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.
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'.)

<table>
<thead>
<tr>
<th>gate</th>
<th>A</th>
<th>B</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>AND</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>OR</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>NAND</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>NOR</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>

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 1c: 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 activation 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 & Unrav 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.
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; Issacs 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 ligand-binding aptamer domain with a different target domain changes the specificity of an antiswitch to new ligands.

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 fluorescent protein reporter. 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 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).
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 natural systems, such as the clock in the cyanobacterium Synechococcus elongatus (Lips and Finlay 2004). As in the endogenous genes of higher eukaryotes, introns may provide an additional layer of control over the expression of target mRNAs if expressed (Rinaudo et al. 2007).

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-α-2a (IFNα-2a) ligand via a flexible linker (Cironi et al. 2008). The EGF ligand acts as a targeting element, binding the EGF receptor (EGFR). The IFNα-2a ligand triggers the desired action of the device, binding IFNα-2a–IFNAR receptor 2 (IFNAR2) and activating the Jak–Stat pathway (Platanias 2005). In the chimeric activator, the IFNα-2a ligand was mutated to reduce its binding affinity for IFNAR2. Reducing IFNα-2a binding affinity had the desired effect: IFNα-2a-mediated activation of the Jak–Stat pathway occurs only when both EGFR and IFNAR2 were present on the cell surface. EGF binding to EGFR brings IFNα-2a closer to the cell surface, increasing the likelihood of IFNα-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; Bhattaeharyya 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 ‘metabolic devices’ requires substantial rewiring of the host cell.

A significant effort in metabolic engineering has been the production of amorpha-4, 11-diene 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 amorpha-4, 11-diene 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 amorpha-4, 11-diene production, albeit with low yields. Adjusting the expression levels of five genes involved in the production of farnesyl pyrophosphate (FPP, converted to amorpha-4, 11-diene by ADS) yielded a five-fold increase in amorpha-4, 11-diene production (figure 4). Screening a library of A. annua cytochrome P450 expressed sequence tags yielded an enzyme that catalysed the conversion of amorpha-4, 11-diene 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 (Keasling 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 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 whose 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

<table>
<thead>
<tr>
<th>acetyl-CoA</th>
<th>ERG10</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetoacetyl-CoA</td>
<td>ERG13</td>
</tr>
<tr>
<td>HMG-CoA</td>
<td>2x tHMGR</td>
</tr>
<tr>
<td>mevalonate</td>
<td>ERG12</td>
</tr>
<tr>
<td>mevalonate-P</td>
<td>ERG8</td>
</tr>
<tr>
<td>mevalonate-PP</td>
<td>ERG19</td>
</tr>
<tr>
<td>IPP</td>
<td>ERG20</td>
</tr>
<tr>
<td>GPP</td>
<td>ERG20</td>
</tr>
<tr>
<td>methionine</td>
<td>FPP</td>
</tr>
<tr>
<td>P_{mec}:ERG9</td>
<td>amorpha-4, 11-diene</td>
</tr>
<tr>
<td>squalene synthesis</td>
<td>CYP71AV1/CPR</td>
</tr>
<tr>
<td></td>
<td>artemisinic acid</td>
</tr>
</tbody>
</table>

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-methyl-glutaryl-CoA; IPP, isopentenylpyrophosphate; GPP, geranyl pyrophosphate.
suggestion 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; Herrgard 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,

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).
ly 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 mode analysis (Schuster et al. 1999, 2000; Klamt & Stelling 2002) and modular MCA (Schuster et al. 1993; Acreenza & Ortég 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 & Signam 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 & Schleif 2004), mitochondria (Truscott et al. 2003) and peroxisomes (León 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 (Bonnefoy & 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 cellulose 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 (Eitman 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 (Eitman 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; Balagadde 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


