine, and the functional T7 protein activates expression of the reporter (Anderson *et al.* 2007). (b) A genetic toggle switch. Bistability is achieved via two mutually repressing genes. Addition of inducer molecules allows switching between stable states (Gardner *et al.* 2000). (c) A cellular memory device. Galactose induces expression of an RFP-tagged transcriptional activator that triggers expression of a YFP-tagged reporter. The YFP-tagged protein then activates itself, maintaining YFP expression (Ajo-Franklin *et al.* 2007).

Riboswitches are regulatory RNAs that bind small molecules or peptides via a specific aptamer domain (Patel *et al.* 1997). Natural riboswitches can be found in the 5' untranslated region (UTR) of mRNA, where ligand binding triggers a conformational change that can negatively or positively modulate transcription or translation (Mandal & Breaker 2004; Tucker & Breaker 2005; Isaacs *et al.* 2006). Catalytic RNAs such as hammerhead ribozymes (HHRz; Khvorova *et al.* 2003) can even be attached to riboswitches to construct ligand-dependent ribozymes (Win & Smolke 2008).

In the context of synthetic devices, riboswitches allow promoter-independent control of gene expression. For example, RNA 'antiswitches' have been engineered in *S. cerevisiae* (Bayer & Smolke 2005). An antiswitch is an RNA molecule containing a ligand-binding aptamer domain and an antisense regulator domain. Antiswitches can be designed to activate or repress translation in response to ligand binding. 'Off antiswitches' feature an antisense domain that is stabilized as a stem loop in the absence of ligand, permitting translation of the targeted mRNA. Ligand binding to the aptamer domain triggers a conformational change, exposing the antisense domain and repressing translation. 'On antiswitches', on the other hand, feature an antisense domain that forms a stem loop when the ligand is bound, releasing the mRNA target from repression. Importantly, the aptamer and antisense domains are functionally modular; exchanging a theophylline aptamer domain for a tetracycline aptamer domain changes the specificity of an antiswitch to tetracycline without altering the antisense targeting.

The modularity of riboswitches has been exploited to construct logic gates (Win & Smolke 2008). Riboswitch logic gates incorporate 'sensor', 'transmitter' and 'actuator' domains (figure 2a). The sensor domain consists of a ligand-dependent aptamer, and the conformational change triggered by ligand binding alters the conformation of the transmitter domain. Attaching

the sensor and transmitter to an actuator domain, in this case an HHRz, confers ligand-dependent ribozyme activity. The devices were embedded in the 3' UTR of mRNA transcripts, such that HHRz activation triggers mRNA cleavage leading to decreased expression. More complex devices were constructed by linking pairs of sensor domains, linked by transmitters, to a single actuator. Rearrangement of dual sensor domains relative to an actuator domain yielded AND, NOR, NAND and OR gates (figure 2b).

Other devices were also constructed by placing two sensor-actuator pairs in series on an mRNA. One such device is a 'bandpass filter', so named because expression of the reporter protein only occurs at intermediate inducer concentrations, with low and high inducer concentrations leading to mRNA degradation. This was accomplished by including a 'buffer gate' that prevents ribozyme cleavage in the presence of theophylline as well as an 'inverter gate' on the same mRNA that activates ribozyme cleavage in the presence of theophylline.

Boolean logic gates have also been designed with small interfering RNAs (siRNAs; Rinaudo *et al.* 2007). In these devices, inputs trigger siRNA expression, with each siRNA targeting a specific mRNA. The targeted mRNA molecules each encode a repressor protein such as LacI that repress the expression of a fluorescent protein reporter. For example, an AND gate consists of two mRNAs driven by the same promoter, but with different 3' UTR targets, A and B. Each mRNA encodes LacI or LacI-KRAB to repress the expression of the reporter protein. The presence of both siRNA A and siRNA B is required to knock down the mRNAs and allow reporter expression. An OR gate was constructed by expressing a single mRNA for the repressor protein, with two 3' UTR targets, such that siRNA A or siRNA B is sufficient for knockdown.

This siRNA-based logic allowed the construction of devices that evaluate complex expressions such as

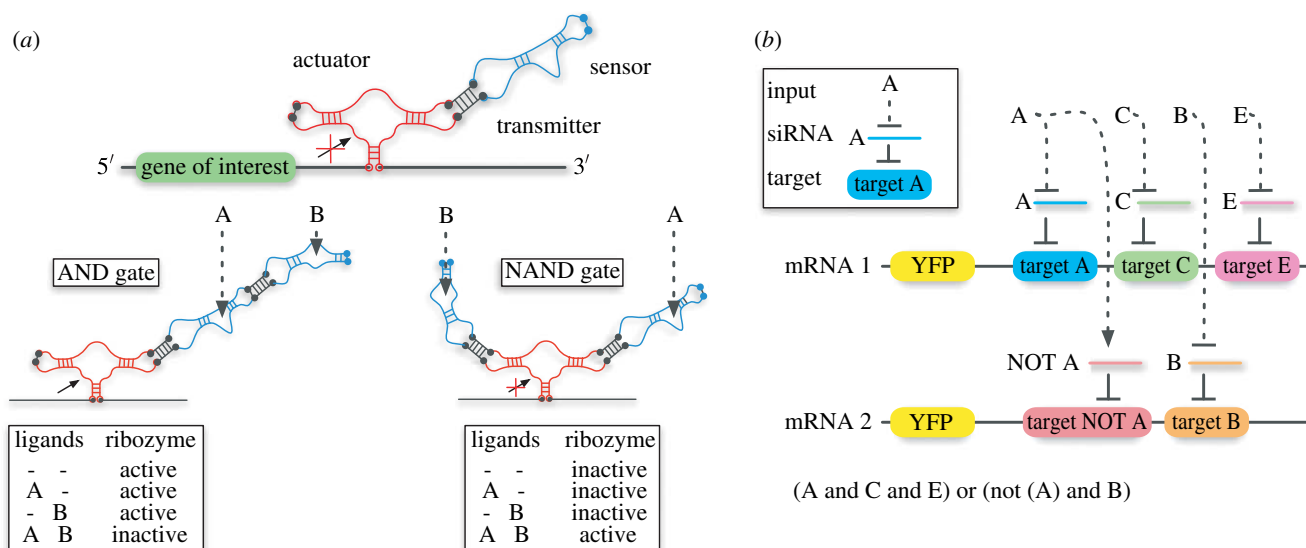


Figure 2. RNA-based logic gates. (a) Modular ribozymes consist of 'actuators' (red) that contain the HHRz ribozyme; 'transmitters' that transmit conformational changes to the actuator; and 'sensors', the RNA aptamers that bind ligands. An arrow next to the actuator denotes that the ribozyme is active in the absence of ligand. A crossed out arrow denotes that the ribozyme is inactive in the presence of ligand. Multiple sensor domains can be added to either end of the actuator. An example AND gate and NAND gate are shown, multiple actuators can also be incorporated in series on a single mRNA (Win & Smolke 2008). (b) Evaluation of complex logic in an RNAi-based device. Input A represses transcription of the 'A' siRNA and activates transcription of the 'NOT A' siRNA. Inputs C, B and E repress transcription of their respective siRNAs. Each siRNA downregulates the expression of its target mRNA if expressed (Rinaudo *et al.* 2007).

'(A AND B AND C) OR (D AND E)' and '(A AND C AND E) OR (NOT(A) AND B)' (figure 2b), with A, B, C, D and E representing different inputs. Every possible input combination was experimentally verified, with the second device yielding the only incorrect output, returning false in the case that A and E were true and B and C were false (Rinaudo *et al.* 2007).

RNA logic devices demonstrate that RNA can regulate complex synthetic networks. Both the riboswitch and RNAi-based devices take advantage of antisense base pairing to confer specificity between synthetic RNA inputs and their mRNA targets. Modular riboswitches may prove to be portable to a wide range of host species, as RNA folding and ribozyme activity occur independently of the host cellular machinery. Complex RNA regulatory devices in natural systems, such as the *S*-adenosyl-methionine/adenosylcobalamin NOR gate in *Bacillus clausii*, demonstrate that RNA aptamers are an effective means of responding to changes in metabolite concentrations (Nahvi *et al.* 2002; Stoddard & Batey 2006).

In higher eukaryotes, introns appear to be an important regulatory element at the mRNA level. Approximately 95 per cent of the average human gene is made up of introns (Lander *et al.* 2001; Venter *et al.* 2001). Both introns and exons must be transcribed; introns are removed from the mRNA prior to translation. In the development of multicellular organisms, one suggested role for introns is in regulating the timing of gene expression, as intron processing increases the time between transcription and translation. Time delay coupled with autoinhibition can produce oscillations, and oscillatory gene expression is often observed in development (Swinburne & Silver 2008).

To study the role of intron length in development, an intron-containing device was constructed (Swinburne

*et al.* 2008). The device contained an intron, a fluorescent reporter and the Tet repressor, which negatively inhibited its own expression. The device demonstrated pulses of expression in mammalian cells, and the frequency of the pulses was dependent on intron length. As in the endogenous genes of higher eukaryotes, introns may provide an additional layer of control to synthetic devices.

#### 4. SYNTHETIC PROTEIN SIGNALLING

Proteins do not interact the way components on a circuit board do. Identical components such as resistors can be placed on the same circuit board and not interfere with one another, because the wiring keeps them connected only to components they were intended to interact with. Proteins, however, can diffuse throughout the cellular compartment that contains them, interacting with any suitable binding partners.

Evolution has found a solution to orthogonal signalling that still allows cells to use the same protein components for multiple processes (Bhattacharyya *et al.* 2006). Classes of proteins, such as kinases, share a common mechanism of action but can act on a variety of targets. This is often achieved by varying combinations of adapter domains and effector domains. Signalling protein interactions are often mediated by events such as phosphorylation that change binding affinities and 'rewire' the network (Pawson 2007).

Prokaryotic two-component signalling systems provide a simple model system for studying protein interactions. Canonical two-component systems consist of a membrane-bound receptor with ligand-dependent histidine kinase (HK) activity and a response regulator (RR) protein, usually a transcription factor (West & Stock 2001).



The rational design of two-component signalling systems has recently been demonstrated (Skerker *et al.* 2008). Multiple sequence alignments of HK and RR pairs identified residues that covaried, representing interacting partners. Modification of as few as three interacting residues in the HK EnvZ (an *E. coli* osmolarity sensor) switched the EnvZ HK phosphorylation specificity from its cognate RR OmpR to the non-cognate RR RstA, which is not normally induced by osmotic changes. Additionally, the RR specificity of EnvZ was switched to that of CC1181, a *Caulobacter crescentus* sensor protein. This research establishes a protocol for the rational design of two-component systems, as well as validating methods of HK–RR interaction prediction. The existence of two-component systems in eukaryotes (Saito 2001) implies that novel HK–RR pairs could augment eukaryotic devices as well.

Eukaryotic signalling cascades are usually more complex than two-component systems, but eukaryotic signalling proteins can still be reprogrammed to accept new inputs. For example, a guanine nucleotide exchange factor (GEF) was modified to be responsive to protein kinase A (PKA; Yeh *et al.* 2007). Active GEFs catalyse the exchange of GDP for GTP bound to Rho GTPases, and, in turn, GTP-bound Rho activates downstream effectors (Rossman *et al.* 2005). To build a synthetic GEF, researchers replaced the cognate autoinhibitory domain of the CDC42-specific GEF Itsn1 with a PKA-dependent autoinhibitory domain, leaving the GEF catalytic domain intact (figure 3; Yeh *et al.* 2007). The addition of forskolin, a PKA activator, induces production of filopodia in mammalian cells, indicating Itsn1 signalling. Substituting different GEF catalytic domains also produced new signalling behaviours, and signalling cascades involving two synthetic GEFs were also functional (Yeh *et al.* 2007).

Protein domains can be combined to produce novel switch-like behaviour. A chimeric protein with two separate ligand-binding domains could act as a switch or an OR gate if only one ligand-binding domain could be occupied at a time. To isolate new protein switches, researchers overlapped functional ligand-binding domains and peptides in chimeric proteins, such that correct folding of one domain would disrupt folding of the other (Sallee *et al.* 2007). Out of 25 candidates, seven chimeric proteins yielded functional switches, with domains that are unstructured in the absence of ligand showing the highest likelihood of success (Sallee *et al.* 2007).

Protein ligands that bind cell surface receptors can also be used as modular regulatory elements. One such device, a chimeric-activating protein, was constructed by connecting an epidermal growth factor (EGF) ligand and an interferon- $\alpha$ -2a (IFN $\alpha$ -2a) ligand via a flexible linker (Cironi *et al.* 2008). The EGF ligand acts as a targeting element, binding the EGF receptor (EGFR). The IFN $\alpha$ -2a ligand triggers the desired action of the device, binding IFN $\alpha$ -2a–IFN $\alpha$  receptor 2 (IFNAR2) and activating the Jak–Stat pathway (Platanias 2005). In the chimeric activator, the IFN $\alpha$ -2a ligand was mutated to reduce its binding affinity for IFNAR2. Reducing IFN $\alpha$ -2a binding affinity had the desired effect: IFN $\alpha$ -2a-mediated activation of the

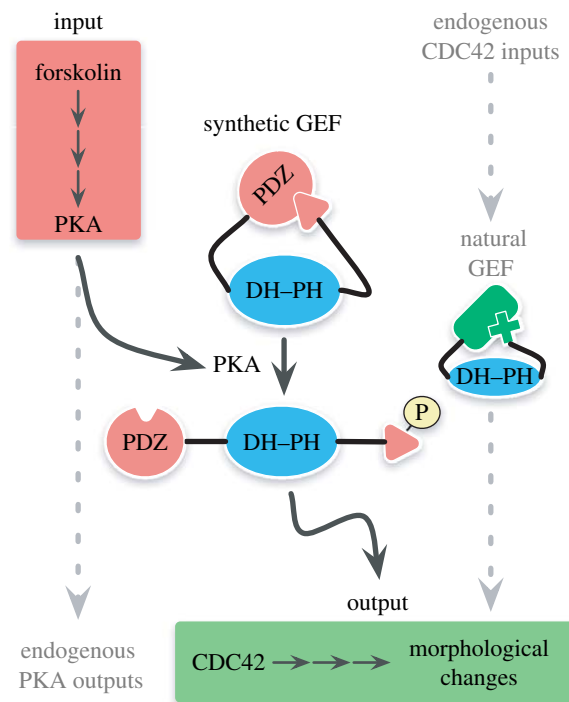


Figure 3. A synthetic GEF re-routes PKA signalling to activate the CDC42 pathway. The Dbl homology–pleckstrin homology domain (DH–PH) from a GEF involved in CDC42 signalling is combined with a PDZ domain and a PKA target peptide. The PDZ domain is normally bound to the target peptide, inactivating the DH–PH. Forskolin activates PKA that phosphorylates the target peptide. Phosphorylation of the target allows the DH–PH to activate CDC42 signalling (Yeh *et al.* 2007).

Jak–Stat pathway occurs only when both EGFR and IFNAR2 were present on the cell surface. EGF binding to EGFR brings IFN $\alpha$ -2a closer to the cell surface, increasing the likelihood of IFN $\alpha$ -2a–IFNAR2 binding and subsequent Jak–Stat signalling. As well as having therapeutic applications, chimeric activators could be incorporated into synthetic devices for intercellular signalling (Cironi *et al.* 2008).

These synthetic protein devices demonstrate that the rearrangement of natural protein modules can yield new behaviours. To achieve protein–protein interactions that are truly orthogonal to an existing protein network, however, it may be necessary to design new protein interactions. There is evidence that new types of signalling, such as tyrosine kinase signalling, evolved in response to the saturation of previous signalling networks (King *et al.* 2003; Bhattacharyya *et al.* 2006). The engineering of novel signalling systems may permit synthetic devices to operate in cells without the interference of the endogenous protein network.

Modification of protein interfaces and binding pockets (as opposed to the rearrangement of modular elements discussed previously) has proven to be an effective method of altering protein specificity. Computational modelling of protein interfaces is often similar to *ab initio* modelling of whole proteins, although the scope of the model is reduced to the interface in question (Kortemme & Baker 2004). By modifying the interacting surface of one protein and predicting

compensating mutations in a binding partner protein, natural protein–protein interfaces have been successfully redesigned (Kortemme *et al.* 2004). Similarly, the rational design of ligand-binding pockets led to engineered periplasmic binding protein receptors capable of binding trinitrotoluene, L-lactate and serotonin (Looger *et al.* 2003).

## 5. METABOLIC ENGINEERING

The rational design of organisms to produce important metabolites such as biofuels and drugs has been labelled as the defining application of synthetic biology (Brenner *et al.* 2006). Rational reconfiguration of metabolism is also an enormous challenge; metabolite levels are regulated in many ways, and small changes in gene regulation can have amplified effects on the metabolome (Raamsdonk *et al.* 2001; Kell 2006). As demonstrated in the following examples, the construction of a ‘metabolic device’ requires substantial rewiring of the host cell.

A significant effort in metabolic engineering has been the production of amorphaadiene in *E. coli* (Martin *et al.* 2003) and artemisinic acid in *S. cerevisiae* (Ro *et al.* 2006), both precursors to the anti-malarial drug artemisinin. In both cases, the amorphaadiene synthase (ADS) gene from *Artemisia annua* L was heterologously expressed, and flux through the host metabolic network was redirected towards ADS. In the case of *S. cerevisiae*, expressing the ADS is sufficient for amorphaadiene production, albeit with low yields. Adjusting the expression levels of five genes involved in the production of farnesyl pyrophosphate (FPP, converted to amorphaadiene by ADS) yielded a 500-fold increase in amorphaadiene production (figure 4). Screening a library of *A. annua* cytochrome P450 expressed sequence tags yielded an enzyme that catalysed the conversion of amorphaadiene to artemisinic acid, which was then integrated into the engineered strain along with NADPH : cytochrome P450 oxidoreductase. In the case of artemisinic acid, as well as in many other metabolic engineering efforts, the redirection of flux via adjustments in gene regulation was paramount in achieving commercially viable yields (Kasling 2008).

Re-routing metabolic flux via gene deletions can force cells to produce more of a desired product. Genome-scale simulations of cellular metabolism such as flux balance analysis (FBA), which we will explore in §6, can predict beneficial gene deletions. This approach has been validated for several metabolic engineering efforts, including the production of lycopene in *E. coli* (Alper *et al.* 2005), and in the case of our own efforts to produce formic acid in *S. cerevisiae* (Kennedy *et al.* in preparation). Results in both studies suggested that further regulatory modifications would boost yields. *In silico* strain design coupled with tight regulation of enzyme levels by synthetic devices will undoubtedly be essential to future metabolic engineering efforts.

As in protein signalling networks, modular manipulation of metabolic enzymes would facilitate the development of new pathways. However, unlike signalling proteins, metabolic enzymes are rarely structurally

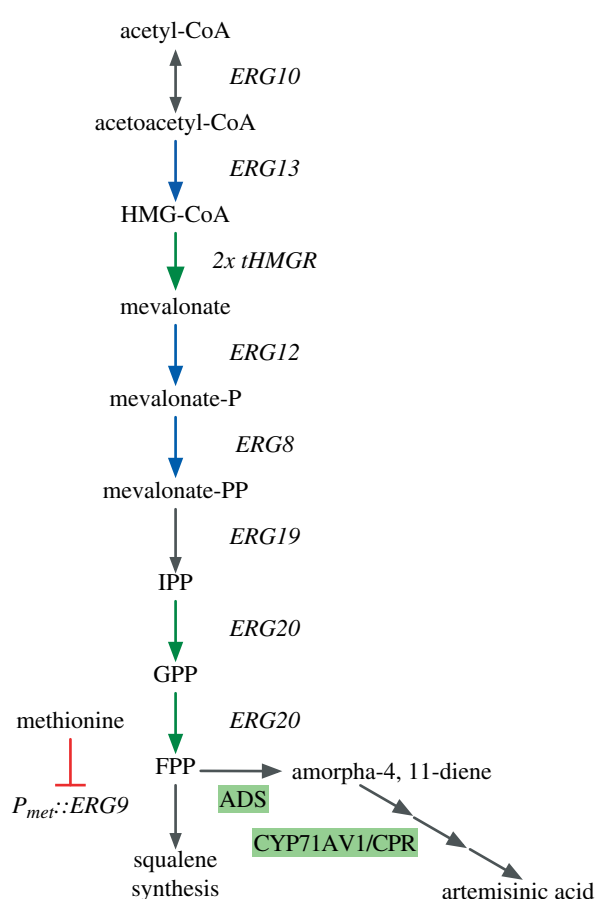


Figure 4. Engineered pathway for artemisinic acid production in *S. cerevisiae*. Blue arrows indicate enzymes indirectly upregulated by expression of *upc2-1*. Green arrows indicate enzymes that were directly upregulated. The red repression arrow indicates that *ERG9* was placed under the control of a methionine-repressed promoter, reducing flux to squalene synthesis. Green boxes indicate the exogenous enzymes ADS and CYP71AV1 and the redox partner protein CPR (Ro *et al.* 2006). CoA, coenzyme A; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; IPP, isopentenylpyrophosphate; GPP, geranyl pyrophosphate.

modular. For example, metabolic enzymes often contain allosteric regulatory sites and the active site within a single domain, confounding efforts to decouple allosteric regulation and catalysis (Bhattacharyya *et al.* 2006). A notable exception is enzymes such as polyketide synthases (PKS), which are not only modular, but their spatial arrangement can also be exploited to create new products.

Rational engineering of PKS assembly lines would open up new possibilities for the synthesis of organic molecules. Polyketides are assembled by linear complexes of PKS proteins, with each PKS performing catalysis on the growing polyketide chain (figure 5). Evolutionary rearrangement of PKS modules has generated a diverse array of natural products, including many antibiotics (Robinson 1991). PKS proteins possess N- and C-terminal ‘docking domains’ for attachment to other PKS (Thattai *et al.* 2007). Combinatorial shuffling of PKS modules resulted in the *in vivo* synthesis of novel polyketides (Menzella *et al.* 2005). Computational modelling of potential PKS products

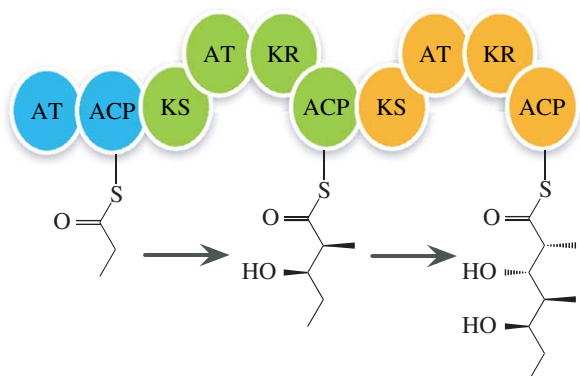


Figure 5. The DEBS1 PKS that catalyses the first steps in synthesizing 6-deoxyerythronolide B. Blue domains comprise the loading module, containing an acyl transferase (AT) and an acyl carrier protein (ACP). Two extender modules, in green and orange, each contain a ketosynthase (KS), an AT, a ketoreductase (KR) and an ACP (Menzella *et al.* 2005).

suggests that billions of possible molecules could be synthesized via engineered PKS combinations, and that it may be possible to predict PKS combinations that will produce a desired compound (González-Lergier *et al.* 2005).

Our ability to conduct ‘retro-biosynthesis’, the rational design of biological routes to target compounds, is limited both by our knowledge of enzyme properties and our ability to override regulation in biological systems (Prather & Martin 2008). The diversity of compounds synthesized by natural organisms suggests that biological chassis are an ideal platform for chemical production. It has been noted that the development of synthetic biology is paralogous to the development of synthetic chemistry; the synthesis of many important organic compounds was achieved before chemists understood covalent bonds (Yeh & Lim 2007). Similarly, efforts to re-engineer metabolism will contribute to a more complete understanding of metabolic systems and cellular regulation.

## 6. METABOLIC MODELLING

The state of the cellular metabolic network is a function of the network topology, the physical properties of enzymes and the regulation of enzyme levels and activity. Rational design of metabolism will require accurate models of the metabolic network and how the network is regulated. The complexity of metabolism has necessitated trade-offs in the formulation of metabolic models. In general, current models of metabolism fall into one of two categories: constraint-based models and kinetic models.

Most constraint-based models of metabolism are based on the framework of FBA, a technique that simulates the entire metabolic network of an organism (Varma & Palsson 1994). The only required parameter for an FBA model is a stoichiometric matrix that contains all known metabolic reactions of an organism. Constraints are placed on certain fluxes, defining nutrient availability and relative uptake rates as well as thermodynamic constraints on the reversibility of reactions. It is assumed that, at steady state, the net

flux of the system is fixed. The model is then solved for the optimization of an objective function such as maximization of biomass. Since FBA models do not consider enzymatic parameters beyond the stoichiometry of each reaction, the availability of comprehensive databases such as the Kyoto Encyclopedia of Genes and Genomes (<http://www.kegg.com>) has fostered the development of FBA models for many organisms (Varma & Palsson 1994; Duarte *et al.* 2004; Becker & Palsson 2005; Feist *et al.* 2007; Lee *et al.* 2008; Senger & Papoutsakis 2008). Owing to the genome-scale nature of constraint-based models, *in silico* screens have been applied to predictions of gene essentiality (Edwards & Palsson 2000; Thiele *et al.* 2005; Samal *et al.* 2006; Becker & Palsson 2008) and the related metabolic engineering problem of predicting gene knockouts for strain optimization (Burgard *et al.* 2003; Alper *et al.* 2005; Kennedy *et al.* in preparation).

The successful application of FBA in a variety of organisms demonstrates the use of constraint-based models in the context of metabolism. However, even in the unlikely case that enzyme kinetics are unimportant to determine metabolic flux, traditional FBA models assume that a cell's entire complement of enzymes is available at all times. Regulation can be modelled implicitly, via methods such as minimization of metabolic adjustment, which assumes that regulation will force mutant flux distributions to be as similar to the wild-type distribution as possible (Segrè *et al.* 2002). Models such as regulatory FBA attempt to explicitly model regulation by switching fluxes on and off, based on the experimental data of enzyme expression in various growth conditions (Covert *et al.* 2001; Covert & Palsson 2002; Herrgård *et al.* 2006).

Experimental efforts to modify transcription factor behaviour have underscored the importance of regulation to metabolic fluxes. One such approach, known as ‘global transcription machinery engineering’, yields improved strains by screening transcription factor mutants. For example, a more ethanol-tolerant strain of *S. cerevisiae* was isolated by mutagenizing the TATA-binding protein *SPT15* (Alper *et al.* 2006). The global influence of transcription factors makes them a powerful tool for strain construction; unfortunately, this has also made it difficult to predict the full effects of transcription factor modification. Future genome-scale models may include genome-scale models of transcriptional regulation, but comprehensive information on transcription factor interactions is not presently available.

Metabolic regulation *in vivo* is influenced by enzyme and substrate concentrations and the kinetic parameters of each enzyme. Kinetic models sacrifice the genome scale of constraint-based models in favour of detailed quantitative modelling of specific pathways. Unlike FBA, metabolic control analysis (MCA) accounts for the kinetic parameters of all enzymes in the pathway, along with the concentrations of the enzymes and the metabolites involved (Fell 1997). Control coefficients for each enzyme, based on experimentally measured parameters, define the amount of influence an enzyme has on a pathway. A key concept of MCA is multisite modulation,



i.e. there is no 'master' enzyme or rate-limiting step in a pathway; instead, each enzyme in the pathway has a non-zero control coefficient, and the control coefficients of all the enzymes in the pathway sum to 1. Thus, MCA can identify enzymes with high control coefficients, and quantitatively predict the impact of adjusting the concentration or rate of those enzymes. Unfortunately, the detailed nature of MCA models makes them impossible to apply on the genome scale until more comprehensive data on enzyme kinetics have been collected.

Even if 'omics' technologies generate a complete list of parameters for genome-scale MCA, the resulting dataset would still be difficult to model (Schuster 1999). One approach to reducing the complexity of a genome-scale model is to group metabolic fluxes into modules, provided that the groupings are functionally relevant. Although individual metabolic enzymes are rarely composed of modular elements, considering groups of enzymes as higher order modules may yield adequate models for understanding metabolic network kinetics at a genome scale.

There is evidence that higher order modularity exists in metabolic networks. Metabolic networks appear to be both scale free and hierarchical (Ravasz *et al.* 2002). Modularity is inherent in such networks, which implies that metabolic networks contain modules of small reaction networks linked to highly connected hubs (Jeong *et al.* 2000; Guimerà & Nunes Amaral 2005). Modelling approaches such as elementary flux modes analysis (Schuster *et al.* 1999, 2000; Klamt & Stelling 2002) and modular MCA (Schuster *et al.* 1993; Acerenza & Ortega 2007; Poolman *et al.* 2007) attempt to identify modular reaction sets, generating simplified network models. Identifying modular elements of metabolism may reveal generalized methods for manipulating metabolic regulation in engineered organisms.

## 7. SUBCELLULAR ENGINEERING

In computer science, abstraction allows the construction of higher orders of modularity (Abelson *et al.* 1996). Computer software usually takes inputs and returns outputs without involving the user in the intervening calculations. Within complex computer programs, there is further abstraction between modules, separating elements such as memory management and input/output control. Similarly, subcellular compartments are an evolutionary form of abstraction; eukaryotic cells possess membrane-bound organelles with specialized and well-defined tasks. Regulated transport limits interaction between the organelle and the rest of the cell to a set of defined inputs and outputs. Metabolite and protein interactions within the organelle are abstracted from the cytosolic environment. Although prokaryotes lack membrane-bound organelles, protein-bound compartments such as carboxysomes exist in certain species (Price *et al.* 2008). In the context of synthetic devices, abstraction afforded by subcellular compartments could limit the interference of endogenous regulation with the engineered regulation of the device.

Lysosomes, peroxisomes and many other organelles contain enzymes or metabolites that are harmful to the host cell (Page *et al.* 1998; Yeldandi *et al.* 2000). Metabolic reactions that would be thermodynamically unfavourable in the cytosol are often found in organelles (Feldman & Sigman 1983). Organelles could conceivably harbour engineered pathways that would be incompatible with the cytoplasm.

The specialized machinery of organelles could augment synthetic devices. Chloroplasts and mitochondria, being of endosymbiotic origin (Embley & Martin 2006), are the most sophisticated eukaryotic organelles and have useful properties for engineered systems. These organelles possess subcompartments of their own, the thylakoids in the case of chloroplast and the matrix in the case of the mitochondrion (Frey & Mannella 2000; Mustárdy *et al.* 2008). In addition, both organelles can generate electrochemical gradients between their subcompartments (Dimroth *et al.* 2000).

Although no completed examples of synthetic organelle devices exist, there is reason to believe that engineering organelles is feasible. In the case of many organelles, such as chloroplasts (Soll & Schleiff 2004), mitochondria (Truscott *et al.* 2003) and peroxisomes (Léon *et al.* 2006), the targeting of arbitrary proteins to these compartments has been described. Chloroplasts and mitochondria also have self-contained genomes, although gene integration into their genomes is more complex than nuclear integration (Bonney & Fox 2007; Verma & Daniell 2007).

Genome-wide metabolic models often have difficulties modelling organelle metabolism (Satish Kumar *et al.* 2007). However, since organelles compartmentalize their metabolic reactions, they can be abstracted from the rest of cellular metabolism. Inputs and outputs from a detailed kinetic model of organelle metabolism could be interfaced with a genome-scale constraint-based metabolic model. In this manner, detailed models of the organelle in question could be combined with global models for the rest of the cell.

## 8. MULTISPECIES DEVICES

In nature, environmental niches are often colonized by microbial consortia rather than a single dominating species. This may confer a group advantage to cooperating species. In the bovine rumen, multispecies microbial biofilms manage to degrade cellulosic biomass (McAllister *et al.* 1994). Engineered co-cultures may be able to achieve similarly difficult tasks (Brenner *et al.* 2008).

Co-cultures have been applied to the breakdown of lignocellulose, an important carbon source for biofuels (Eiteman *et al.* 2008). In current industrial processes, hydrolysis of lignocellulose yields a mix of five-carbon sugars, such as xylose, and six carbon sugars, such as glucose. In the presence of both xylose and glucose, *E. coli* will preferentially feed on glucose first. A co-culture of two *E. coli* strains, one that is deficient in xylose usage and the other that is deficient in glucose usage, breaks down the sugar mixture more effectively than a monoculture (Eiteman *et al.* 2008).

Synthetic biology offers a range of devices that may allow the coordination of gene regulation between two species, such as quorum-sensing cell–cell signalling devices (Weber *et al.* 2007; Balagaddé *et al.* 2008; Brenner *et al.* 2008). The major task of synthetic devices in co-culture systems may be to exert a selective pressure to maintain a co-culture. Advances in metabolic modelling will improve our ability to design dependencies between strains, such as a requirement for cross-feeding.

## 9. CONCLUSIONS

The defining question of synthetic biology research moving forward will not be whether biology can be engineered, but how to develop engineering principles for biological systems. Understanding natural regulatory systems, developing improved regulatory systems for synthetic devices and properly interfacing synthetic devices with host cells will play a large role in this process.

The synthetic devices presented here have demonstrated that functioning devices can be constructed, even though our understanding of biological systems is incomplete. In the most promising cases, engineering progress has also provided new biological insights. There is also much to be learned from building synthetic devices that do not work as planned. Designing synthetic devices in an iterative fashion, with experimental results allowing improved models and vice versa, will allow increasingly complex device designs. In addition, modelling shortfalls and unexpected experimental outcomes may shed light on new mechanisms of endogenous biological regulation.

The success of synthetic biology endeavours depends heavily on understanding how biological systems are regulated. In natural systems, there are regulated interactions between DNA, RNA, proteins and metabolites. Identifying modular regulatory elements such as promoters and riboswitches has been essential to the progress of synthetic biology. There is considerable evidence that genomic rearrangements and horizontal gene transfer have driven the evolution of new biological capabilities. Similarly, the identification of biological modules that confer new functionality when assembled in different contexts will drive the progress of synthetic biology.

We thank Christina Agapakis and Caleb Kennedy for reviewing the paper. This work was supported by Harvard University Center for the Environment (HUCE) and National Institutes of Health (NIH) funding to P.A.S., and a HUCE Graduate Fellowship and an NIH Cell and Developmental Biology training grant to P.M.B.

## REFERENCES

- Abelson, H., Sussman, G. J. & Sussman, J. 1996 *Structure and interpretation of computer programs*. Cambridge, MA: The MIT Press.
- Acerenza, L. & Ortega, F. 2007 Modular metabolic control analysis of large responses. *FEBS J.* **274**, 188–201. (doi:10.1111/j.1742-4658.2006.05575.x)
- Ajo-Franklin, C. M., Drubin, D. A., Eskin, J. A., Gee, E. P. S., Landgraf, D., Phillips, I. & Silver, P. A. 2007 Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276. (doi:10.1101/gad.1586107)
- Alper, H., Jin, Y.-S., Moxley, J. F. & Stephanopoulos, G. 2005 Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* **7**, 155–164. (doi:10.1016/j.ymben.2004.12.003)
- Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. 2006 Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568. (doi:10.1126/science.1131969)
- Anderson, J. C., Voigt, C. & Arkin, A. P. 2007 Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3**, 133. (doi:10.1038/msb4100173)
- Arkin, A. P. & Fletcher, D. A. 2006 Fast, cheap and somewhat in control. *Genome Biol.* **7**, 114. (doi:10.1186/gb-2006-7-8-114)
- Balagaddé, F. K., Song, H., Ozaki, J., Collins, C. H., Barnet, M., Arnold, F. H., Quake, S. R. & You, L. 2008 A synthetic *Escherichia coli* predator–prey ecosystem. *Mol. Syst. Biol.* **4**, 187. (doi:10.1038/msb.2008.24)
- Bartel, D. P. & Unrau, P. J. 1999 Constructing an RNA world. *Trends Cell Biol.* **9**, M9–M13. (doi:10.1016/S0962-8924(99)01669-4)
- Bayer, T. S. & Smolke, C. D. 2005 Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.* **23**, 337–343. (doi:10.1038/nbt1069)
- Becker, S. A. & Palsson, B. Ø. 2005 Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.* **5**, 8. (doi:10.1186/1471-2180-5-8)
- Becker, S. A. & Palsson, B. Ø. 2008 Three factors underlying incorrect *in silico* predictions of essential metabolic genes. *BMC Syst. Biol.* **2**, 14. (doi:10.1186/1752-0509-2-14)
- Bhattacharyya, R. P., Reményi, A., Yeh, B. J. & Lim, W. A. 2006 Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75**, 655–680. (doi:10.1146/annurev.biochem.75.103004.142710)
- Bonnefoy, N. & Fox, T. D. 2007 Directed alteration of *Saccharomyces cerevisiae* mitochondrial DNA by biolistic transformation and homologous recombination. *Methods Mol. Biol.* **372**, 153–166. (doi:10.1007/978-1-59745-365-3\_11)
- Brenner, K., You, L. & Arnold, F. H. 2008 Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* **26**, 483–489. (doi:10.1016/j.tibtech.2008.05.004)
- Brenner, M. P. *et al.* 2006 Engineering microorganisms for energy production. Report JSR-05-300, US Department of Energy, Washington, DC.
- Burgard, A., Pharkya, P. & Maranas, C. 2003 Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657. (doi:10.1002/bit.10803)
- Casadaban, M. J. 1975 Fusion of the *Escherichia coli* lac genes to the ara promoter: a general technique using bacteriophage Mu-1 insertions. *Proc. Natl Acad. Sci. USA* **72**, 809–813. (doi:10.1073/pnas.72.3.809)
- Cironi, P., Swinburne, I. A. & Silver, P. A. 2008 Enhancement of cell type specificity by quantitative modulation of a chimeric ligand. *J. Biol. Chem.* **283**, 8469–8476. (doi:10.1074/jbc.M708502200)
- Covert, M. W. & Palsson, B. Ø. 2002 Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277**, 28 058–28 064. (doi:10.1074/jbc.M201691200)

- Covert, M. W., Schilling, C. H. & Palsson, B. Ø. 2001 Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* **213**, 73–88. (doi:10.1006/jtbi.2001.2405)
- Dimroth, P., Kaim, G. & Matthéy, U. 2000 Crucial role of the membrane potential for ATP synthesis by F(1)F(o) ATP synthases. *J. Exp. Biol.* **203**, 51–59.
- Drubin, D. A., Way, J. C. & Silver, P. A. 2007 Designing biological systems. *Genes Dev.* **21**, 242–254. (doi:10.1101/gad.1507207)
- Duarte, N. C., Herrgård, M. J. & Palsson, B. Ø. 2004 Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14**, 1298–1309. (doi:10.1101/gr.2250904)
- Edwards, J. S. & Palsson, B. Ø. 2000 Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinform.* **1**, 1. (doi:10.1186/1471-2105-1-1)
- Eiteman, M. A., Lee, S. A. & Altman, E. 2008 A co-fermentation strategy to consume sugar mixtures effectively. *J. Biol. Eng.* **2**, 3. (doi:10.1186/1754-1611-2-3)
- Elowitz, M. B. & Leibler, S. 2000 A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338. (doi:10.1038/35002125)
- Embley, T. M. & Martin, W. 2006 Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630. (doi:10.1038/nature04546)
- Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V. & Palsson, B. Ø. 2007 A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3**, 121. (doi:10.1038/msb4100155)
- Feldman, R. I. & Sigman, D. S. 1983 The synthesis of ATP by the membrane-bound ATP synthase complex from medium 32Pi under completely uncoupled conditions. *J. Biol. Chem.* **258**, 12 178–12 183.
- Fell, D. A. 1997 *Understanding control of metabolism*. London, UK: Portland Press.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. 1998 Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811. (doi:10.1038/35888)
- Frey, T. G. & Mannella, C. A. 2000 The internal structure of mitochondria. *Trends Biochem. Sci.* **25**, 319–324. (doi:10.1016/S0968-0004(00)01609-1)
- Gardner, T. S., Cantor, C. R. & Collins, J. J. 2000 Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342. (doi:10.1038/35002131)
- Gibson, D. G. *et al.* 2008 Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220. (doi:10.1126/science.1151721)
- González-Lergier, J., Broadbelt, L. J. & Hatzimanikatis, V. 2005 Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *J. Am. Chem. Soc.* **127**, 9930–9938. (doi:10.1021/ja051586y)
- Guimerà, R. & Nunes Amaral, L. A. 2005 Functional cartography of complex metabolic networks. *Nature* **433**, 895–900. (doi:10.1038/nature03288)
- Herrgård, M. J., Lee, B. S., Portnoy, V. & Palsson, B. Ø. 2006 Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* **16**, 627–635. (doi:10.1101/gr.4083206)
- Irving, A. 1961 *Thinking machines, a layman's introduction to logic, Boolean algebra, and computers*. New York, NY: John Day Co.
- Isaacs, F. J., Dwyer, D. J. & Collins, J. J. 2006 RNA synthetic biology. *Nat. Biotechnol.* **24**, 545–554. (doi:10.1038/nbt1208)
- Jacob, F. & Monod, J. 1961 Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. 2000 The large-scale organization of metabolic networks. *Nature* **407**, 651–654. (doi:10.1038/35036627)
- Keasling, J. D. 2008 Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* **3**, 64–76. (doi:10.1021/cb7002434)
- Kell, D. B. 2006 Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov. Today* **11**, 1085–1092. (doi:10.1016/j.drudis.2006.10.004)
- Kennedy, C. J., Boyle, P. M., Waks, Z. & Silver, P. A. In preparation. Systems-level engineering of non-fermentative metabolism in yeast.
- Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S. D. 2003 Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.* **10**, 708–712. (doi:10.1038/nsb959)
- King, N., Hittinger, C. T. & Carroll, S. B. 2003 Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301**, 361–363. (doi:10.1126/science.1083853)
- Klamt, S. & Stelling, J. 2002 Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **29**, 233–236. (doi:10.1023/A:1020390132244)
- Kortemme, T. & Baker, D. 2004 Computational design of protein–protein interactions. *Curr. Opin. Chem. Biol.* **8**, 91–97. (doi:10.1016/j.cbpa.2003.12.008)
- Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. 2004 Computational redesign of protein–protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371–379. (doi:10.1038/nsmb749)
- Lander, E. S. *et al.* 2001 Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921. (doi:10.1038/35057062)
- Lee, J., Yun, H., Feist, A. M., Palsson, B. Ø. & Lee, S. Y. 2008 Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Appl. Microbiol. Biotechnol.* **80**, 849–862. (doi:10.1007/s00253-008-1654-4)
- Léon, S., Goodman, J. M. & Subramani, S. 2006 Uniqueness of the mechanism of protein import into the peroxisome matrix: transport of folded, co-factor-bound and oligomeric proteins by shuttling receptors. *Biochim. Biophys. Acta* **1763**, 1552–1564. (doi:10.1016/j.bbamer.2006.08.037)
- Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. 2003 Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185–190. (doi:10.1038/nature01556)
- Mandal, M. & Breaker, R. R. 2004 Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463. (doi:10.1038/nrm1403)
- Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. 2003 Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802. (doi:10.1038/nbt833)
- McAllister, T. A., Bae, H. D., Jones, G. A. & Cheng, K. J. 1994 Microbial attachment and feed digestion in the rumen. *J. Anim. Sci.* **72**, 3004–3018.
- Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A. & Santi, D. V. 2005 Combinatorial polyketide biosynthesis by



- de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176. (doi:10.1038/nbt1128)
- Mustárdy, L., Buttle, K., Steinbach, G. & Garab, G. 2008 The three-dimensional network of the thylakoid membranes in plants: quasi-helical model of the granum-stroma assembly. *Plant Cell* **20**, 2552–2557. (doi:10.1105/tpc.108.059147)
- Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L. & Breaker, R. R. 2002 Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043. (doi:10.1016/S1074-5521(02)00224-7)
- Page, L. J., Darmon, A. J., Uellner, R. & Griffiths, G. M. 1998 L is for lytic granules: lysosomes that kill. *Biochim. Biophys. Acta* **1401**, 146–156. (doi:10.1016/S0167-4889(97)00138-9)
- Patel, D. J., Suri, A. K., Jiang, F., Jiang, L., Fan, P., Kumar, R. A. & Nonin, S. 1997 Structure, recognition and adaptive binding in RNA aptamer complexes. *J. Mol. Biol.* **272**, 645–664. (doi:10.1006/jmbi.1997.1281)
- Pawson, T. 2007 Dynamic control of signaling by modular adaptor proteins. *Curr. Opin. Cell Biol.* **19**, 112–116. (doi:10.1016/j.ceb.2007.02.013)
- Platanias, L. C. 2005 Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **5**, 375–386. (doi:10.1038/nri1604)
- Poolman, M. G., Sebu, C., Pidcock, M. K. & Fell, D. A. 2007 Modular decomposition of metabolic systems via null-space analysis. *J. Theor. Biol.* **249**, 691–705. (doi:10.1016/j.jtbi.2007.08.005)
- Prather, K. L. & Martin, C. H. 2008 *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.* **19**, 468–474. (doi:10.1016/j.copbio.2008.07.009)
- Price, G. D., Badger, M. R., Woodger, F. J. & Long, B. M. 2008 Advances in understanding the cyanobacterial CO<sub>2</sub>-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot.* **59**, 1441–1461. (doi:10.1093/jxb/erm112)
- Raamsdonk, L. M. *et al.* 2001 A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **19**, 45–50. (doi:10.1038/83496)
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A. L. 2002 Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555. (doi:10.1126/science.1073374)
- Reznikoff, W. S., Miller, J. H., Scaife, J. G. & Beckwith, J. R. 1969 A mechanism for repressor action. *J. Mol. Biol.* **43**, 201–213. (doi:10.1016/0022-2836(69)90089-8)
- Rinaudo, K., Bleris, L., Maddamsetti, R., Subramanian, S., Weiss, R. & Benenson, Y. 2007 A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.* **25**, 795–801. (doi:10.1038/nbt1307)
- Ro, D. K. *et al.* 2006 Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943. (doi:10.1038/nature04640)
- Robinson, J. A. 1991 Polyketide synthase complexes: their structure and function in antibiotic biosynthesis. *Phil. Trans. R. Soc. B* **332**, 107–114. (doi:10.1098/rstb.1991.0038)
- Rossmann, K. L., Der, C. J. & Sondek, J. 2005 GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat. Rev. Mol. Cell Biol.* **6**, 167–180. (doi:10.1038/nrm1587)
- Saito, H. 2001 Histidine phosphorylation and two-component signaling in eukaryotic cells. *Chem. Rev.* **101**, 2497–2509. (doi:10.1021/cr000243+)
- Saito, H. & Inoue, T. 2008 Synthetic biology with RNA motifs. *Int. J. Biochem. Cell Biol.* **41**, 398–404. (doi:10.1016/j.biocel.2008.08.017)
- Sallee, N. A., Yeh, B. J. & Lim, W. A. 2007 Engineering modular protein interaction switches by sequence overlap. *J. Am. Chem. Soc.* **129**, 4606–4611. (doi:10.1021/ja0672728)
- Samal, A., Singh, S., Giri, V., Krishna, S., Raghuram, N. & Jain, S. 2006 Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC Bioinform.* **7**, 118. (doi:10.1186/1471-2105-7-118)
- Satish Kumar, V., Dasika, M. S. & Maranas, C. D. 2007 Optimization based automated curation of metabolic reconstructions. *BMC Bioinform.* **8**, 212. (doi:10.1186/1471-2105-8-212)
- Schuster, S. 1999 Use and limitations of modular metabolic control analysis in medicine and biotechnology. *Metab. Eng.* **1**, 232–242. (doi:10.1006/mben.1999.0119)
- Schuster, S., Kahn, D. & Westerhoff, H. V. 1993 Modular analysis of the control of complex metabolic pathways. *Biophys. Chem.* **48**, 1–17. (doi:10.1016/0301-4622(93)80037-J)
- Schuster, S., Dandekar, T. & Fell, D. A. 1999 Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* **17**, 53–60. (doi:10.1016/S0167-7799(98)01290-6)
- Schuster, S., Fell, D. A. & Dandekar, T. 2000 A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* **18**, 326–332. (doi:10.1038/73786)
- Segrè, D., Vitkup, D. & Church, G. M. 2002 Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99**, 15 112–15 117. (doi:10.1073/pnas.232349399)
- Senger, R. S. & Papoutsakis, E. T. 2008 Genome-scale model for *Clostridium acetobutylicum*: part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng.* **101**, 1036–1052. (doi:10.1002/bit.22010)
- Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M. & Laub, M. T. 2008 Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043–1054. (doi:10.1016/j.cell.2008.04.040)
- Soll, J. & Schleiff, E. 2004 Protein import into chloroplasts. *Nat. Rev. Mol. Cell Biol.* **5**, 198–208. (doi:10.1038/nrm1333)
- Stoddard, C. D. & Batey, R. T. 2006 Mix-and-match riboswitches. *ACS Chem. Biol.* **1**, 751–754. (doi:10.1021/cb600458w)
- Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S. & Hasty, J. 2008 A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519. (doi:10.1038/nature07389)
- Swinburne, I. A. & Silver, P. A. 2008 Intron delays and transcriptional timing during development. *Dev. Cell* **14**, 324–330. (doi:10.1016/j.devcel.2008.02.002)
- Swinburne, I. A., Miguez, D. G., Landgraf, D. & Silver, P. A. 2008 Intron length increases oscillatory periods of gene expression in animal cells. *Genes Dev.* **22**, 2342–2346. (doi:10.1101/gad.1696108)
- Thattai, M., Burak, Y. & Shraiman, B. I. 2007 The origins of specificity in polyketide synthase protein interactions. *PLoS Comput. Biol.* **3**, 1827–1835. (doi:10.1371/journal.pcbi.0030186)
- Thiele, I., Vo, T. D., Price, N. D. & Palsson, B. Ø. 2005 Expanded metabolic reconstruction of *Helicobacter*



- pylori* (iT341 GSM/GPR): an *in silico* genome-scale characterization of single- and double-deletion mutants. *J. Bacteriol.* **187**, 5818–5830. (doi:10.1128/JB.187.16.5818-5830.2005)
- Truscott, K. N., Brandner, K. & Pfanner, N. 2003 Mechanisms of protein import into mitochondria. *Curr. Biol.* **13**, R326–R337. (doi:10.1016/S0960-9822(03)00239-2)
- Tucker, B. J. & Breaker, R. R. 2005 Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* **15**, 342–348. (doi:10.1016/j.sbi.2005.05.003)
- Varma, A. & Palsson, B. Ø. 1994 Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* **60**, 3724–3731.
- Venter, J. C. *et al.* 2001 The sequence of the human genome. *Science* **291**, 1304–1351. (doi:10.1126/science.1058040)
- Verma, D. & Daniell, H. 2007 Chloroplast vector systems for biotechnology applications. *Plant Physiol.* **145**, 1129–1143. (doi:10.1104/pp.107.106690)
- Weber, W., Daoud-El Baba, M. & Fussenegger, M. 2007 Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proc. Natl Acad. Sci. USA* **104**, 10 435–10 440. (doi:10.1073/pnas.0701382104)
- West, A. H. & Stock, A. M. 2001 Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci.* **26**, 369–376. (doi:10.1016/S0968-0004(01)01852-7)
- Win, M. N. & Smolke, C. D. 2008 Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460. (doi:10.1126/science.1160311)
- Yeh, B. J. & Lim, W. A. 2007 Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.* **3**, 521–525. (doi:10.1038/nchembio0907-521)
- Yeh, B. J., Rutigliano, R. J., Deb, A., Bar-Sagi, D. & Lim, W. A. 2007 Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. *Nature* **447**, 596–600. (doi:10.1038/nature05851)
- Yeldandi, A. V., Rao, M. S. & Reddy, J. K. 2000 Hydrogen peroxide generation in peroxisome proliferator-induced oncogenesis. *Mutat. Res.* **448**, 159–177. (doi:10.1016/S0027-5107(99)00234-1)