The second wave of synthetic biology: from modules to systems

Priscilla E. M. Purnick and Ron Weiss

Abstract | Synthetic biology is a research field that combines the investigative nature of biology with the constructive nature of engineering. Efforts in synthetic biology have largely focused on the creation and perfection of genetic devices and small modules that are constructed from these devices. But to view cells as true 'programmable' entities, it is now essential to develop effective strategies for assembling devices and modules into intricate, customizable larger scale systems. The ability to create such systems will result in innovative approaches to a wide range of applications, such as bioremediation, sustainable energy production and biomedical therapies.

Ribosome binding sites
A messenger RNA sequence that is recognized by the ribosome for protein translation initiation.

Directed evolution
An adaptation of the natural process of evolution — consisting of mutation and selection — to laboratory settings, in which the goal is to customize the behaviours of individual proteins as well as whole pathways for specific functions.

Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA.
e-mails:
rweiss@princeton.edu;
ppurnick@princeton.edu
doi:10.1038/nrm2698

Synthetic biology has the potential to transform how we interact with our environment and how we approach human health. Conventional genetic engineering approaches for solving complex problems typically focus on tweaking one or a few genes. Synthetic biology, by contrast, approaches these problems from a novel, engineering-driven perspective that focuses on wholesale changes to existing cellular architectures and the construction of elaborate systems from the ground up. Synthetic biology has the potential to fabricate practical organisms that could clean hazardous waste in inaccessible places[1], to use plants to sense chemicals and respond accordingly[2,3], to produce clean fuel in an efficient and sustainable fashion[4], or to recognize and destroy tumours[5]. Whether addressing an existing problem or creating new capabilities, effective solutions can be inspired by, but need not mimic, natural biological processes. Our new designs can potentially be more robust or efficient than systems that have been fashioned by evolution.

As yet, however, these goals are difficult to achieve. We begin this Review by examining the 'first wave' in synthetic biology, a phase that has focused on creating and perfecting genetic devices and small modules. We do not provide a comprehensive discussion of synthetic biology projects, but rather a description of several informative examples. The tremendous increase in the availability and characterization of devices and modules provides an important foundation for the field. These efforts have improved our quantitative understanding of natural biological processes and have helped us to establish design principles that work for small modules. We then describe the 'second wave', a phase that will help us to combine parts and modules to create more sophisticated systems. Finally, we discuss the remaining challenges and open questions for synthetic biologists.

The first wave of synthetic biology

In the first wave of synthetic biology, basic elements — for example promoters, ribosome binding sites and transcriptional repressors — were combined to form small modules with specified behaviours. Currently, modules include switches[6–9], cascades[10], pulse generators[11], time-delayed circuits[12–14], oscillators[8,15–18], spatial patterning[19] and logic formulas[20,21]. These and other modules can be used to regulate gene expression, protein function, metabolism and cell–cell communication. Synthetic biologists expanded on existing genetic engineering techniques and developed new circuit design principles that seem to work well for constructing small biological modules. For example, one such design principle — iterative rational design — involves creating and analysing a computational model of a system, constructing a corresponding genetic circuit, experimentally evaluating circuit performance and refining the design until a performance objective is achieved. Another commonly used strategy is to construct circuit variants with parts in different combinations and configurations, and to then select the variants that exhibit suitable performance. Directed evolution provides a third method for part and circuit optimization[22–24].

The creation of functional and robust modules was not a trivial process. An important initial challenge was to manipulate these basic elements, which often came from disparate sources, so that they could work synergistically towards the desired goal. Efforts are underway to

characterize and standardize parts and small modules by measuring, manipulating and matching their input–output thresholds[25,26]. We now have a growing library of parts (TABLE 1) and modules that have been verified and characterized in different cellular contexts[26,27]. The end goal is to create a catalogue of interchangeable parts that can be easily mixed and matched for circuit construction, and that is accessible to biological engineers at all levels (BOX 1).

Genetic parts and modules allow control over cellular behaviour through various biochemical processes, including transcription, translation and post-translational processes (FIG. 1).

Transcriptional control. In an engineered circuit, one can create new promoters, or pre-existing promoters and their regulatory sites can be modified to provide a transcriptional means to control gene expression. Only a few engineered promoters exist and are routinely used. However, one can add various operator sites to existing promoters, creating additional regulatory interactions with endogenous or engineered elements. Two recent examples of synthetic promoters in yeast include modifications of the *GAL1* promoter to yield different Tet protein transcriptional repression efficiencies[28], and combinatorial promoters that require three chemical inputs[29].

Ham *et al.*[30] developed an inducible, unidirectional, invertible promoter using the FimE invertase[30]. In its initial configuration, a constitutive *trc* promoter is directed away from a gene of interest, resulting in essentially no gene expression. When arabinose is added to the system, the *trc* promoter is inverted, permanently flipping the expression switch from OFF to ON. Subsequent removal or addition of arabinose does not alter the state of the switch. This regulatory mechanism is particularly valuable in situations in which leaky gene expression is undesirable (such as with the expression of a toxic protein) or when short transient induction is preferable to long-term induction (to reduce costs, for example).

Although transcriptional control is often considered to be slow, a recent implementation of a genetic oscillator showed that transcription can be used to attain circuits that are fast and tunable[16] (FIG. 1a).

Table 1 | **Genetic elements used as components of synthetic regulatory networks**

Genetic part	Examples	Purpose
Transcriptional control		
Constitutive promoters	*laclq*[47], *CMV*[50], *EF1α*[5], *UBC*, *SV40* (REF. 7), *T7* (REF. 112), *sp6* (REF. 112), *RSV*[36] and *U6* (REF. 36)	'Always on' transcription
Regulatory regions	*tetO*[6,28,113], *lacO*[6,16,113], *cuO*, *ara*[16], *EST*[58,13], *glnA*[48], *Or*[11], *UASG*[67], *NRI*[8], *gal4* (REF. 46) and *rhl* box[53]	Repressor and activator sites
Inducible and tissue-specific promoters	*ara*[114], stress[115], *trc*[30], nitric oxide[50], *FOS*[50], ethanol[58], *lac*[47,24], *gal*[28,49], *rhl*[53], *lux*[24,47], *FUS1* (REF. 14), *FIG1* (REF. 14), *ste5* (REF. 14), *ade2* (REF. 46), *lys21* (REF. 52), *fdhH*[5], *TEF*[49], *SSRE*[49], *sal*[59], *glnK*[8], *cyc1* (REFS 14,49) and *CAG*[21]	Control of the promoter by induction or by cell state
Cell fate regulators	GATA factors, MYOD and NGN1	Control cell differentiation
Translational control		
RNAi	Logic functions[21] and RNAi repressor[36]	Genetic switch, logic evaluation and gene silencing
Riboregulators	Ligand-controlled ribozymes[31,116–120]	Switches for detection and actuation
Ribosome-binding site	Kozak consensus sequence mutants[6,121–123]	Control the level of translation
Post-translational control		
Phosphorylation cascades	Yeast phosphorylation pathway[49] and MAPK signalling scaffolds[14]	Modulate genetic circuit behaviour
Protein receptor design	TNT receptor[3], ACTR[46] and EST receptor[13]	Control detection thresholds and combinatorial protein function
Protein degradation	Ssra tags[113] and peptides rich in Pro, Glu, Ser and Thr[124]	Protein degradation at varying rates
Localization signals	NLS, NES and mito	Import or export from nucleus and mitochondria
Miscellaneous		
Cell–cell communication	AHL derivatives[19,53], ADH/acetaldehyde[58], NRL/acetate[48] and CRE1 /IPT4 (REF. 49) (these are written in the form receptor/activator)	Cell–cell communication with small diffusible molecules
Colorimetric expression	EGFP, EYFP, ECFP, LacZ, DsRed and ZsYellow	Detection of expression
Antibiotic resistance	amp, chlor, kan, bla, puro, bleo and neo	Selection of cell lines

ACTR, ACT receptor; ADH, alcohol dehydrogenase; AHL, acyl homoserine lactone; amp, ampicillin; Ara, arabinose; bla, blasticidin; bleo, bleomycin; chlor, chloramphenicol; CMV, cytomegalovirus; ECFP, enhanced cyan fluorescent protein; *EF1α*, elongation factor 1α; EGPF, enhanced green fluorescent proteins; EYFP, enhanced yellow fluorescent protein; kan, kanamycin; *lac*, lactase; MAPK, mitogen-activated protein kinase; mito, mitochondrial localization signal; MYOD, myoblast determination protein; neo, neomycin; NES, nuclear export signal; NGN1, neurogenin 1; NLS, nuclear localization signal; NRL, neural retina-specific Leu zipper protein; puro, puromycin; RNAi, RNA interference; *tetO*, tetracycline resistance protein O; TNT, trinitrotoluene; *UBC*, human ubiquitin C.

Burnt pancake problem
The challenge of this problem is to develop an efficient mathematical sorting algorithm that reverses the sequence of elements with as few reversals as possible. This problem adds complexity by assigning a sidedness to each element: each element has a 'burnt' side, and the final objective is to have all of these burnt sides facing down.

Boolean logic
The mathematical foundation for digital systems describing rules for input–output functions. Basic operations include AND, OR, NOT and NOR (NOT OR), and can be combined to form arbitrarily complex expressions.

RNA aptamer
An oligonucleotide that specifically binds a small molecule. DNA aptamers also exist.

Translational control. Classical translational control consists of mutating ribosome-binding sites to increase or decrease expression levels of engineered proteins. Recent methods of classical translational control use artificial ribozymes and riboswitches that sense and respond to small molecules and small interfering (si)RNA to silence, degrade or titrate pre-existing mRNA[20,31–36] (FIG. 1b). For example, Rackham and Chin[37] constructed orthogonal ribosome–mRNA pairs that can silence endogenous mRNA translation and also form information-processing networks that obey Boolean logic[37]. Ellington's group created RNA aptamer biosensors to detect the presence of environmental metals, such that cells would produce a strong fluorescent signal in the presence of zinc[38]. In another study, siRNA was used to create a multi-input logic evaluator in cells, which titrates mRNA levels only if particular logic statements are true (for example, enhanced yellow fluorescent protein (EYFP) is expressed only when '$siRNA_x$ AND $siRNA_y$' are inactive). In this study, the circuit could correctly evaluate up to five cellular inputs simultaneously using AND/OR/NOT Boolean logic and express EYFP under the appropriate conditions[21]. Any Boolean logic formula can theoretically be evaluated using this approach.

Post-translational control. Methods for computational design and prediction of engineered protein structure and folding are becoming more refined[39–45]. For example, Bowen *et al.* computationally designed a receptor, specifically an artificial trinitrotoluene (TNT) receptor[3], that allows plants to monitor hazardous substances in the environment. The TNT receptor is part of a 'detect and respond' circuit that makes use of pre-existing cell–cell communication elements in *Arabidopsis thaliana* (through the His kinase pathways) to determine when the receptor is activated. Upon TNT receptor activation, the genetic circuit causes rapid chlorophyll loss, thus changing the colour of the plant. One can imagine a myriad of other responses that could be wired into this genetic circuit[3]. Another mechanism for post-translational control is to use orthogonal ribosome–mRNA pairs that encode for synthetic amino acids for unique proteins[46]. Synthetic protein scaffolds provide yet another level of control and can be coupled to feedback motifs[14] (FIG. 1c). As we learn more about biological systems and improve our ability to design proteins, we can look forward to the creation of additional highly imaginative synthetic genetic parts.

The second wave of synthetic biology

We are now on the cusp of the second wave of synthetic biology, in which basic parts and modules need to be integrated to create systems-level circuitry. Many of the genetic circuits described thus far have been simple and are usually aimed at controlling isolated cellular functions. Over the past few years, however, activity in the field has intensified (TIMELINE), as reflected by an increased number of published experimental circuits (FIG. 2a). Surprisingly, the actual complexity of synthetic biological circuitry over this time period, as measured by the number of regulatory regions, has only increased slightly (FIG. 2b); it is possible that existing engineering design principles are too simplistic to efficiently create complex biological systems and have so far limited our ability to exploit the full potential of this field.

The challenges faced in creating larger functional systems out of modules are both exciting and daunting. The precise details of most biological environments are poorly understood. Thus, engineering biological systems probably requires both new design principles and the simultaneous advance of scientific understanding. To be effective, we need to learn more about the systems that we are manipulating and to dynamically incorporate this information into our design strategies. In cases in which precise information about the relevant biological processes is not available during the design phase, our synthetic circuits should be adaptive and intelligently account for unknowns. Whereas traditional engineering practices typically rely on the standardization of parts, the uncertain and intricate nature of biology makes standardization in the synthetic biology field difficult. Beyond typical circuit design issues, synthetic biologists must also account for cell death, crosstalk, mutations, intracellular,

Box 1 | **The international genetically engineered machine competition**

The international genetically engineered machine (iGEM) competition provides teams of undergraduates with an opportunity to design and create unique synthetic biology projects. Each year, the programme culminates in a meeting at the Massachusetts Institute of Technology, Cambridge, USA, where the teams present their work and receive recognition for their efforts. iGEM has resulted in numerous exciting projects since its inception in 2004 (see the iGEM website), some of which have led to publications. For example, the 2007 iGEM team from Brown University, Providence, Rhode Island, USA, created a tri-stable toggle switch[93]. Their proof-of-principle switch controls the stable expression of three reporter genes. Each of the three promoters expresses repressors for the other two promoters in the system. Such switches might be incorporated into various synthetic systems, for example to guide cellular patterning and to regulate development. The 2007 iGEM team from Davidson College, Davidson, North Carolina, USA, constructed a switch that operates by inverting DNA sequences between sites recognized by the invertase Hin[94]. Using Hin-based switches, their goal is to engineer a multicellular bacterial system that can solve the burnt pancake problem, a mathematical puzzle. The 2007 iGEM team from The University of Edinburgh, UK, developed a biosensor in *Escherichia coli* that detects low concentrations of arsenate, a toxic by-product of water contaminated with arsenic[95]. In the design of the Edinburgh team, the presence of arsenate activates the engineered expression of β-galactosidase, which causes a subsequent decrease in water pH[95]. This biosensor would allow simple and efficient detection of arsenic contamination in drinking water with the use of a pH indicator dye or electrode. Several other iGEM projects have focused on sensing hydrogen[96], temperature change[97] or iron[98], and these have led to full-scale laboratory projects.

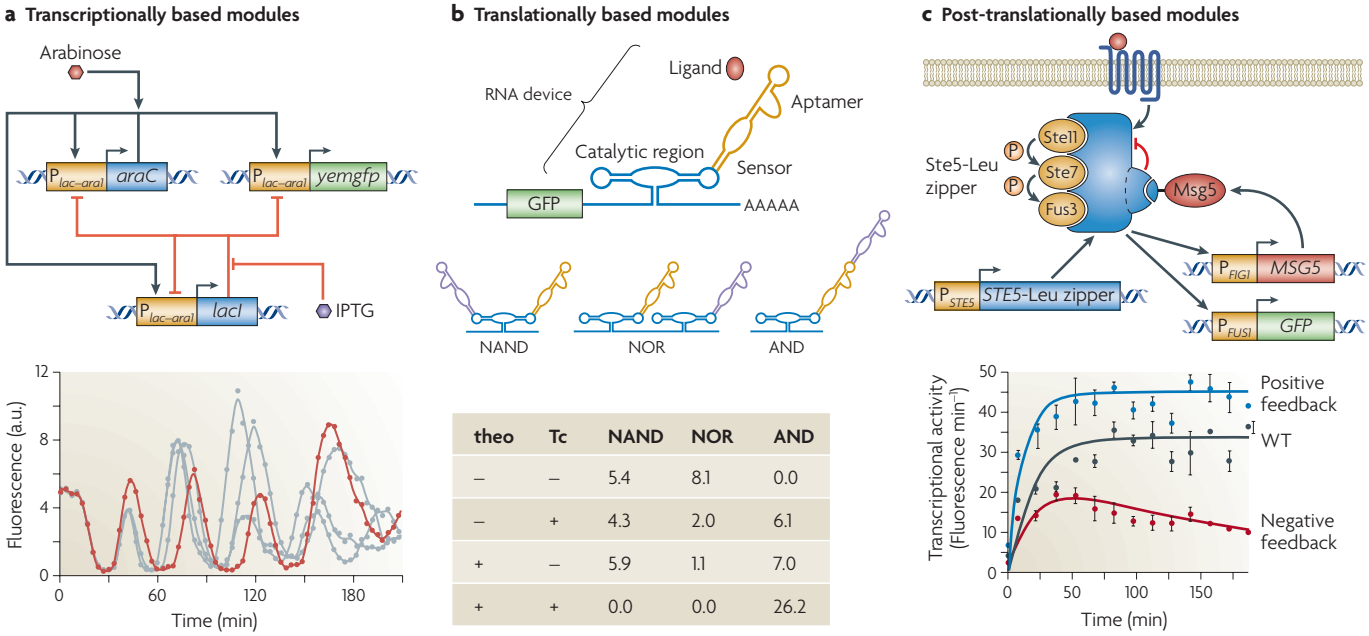


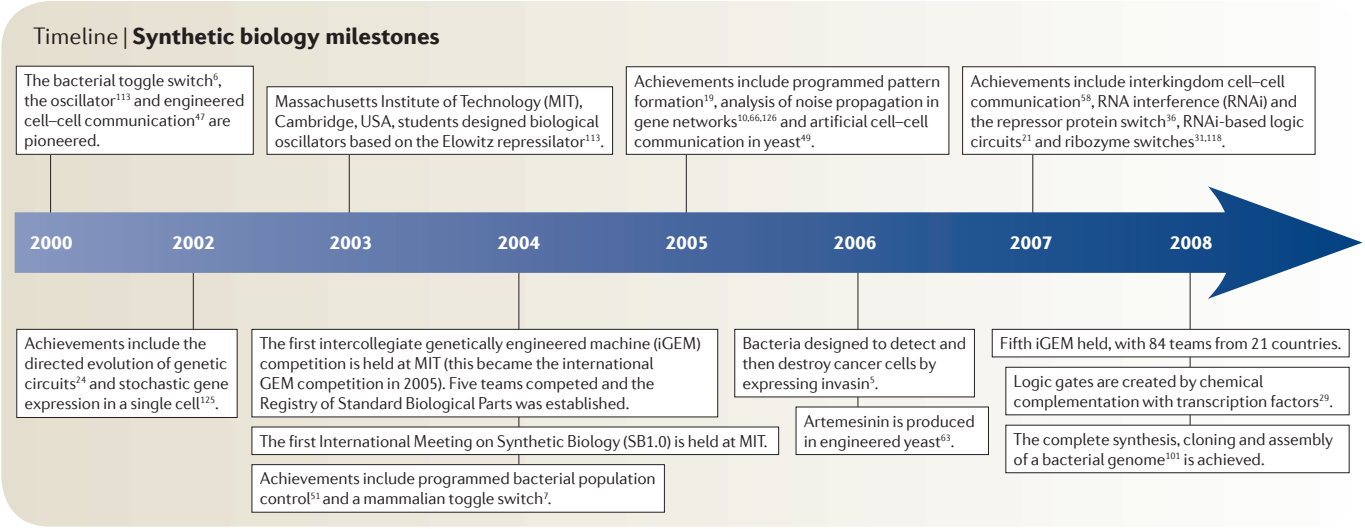
Figure 1 | **Modules based on transcriptional, translational and post-translational control. a** | The dual-feedback oscillator circuit[16] was constructed by placing a transcriptional repressor (*lacI*), a transcriptional activator (*araC*) and a reporter (monomeric yeast enhanced green fluorescent protein (*yemgfp*)) each downstream of an engineered promoter ($P_{lac–ara1}$). A network motif comprises positive feedback that is mediated by AraC (and modulated by arabinose) coupled with the negative feedback that is mediated by LacI (and modulated by isopropyl-β-D-thiogalactopyranoside (IPTG)), and can produce oscillations in a manner similar to transcriptional regulatory control mechanisms in certain circadian rhythms. Single cell fluorescence traces in the presence of 0.7% arabinose and 2 mM IPTG depict synchrony between the cells. Each trace (gray or red; the red trace is used to emphasize one trace for clarity) represents an individual cell. **b** | Engineered RNA devices[20] regulate protein translation by sensing and processing molecular inputs. The binding of a ligand to the RNA aptamer propagates a structural change to the catalytic region, either activating or deactivating ribozyme self cleavage. Self cleavage, in turn, results in quick degradation of the transcript by exonucleases. Shown are several multi-input configurations with different signal integration schemes that implement various logic functions: AND, NOR (NOT OR) and NAND (NOT AND). Any given scheme can implement multiple logic functions based on the choice of activation or deactivation (not shown). The chart depicts the response of each logic configuration in relative fluorescence units when the ligands theophylline (theo) and tetracycline (Tc) are present or absent. **c** | A scaffold protein phosphorylation system from yeast with a synthetic Leu zipper and modulator binding site[14]. Scaffold activation results in GFP expression. Engineered negative feedback from Msg5 results in delayed attenuation of output and downregulation of the phosphorylation cascade (red curve), whereas positive feedback with Ste50 (not shown) accelerates and amplifies the response (blue curve). Over time, the positive-feedback system reaches a steady state expression that is approximately 1.5 times greater than wild type (WT), whereas expression in the negative-feedback system steadily decreases. a.u, arbitrary units. Figure part **a** is modified, with permission, from *Nature* REF. 16 © (2008) Macmillan Publishers Ltd. All rights reserved. Figure part **b** is modified, with permission, from REF. 20 © (2008) American Association for the Advancement of Science. Figure part **c** is modified, with permission, from REF. 14 © (2008) American Association for the Advancement of Science.

intercellular and extracellular conditions, noise and other biological phenomena. A further difficult task is to correctly match suitable components in a designed system. As the number of system components grows, it becomes increasingly difficult to coordinate component inputs and outputs to produce the overall desired behaviour.

Our hope is that, in the second wave, synthetic biologists will formulate new and effective bioengineering design principles to address these challenges. This will allow us to readily combine modules into complex synthetic pathways and thereby create sophisticated cellular behaviours. Such systems-level bioengineering can synergistically target multiple pathways, symptoms or targets — such as multiple cell populations or organs — creating the potential for innovative environmental and therapeutic applications.

Synthetic ecosystems. Designing multicellular systems to exhibit finely tuned coordinated behaviour is a major challenge for synthetic biologists. Rather than being isolated in nature, organisms (whether eukaryotic or prokaryotic) sense their environment and frequently communicate with one another. Synthetic biologists have implemented artificial signalling pathways in various organisms, including acyl homoserine lactone (AHL) and acetate-based signalling in bacteria[11,47,48], plant hormone signalling in yeast[49] and nitric oxide (NO) signalling in mammalian cells[50]. By constructing and analysing synthetic multicellular systems that use such artificial signalling, we can improve our understanding of naturally occurring systems and devise effective design principles for building unique systems with new capabilities.

AHL
(Acyl-homoserine lactone.) A class of small signalling molecule that is commonly used in bacterial quorum sensing pathways.



In one project, *Escherichia coli* was engineered with a LuxI and LuxR AHL quorum sensing system coupled to the expression of a toxic protein (CcdB)[51] (FIG. 3a). Higher cell densities result in higher levels of CcdB expression in each cell (and subsequent cell death). However, different cells in the population have different levels of CcdB owing to signalling variations and gene expression noise. The system design exploits this noise to maintain a stable cell population over a long period of time. Engineered population control could be used in many situations, including industrial fermentation, immune responses and bioremediation.

Using the cooperative yeast ecosystem CoSMO (cooperation that is synthetic and mutually obligatory), scientists have engineered two normally non-interacting strains of *S. cerevisiae* to each synthesize and secrete a metabolite that is vital for the survival of the other strain, demonstrating artificial symbiotic behaviour between different strains of yeast[52] (FIG. 3b). The microbial consensus consortium uses a different approach to achieve coordinated behaviour between two types of cells. The system consists of two *E. coli* strains that have been engineered to communicate in a bidirectional manner using AHL signals, such that targeted gene expression is activated only if both cell populations are present at sufficient densities[53]. Cooperation between different coexisting cell types allows multicellular organisms to function and survive. A related synthetic bacterial predator–prey system illustrates the ability of two engineered *E. coli* populations to regulate each others' growth dynamics through bidirectional communication. Extensive theoretical and computational analysis on predator–prey systems in the literature shows that this type of interaction often generates interesting and complex oscillatory population dynamics[54–56]. The predator decreases the population of the prey, leading to a decrease in the population of predators owing to a lack of prey, thereby allowing reconstitution of the prey population[57]. Often, we gain important insights by constructing biological systems and comparing the experimental observations of these systems with long-standing computational and theoretical models.

Several other synthetic ecosystems have also been created, based on modulation of engineered mammalian and interkingdom cell–cell communication (between mammalian, yeast and bacterial cells) (FIG. 3b). Various mammalian strains were created to either produce or respond to compounds such as ampicillin, biotin and volatile acetaldehyde[58]. In some configurations, the engineered mammalian cells were co-cultured with bacteria and yeast that produce acetaldehyde, bacteria that produce erythromycin or bacteria that respond to ampicillin. Different multicellular configurations were tested and analysed, including commensalism, amensalism, mutualism, parasitism, third party inducible parasitism and predator–prey relationships[58]. Based on the lessons learned from these engineered systems, and by incorporating additional parts and modules into the circuits, we might be able to create even more complex multicellular systems with practical purposes, such as tissue development and the creation of auxiliary control organs.

Application-orientated systems. The notion of programmable cells, or programmable biology in general, is inspiring researchers to devise innovative solutions to currently unsolved problems. For example, one system aims to destroy tumours by using bacteria as a living computational therapeutic tool[5,59]. Upon the simultaneous detection of two conditions, using a two-input logical AND gate, engineered bacteria will invade and kill tumorous cells. Engineered bacteria must first detect that they inhabit a hypoxic environment (which is similar to the environment surrounding tumours *in vivo*). Second, exploiting the fact that some bacteria localize and thrive naturally in tumours, engineered bacteria use a synthetic quorum sensing pathway to detect high population density. When both conditions are satisfied, the bacteria express invasin, they bind specific mammalian integrin receptors and they initiate endocytosis. This system comprises sensors, actuators and responses that are modular and can be swapped with different engineered parts, allowing researchers to tailor the system to different cancers. Perhaps the most novel aspect of

Quorum sensing
Density-dependent bacterial behaviour that is regulated by cell–cell communication.

CcdB
A toxic protein that targets the *Escherichia coli* DNA gyrase, a bacterial topoisomerase II.

Commensalism
Non-competitive existence and growth.

Amensalism
The presence of one organism adversely affects the other.

Mutualism
The presence of each organism benefits the other.

Parasitism
One organism enables the other to survive at the expense of the first organism.

Third party inducible parasitism
One organism directs a second organism to allow a third organism to act as a parasite.

Integrin receptor
A human cell surface receptor that interacts with several components of the extracellular matrix, including fibronectin.

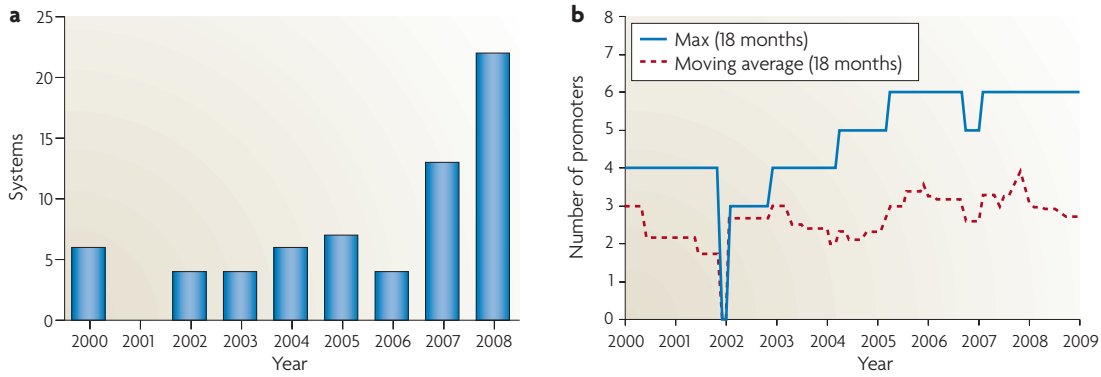


Figure 2 | **The progression of synthetic biology.** We sampled publications that describe the construction and characterization of complete synthetic biological circuits from 2000 to 2008. Each publication can contain more than one circuit (each circuit is referenced here as a system). **a** | The number of synthetic systems in publication. The number of new synthetic biological systems increased moderately from 2000 to 2008. **b** | The complexity of synthetic systems in publication. For the purposes of this analysis, we define complexity as the number of regulatory regions (promoters) comprising any given synthetic system. Shown are 18-month moving window averages and maximum values. Although the overall number of synthetic systems has increased over a 9 year span (as shown in part **a**), the complexity of published systems seems to have reached a plateau (at least for now). The analysis does not include partially constructed systems (such as unpublished international genetically engineered machine (iGEM) competition projects; BOX 1) or patent applications.

this system is the ability to engineer cells that integrate multiple sources of information to make decisions.

Another exciting domain for applications of synthetic biology emerges from the combination of synthetic biology with metabolic engineering. Groups are engineering parallel metabolic systems that interface with cellular metabolic machinery to provide cost effective chemical and drug synthesis[60]. The Keasling group, for example, successfully engineered a synthetic metabolic pathway based on the mevalonate-dependent isoprenoid pathway of *S. cerevisiae* into *E. coli*[61]. Isoprenoid is a viable terpenoid precursor that is used for the synthesis of various drugs, including artemisinin, an expensive antimalarial compound that is currently harvested from the rare *Artemisia annua* plant. By modifying their isoprenoid system, Keasling and colleagues constructed an artemisinin biosynthetic pathway in yeast, potentially providing an affordable and reliable source of highly potent antimalarial drugs[62,63]. Such transfer of genetic elements and pathways from one organism to another, and subsequent adaptation to an altogether different environment, is one of the unique features and greatest challenges of systems-level synthetic biology.

Methods that seem to be generally effective for systems-level bioengineering include: the optimization of heterologous gene expression, the creation and expression of novel enzymes tailored to new tasks and attention to distinctive features of the host organism. The tumour-killing bacteria and the creation of sophisticated and efficient metabolic pathways for drug and chemical synthesis are just two examples of applications of synthetic biology.

Minimal genomes and synthetic life. Another ambitious endeavour is the attempt to create minimized *de novo* genomes with the bases or genes that are required to support life[64]. For engineering purposes, minimized cells should provide more simplified 'chassis' than are currently known in nature. These will enable synthetic biologists to build less encumbered pathways, resulting in fewer undesired interactions with endogenous systems. These efforts might also lead to the construction of synthetic chromosomes that are easier to understand and manipulate than those fashioned by nature.

The creation of a minimized cell can be accomplished using either a top-down elimination approach or a bottom-up forward engineering approach. In a top-down approach, scientists begin with living cells and determine how much genetic material can be eliminated while still maintaining cell viability. For example, Venter's group is using a top-down approach with *Mycoplasma* spp. bacteria to help understand the minimal genome that is required for growth in culture. In a bottom-up approach, scientists attempt to create a cell *de novo* by constructing a membrane-bound compartment and then adding components. By using such an approach, the Murtas group currently estimates that approximately 100 genes are required to support basic life functions[64] (BOX 2). Importantly, there could be many definitions of a functional minimized cell. There will inevitably be multiple solutions to this complex problem, eventually providing synthetic biologists with a flexible toolbox of minimal genomes.

Open questions and challenges

Additional challenges still exist at the basic levels of synthetic biology. We still do not fully understand how to adequately handle noise, to efficiently design novel proteins with desired functions, or to design three-dimensional multimeric molecular structures. It is still not clear how small engineered modules that operate well in a given cellular context can be transferred readily into other contexts or organisms. Here, we explore several important remaining challenges and research opportunities for the second wave.

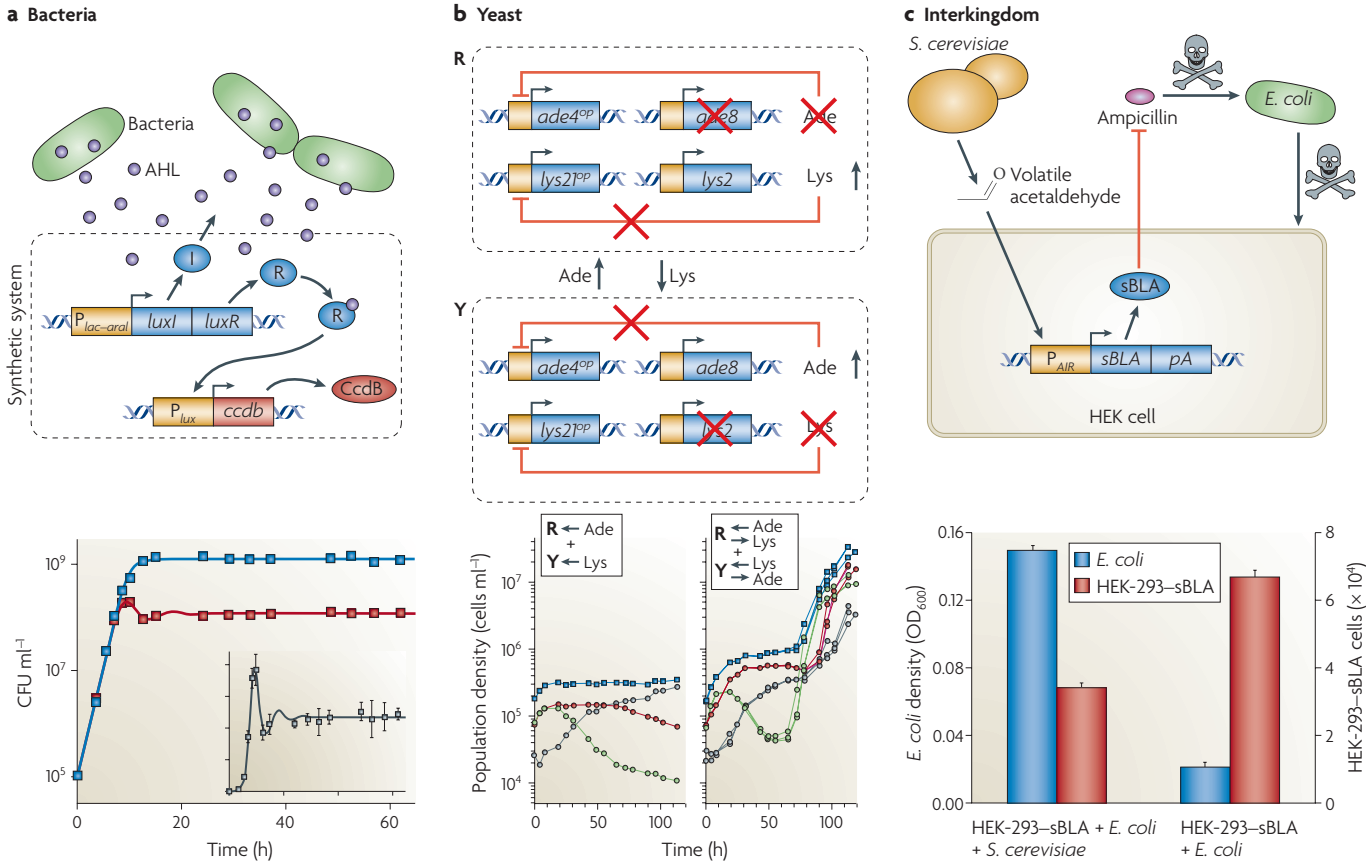


Figure 3 | **Synthetic multicellular systems. a** | Bacterial population control[51]. *Escherichia coli* engineered with *luxI* and *luxR* under control of the synthetic promoter $P_{lac-ara1}$. In the presence of isopropyl-β-D-thiogalactopyranoside (IPTG), LuxI is expressed and synthesizes the diffusible molecule acyl homoserine lactone (AHL). LuxR, a transcriptional regulator, is also expressed and activated by AHL. The LuxR–AHL complex activates P_{luxI} and induces expression of a killer protein, CcdB. High CcdB expression results in cell death. The graph shows experimental growth curves when the system is induced (red line) and uninduced (blue line). The inset shows the growth curve in linear scale for the induced case. Induced cells grow at the same rate as uninduced cells until a threshold density is reached. **b** | The CoSMO (cooperation that is synthetic and mutually obligatory) yeast system[52]. Yeast strain R synthesizes Lys but cannot synthesize adenine (Ade). A mutant enzyme involved in synthesis of Lys ($Lys21^{op}$) does not respond to negative feedback from high Lys levels. Conversely, yeast strain Y synthesizes adenine but cannot synthesize Lys. A similar mutation of an adenine synthesis pathway enzyme ($Ade4^{op}$) allows overproduction of adenine. By overproducing Lys and adenine, the two yeast strains provide each other with their missing metabolite. Population dynamics of fluorescent live R (red), fluorescent live Y (green), non-fluorescent dead (gray) and total (blue) cells are shown. The first graph shows co-cultures of two mutant strains washed free of supplements. The second graph depicts co-cultures of the overexpression mutants, showing that they can sustain each other. Multiple replicate culture data are superimposed in the second graph. **c** | Interkingdom third party inducible parasitism[58]. *Saccharomyces cerevisiae* metabolizes glucose, producing volatile acetaldehyde. Acetaldehyde activates the synthetic P_{AIR} promoter in cultured HEK-293 cells, which turns on expression of sBLA (S-β-lactamase). sBLA breaks down ampicillin in the media, which allows *E. coli* proliferation, depletion of nutrients and subsequent death of mammalian HEK-293 cells. The graph depicts *E. coli* and HEK-293–sBLA cell density in the presence and absence of *S. cerevisiae*-dependent activation of the circuit. CFU, colony-forming unit. Figure part **a** is modified, with permission, from *Nature* REF. 51 © (2004) Macmillan Publishers Ltd. All rights reserved. Figure part **b** is modified, with permission, from REF. 52 © (2007) National Academy of Sciences. Figure part **c** is modified, with permission, from REF. 58 © (2007) National Academy of Sciences.

Characterization, standardization and modularity. How can synthetic biologists combine many basic components effectively so that cellular behaviour is optimized and 'made to order'? The BioBrick standard[26,27] contains a useful DNA cloning mechanism to combine sequences of genetic elements (for the complete set of rules comprising the standard, see the OpenWetWare BioBrick standard web page). This standard is quickly gaining popularity because of the simplicity of the cloning mechanism and because comprehensive cataloguing of parts and modules is available at the Registry of Standard Biological Parts. However, extending the BioBrick standard to support the functional composition of these elements remains an important challenge. At the very least, we need to know the behavioural characteristics of each part (whether basic or composite), because, depending on the part, different information might be required. For example, basic transcriptional

BioBrick standard
A set of rules that define the assembly of DNA pieces or parts such that parts can be easily combined to form more complex parts.

Box 2 | **Construction of minimal genomes**

Mycoplasma spp. bacteria have extremely small genomes and are thought to approach the size of a predicted minimal genome. Using global transposon mutagenesis, Venter's group determined in single knockout experiments that at least 100 of 482 protein-coding genes in the genome of *Mycoplasma genitalium* can be eliminated individually while still sustaining bacterial life in culture[99]. By using mainly genes for membrane, DNA, RNA and protein synthesis, Murtas estimates that it might be possible to build a minimized genome of approximately 100 genes[64]. Forster and Church suggest that a minimal genome based on *Mycoplasma* spp. bacteria will be slightly larger and require at least 151 genes with a genome size of 113,000 base pairs[100].

Recently, Venter and colleagues constructed the full genome of *M. genitalium* and plan to transplant the chromosome into *Mycoplasma* cytoplasm for further testing[101]. The Murtas group is creating a liposomal structure for a minimal cell by using the fatty acid synthesis (FAS) type I pathway from bacteria[64] and testing the essential requirements for protein synthesis[102]. Rasmussen *et al.* is developing a protocell, a minimal self-replicating cellular machine, composed of a small lipid aggregate container with a lipophilic peptide nucleic acid anchored to the outside of the protocell[103]. Noireaux and Libchaber created a synthetic vesicle bioreactor, a step towards an artificial cell that uses plasmid-based transcription and translation systems[104,105]. Szostak's group is studying the permeability properties and requirements of primitive synthetic protocells[106], and Stano, Kuruma and Luisi are investigating properties such as compartmentalization, evolution and macromolecular synthesis in the context of the minimal cell[102,107–111]. Individually, these elements have been deemed important to the function of a living cell, but not all of these elements might be necessary in all designs.

regulatory elements might be characterized using PoPs[27], whereas certain basic post-translational regulatory elements might be characterized by their phosphorylation activity. Composite parts will need more sophisticated descriptions, perhaps using mathematical models that combine basic part characterizations. Crucially, models of composite parts must also be able to be combined themselves to describe even larger composite parts.

Current standardization efforts focus on creating basic part libraries with elements that can be easily combined and that function well together. One approach is to create parts that have similar kinetic characteristics and input–output thresholds. This can be achieved through genetic manipulations that are either simple (such as single base mutations) or more complex (such as domain shuffling). Standardization in other engineering disciplines allows components to be easily combined to form larger systems, but this approach relies on modularity between these components. A prevailing assumption in synthetic biology is that biological components are, or should be, modular as well. However, characterization, standardization and modularity are affected by cellular context. We cannot assume that a functional module in one cell type will work the same way in even a closely related cell type[65]. Researchers will always need to be mindful of intracellular, intercellular and extracellular environments during the design process. In addition, integrating synthetic components into a new cellular environment itself can significantly change the operational context of the cell. Parts and modules need to be characterized in systems and contexts of interest. These components will be most useful if they account for and adapt to the dynamics of the system. Alternatively, we could set out to engineer a completely new cellular environment or a subcellular environment (for example, organelle) in which synthetic parts and modules function orthogonally to one another.

Noise. Given the extent of genetic noise in biological systems, is it reasonable to expect that we can construct reliable, robust and predictable systems? Noise originates from many sources, including extrinsic environmental variations, fluctuations in gene expression, cell cycle variations, differences in the concentrations of metabolites and continuous mutational evolution[66]. The stochastic nature of biochemical reactions, even at the level of a single gene, can induce significant intrinsic genetic noise[10,66,67].

One way that synthetic biologists attempt to study noise is by building regulatory networks. Synthetic transcriptional cascades, for example, attenuate noise under some conditions and amplify noise under others[10,68]. The role of feedback in attenuating noise is complex and depends on both input level and feedback strength[69,70]. Noise is often assumed to have a negative influence on cellular processes and should be avoided when engineering genetic circuits that require exquisite control. However, for the synthetic population control system described above, noise has an integral role in maintaining stable cell densities[48]. Thus, noise can be viewed as a naturally sophisticated way for an isogenic population of cells to sense and respond to their environment. This can be advantageous to the survival of a cellular population during development and during times of stress.

Recent discoveries reveal that asynchrony in the cell cycle and apoptotic pathways can maintain different cellular populations with differing characteristics. The natural oscillations of p53 are asynchronous between cells and serve to ensure that the cells 'trickle' into apoptosis when a population is damaged, instead of killing off an entire population at once[71]. This noisy expression of p53 and the highly variable nuclear–cytoplasmic oscillations of the nuclear factor-κB (NF-κB) system are thought to have a role in stress response feedback loops[72,73]. Some viruses are thought to make use of noise to vary their phenotypes. Noise can determine the decision of HIV-1 to replicate or to remain latent in the genome, in part based on the site of pseudo-random genomic integration[74]. This decision is stochastic and is dependent on activation of the long terminal repeat region. Other regulatory networks, such as circadian rhythms, tend to be highly robust and resistant to noise[75]. Therefore, depending on the goal, noise and genetic fluctuation can either be useful or undesirable.

Global transposon mutagenesis
A top-down method by which non-essential genes in the chromosome of an organism are identified using random transposon insertion. Viable insertions are then sequenced.

PoPs
(Polymerase operations per second.) A measurement of the transcriptional activity of a gene that is defined as the number of RNA polymerase molecules passing a predefined point on the DNA each second.

Long terminal repeat
A sequence of DNA found in retroviruses that flanks genes. The sequence aids the process of integrating the retroviral DNA into the host genome.

Epigenetics. Epigenetic changes in gene expression occur in all organisms[76–78] and might be problematic for synthetic biologists. Epigenetics is typically defined as the heritable changes that propagate without changing the underlying DNA sequence, and might be an important mechanism for establishing temporal programmes of gene expression. How do synthetic biologists circumvent epigenetic processes or, alternatively, use epigenetic processes for their own purposes? For example, co-opting the processes of chromatin remodelling and DNA methylation to modulate heritable gene expression could be an effective technique to propagate information from a cell to daughter cells[79]. Furthermore, if an organism has already propagated an epigenetic modification, how can synthetic biologists change or release the modification and return a gene to its default state? We may be able to address these issues by designing systems that are sensitive to cross-generational changes in the state of DNA (for example, DNA methylation) or by designing cells that can synthetically maintain or modulate epigenetic 'memory' on demand.

Computational tools. Small modules can often be designed and optimized using intuition alone. Nonetheless, simple computational models are used extensively to refine and better understand such circuitry[80–85]. Perfect models of biology do not currently exist, so system optimization is carried out through iterative steps of mathematical design, genetic manipulation, experimental observation and model refinement. At each iterative step, the designer chooses to genetically manipulate a particular set of components. However, as engineered biological systems become more complex, it becomes impractical to test all or even a small percentage of the possible component combinations. Global sensitivity analysis is a potentially powerful computational technique with which we can identify the best elements to manipulate genetically — that is, the elements that have the largest effect on overall system performance[86].

Other computational algorithms developed by those who study system control theory are also likely to be useful design tools for synthetic biology. Reachability analysis and safety verification provide formal mechanisms to guarantee the behaviour and performance of systems[87]. For a given biological system, both algorithms first generate an exhaustive list of the states (within realistic boundaries): a single biological state can contain information about the expression levels of all relevant genes. A simple state transition occurs when the expression level of a single gene changes slightly. For each state, these algorithms then determine the set of possible simple state transitions. With this basic information, the algorithms can compute whether a transition from any given state to another is possible (perhaps in multiple steps). Reachability analysis enumerates all possible outcomes given an initial set of states, whereas safety verification ensures that a system never reaches undesired states or outcomes. Such tools can help to choose appropriate rate constants that function well in uncertain environments[87]; for example, by helping to determine which mutations to perform or by predicting behaviour when certain genetic parts are replaced by other parts.

Several important issues must be considered when using computational tools to engineer systems. First, the precise rate constants of genetic components are rarely known. Even if these are known, they might differ significantly when the cellular context changes. Second, computational design methods must be scalable to accommodate possible increases in the number of components and system complexity. Third, biological design tools should integrate all relevant aspects of the system, such that regulatory processes, metabolism, and physical and structural properties (such as membranes and organelles) are unified into a single modelling framework. Software developers are now beginning to devise and develop solutions to such issues in both existing and new design tools. TABLE 2 provides a partial list of relevant software tools and algorithms.

Programming abstractions. Another important consideration for biological design is the limited ability of humans to manage the details of increasingly complex systems. Although it is reasonable to expect a designer of small synthetic systems to choose and keep track of all genetic components (such as ribosome-binding sites, degradation tails and regulatory regions in promoters), this is not feasible for larger systems. To manage complexity, biological system designers must be able to create bioprograms using intuitive high level abstractions. Computational tools should automatically convert high level bioprograms into corresponding low level representations (such as genetic sequences). The bioprogramming language Proto was developed by Beal and Bachrach and used to simulate communities of cells programmed to form various spatial patterns. Proto supports programming statements for logic, arithmetic, comparisons, communication and sensing[88]. Programs written in Proto can be directly implemented with genetic regulatory networks, specifically using BioBricks. An important aspect of Proto is network optimization — the ability to automatically compile a complex and inefficient network design into a much simpler, yet functionally equivalent, network design. The Genetic Engineering of living Cells (GEC) project at Microsoft Research has similar aims[89].

Two other programming languages that are adaptable to biology are the Growing Point Language[90] (GPL) and the Origami Shape Language[91] (OSL). Similar to Proto, programs written in these languages can be automatically compiled (or transformed) to simpler representations that can be encoded as genetic sequences and tested experimentally. GPL uses a botanical metaphor of tropism where 'cells' both grow and propagate information using chemical gradients as guides for directionality. The growth process includes branching, death, length limitations, chemical secretions and merging. GPL provides a high-level descriptive language to specify pattern formation, which could be useful for cell-directed microfabrication. OSL programs use six types of folding operations on sheets of cells to create 2D and 3D structures.

Global sensitivity analysis
A quantitative evaluation of how perturbations in system components affect the overall behaviour of a system (for example, analysing how a change in the DNA-binding constant of a transcription factor affects the overall gene expression output of signal transduction pathway).

Table 2 | **Currently available computational tools for the design and analysis of genetic networks**

Computational tool	Use	Website‡
21U-RNA	Scoring 21U-RNA-associated upstream motifs	Bartel laboratory introduction to 21U-RNAs
Antimony*	Programming language describing synthetic biological devices	Deepak laboratory syntax guide
Athena*	Build and simulate genetic circuits (implemented in C++)	Deepak laboratory downloads
BioJade*	Synthetic biology design and simulation (implemented in Java)	BioJade
CAD of modular protein devices	Modular protein device algorithm using a backbone of scaffold proteins[82]	None
ESSA	RNA secondary structure analysis	ESSA
Evolutionary design of genetic networks *in silico*	Algorithm to evolve small gene networks (modules) that perform basic tasks, such as toggle switches or oscillators[81]	None
GeneDesign*	Editing protein sequences and generating oligos for protein construction (implemented in Perl)	Gene Design
GeNetDes*	Transcriptional network design tool using simulated annealing optimization	Genetdes
GenoCAD*	Design of complex genetic constructs from standard parts library	GenoCAD
MiRscan	Scoring of hairpins versus some experimentally verified microRNAs from *Caenorhabditis elegans or Caenorhabditis briggsae*	MiRscan
OptCircuit	Identifies circuit components and suggests circuit topologies to attain desired outcome[85]	None
PCEnv*	Environment for simulating various types of CellML models	OpenCell
PROTDES*	Computational protein design	PROTDES
Random Sampling-High Dimensional Model Representation	Global sensitivity analysis algorithm that is useful in optimizing genetic circuit properties not available from experiments or modelling[86]	None
Registry of Standard Biological Parts and Clotho*	Creation, cataloguing and public availability of modular biological parts that are extensively characterized; Clotho is a database for managing these parts	Registry of Standard Biological Parts and Clotho Development
RNA world website	Compendium of RNA software	RNA world
RNAdraw	RNA secondary structure analysis	RNAdraw
RNAMotif	Database search for RNA sequences that match a secondary structure motif	Rutgers Case Group
RNAstructure	RNA secondary structure analysis	RNAstructure
RnaViz	RNA secondary structure images	RnaViz
Rosetta package	Design of protein-binding peptide sequences and protein engineering	Rosetta @ home and Rosetta Commons
RoVerGeNe*	Tool to analyse and tune gene networks	RoVerGeNe
SynBioSS*	Suite of programs to generate and simulate synthetic biological networks	SynBioSS
Tinkercell	Synthetic biology CAD program	Tinkercell
UNAFold software	Nucleic acid folding and hybridization	UNAFold software
Vienna RNA package	RNA secondary structure	Vienna RNA package

*Described in detail at the OpenWetWare site. ‡See Further information for full addresses. CAD, computer-aided design.

The folding operations can take points as references and fold cellular sheets between or on to these points. Similarly, lines can serve as references for sheet-folding operations. The ability of OSL to describe 3D shapes might be a convenient tool for specifying the creation of whole tissues without the need to consider the specific genetic engineering requirements for each individual cell.

To be effective, these and other high-level design tools must not only provide useful programming metaphors, they must also manage the subtleties and challenges of the biological substrate. Programming tools should take into account noise, crosstalk, cell death, morphogen gradient effects, feedback mechanisms, threshold matching between parts and modules, mutations, cellular motility, environmental conditions and cellular context (among other factors). Until we obtain a more detailed and precise understanding of biology, iteration between high-level design and experimentation will have an important role in the implementation and optimization of complex biological systems.

Ramifications of synthetic biological treatments. As responsible researchers, we need to understand the logical consequences of our actions when we design synthetic circuitry in live cells or organisms. In general, we need to be able to engineer systems that withstand or correct mutations and hence allow engineered cells or organisms to remain operational over longer periods of time. The relevant timescale of industrial applications of engineered bacteria could be days to weeks, whereas for tissue engineering it might be years or decades. Each application of synthetic biology will require unique considerations. For example, what are the consequences of adding artificial circuits to plants? As the international debate over genetically modified organisms informs us, we need to be mindful of cross-pollination with native species, the possible disruption of ecosystems and the introduction of non-native and perhaps deleterious resistances. Similar concerns must be addressed when engineering bacteria for environmental remediation. Many synthetic biology workshops and conferences are devoting entire tracks to such discussions, and several synthetic biology groups and centers are engaged in discussion with the public and policy makers regarding ethical and safety issues.

Another important issue is whether synthetic biology provides a legitimate way to cure disease. As we move towards one of our intended goals of human therapeutics, we need to be exceptionally careful when making claims, reporting results and designing systems. What are the possible ramifications of a human with 'mosaic' genetically engineered cells and how is the transplantation of these cells any different from, for example, an artificial knee, arm or hip? The initial systems will inevitably be developed in the laboratory where many of these questions cannot be addressed. It is important to anticipate potentially dangerous problems, such as immune responses, lack of compartmentalization or cancerous growth *in vivo*, before the system is even tested in animals. The failure of early gene therapy attempts was due to events that might have been tested or anticipated in longer trials[92]. Notably, synthetic biology has at least one important advantage over other approaches: researchers can design sophisticated checkpoints and fail-safe switches to ensure that the system operates correctly, even in highly complex environments.

Conclusions and future perspectives

Thus far, the behaviour of most synthetic modules and systems has been studied in isolated cellular populations in the laboratory. Some systems, such as synthetic metabolic pathways for drug production, are inherently suitable for use in *in vitro* conditions. Many other systems, however, will have their greatest impact when introduced into a living organism or a wider ecological setting. Whether engineered cells create new sources of energy, perform bioremediation or sense and destroy pathogens, the environment in which the cells operate will have a significant impact on functionality. How will researchers ensure that devices, modules and systems that work well in a laboratory environment also work well in much more complex natural environments? One method is to gradually step through increasing levels of complexity that attempt to more closely mimic the natural environment, evaluating and adapting the engineered system to account for each new set of conditions in a logical progression. A more elegant solution might be to design systems that can evolve and adapt to such environmental and contextual differences.

As we move into the second wave of synthetic biology, it is important to incorporate classical systems-level engineering practices, such as modularity, component testing, standards, interfaces, libraries of parts and computer-aided design. However, we cannot ignore the unique features of biological systems that might hamper the straightforward use of these well-meaning and well-organized practices. For example, combining well-characterized modules in biological systems often creates behaviours that are not intuitive or easily predictable and, therefore, that are not easy to model in a classical sense. Current computational models and techniques are inadequate for biological system design, especially given the inherent level of uncertainty in biology. However, rather than viewing certain features of biology as problematic, it might be better to take advantage of them. We need to create new design and computational tools that take biological variability, uncertainty and evolution into account and allow us to develop systems that are more reliable. By addressing all of these issues, the second wave of synthetic biology will have a significant and pivotal impact on our ability to solve biological and environmental problems in a more predictable, robust and efficient manner.

1. Cases, I. & de Lorenzo, V. Genetically modified organisms for the environment: stories of success and failure and what we have learned from them. *Int. Microbiol.* **8**, 213–222 (2005).
2. Antunes, M. S. *et al.* A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity. *Plant Biotechnol. J.* **4**, 605–622 (2006).
3. Bowen, T. A., Zdunek, J. K. & Medford, J. I. Cultivating plant synthetic biology from systems biology. *New Phytol.* **179**, 583–587 (2008).
4. Savage, D. F., Way, J. & Silver, P. A. Defossiling fuel: how synthetic biology can transform biofuel production. *ACS Chem. Biol.* **3**, 13–16 (2008).
5. Anderson, J. C., Clarke, E. J., Arkin, A. P. & Voigt, C. A. Environmentally controlled invasion of cancer cells by engineered bacteria. *J. Mol. Biol.* **355**, 619–627 (2006).
Introduces the novel idea of engineering bacteria that integrate information from multiple sources to determine whether mammalian cells are cancerous and, if so, invade and destroy these cells.
6. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
A seminal paper that helped to establish the field of synthetic biology, demonstrating how a bistable synthetic gene network can be constructed by connecting two mutually inhibitory transcriptional repressors.
7. Kramer, B. P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nature Biotech.* **22**, 867–870 (2004).
8. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
9. Ham, T. S., Lee, S. K., Keasling, J. D. & Arkin, A. P. Design and construction of a double inversion recombination switch for heritable sequential genetic memory. *PLoS ONE* **3**, e2815 (2008).
10. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA* **102**, 3581–3586 (2005).
11. Basu, S., Mehreja, R., Thiberge, S., Chen, M. T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc. Natl Acad. Sci. USA* **101**, 6355–6360 (2004).
12. Weber, W. *et al.* A synthetic time-delay circuit in mammalian cells and mice. *Proc. Natl Acad. Sci. USA* **104**, 2643–2648 (2007).
13. Weber, W., Kramer, B. P. & Fussenegger, M. A genetic time-delay circuitry in mammalian cells. *Biotechnol. Bioeng.* **98**, 894–902 (2007).
14. Bashor, C. J., Helman, N. C., Yan, S. & Lim, W. A. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* **319**, 1539–1543 (2008).

Fused a Leu-zipper domain to the Ste5 scaffold protein and connected positive- and negative-feedback elements to this pathway. This shows how synthetic signalling pathways can be engineered using new phosphorylation interactions.
15. Goh, K. I., Kahng, B. & Cho, K. H. Sustained oscillations in extended genetic oscillatory systems. *Biophys. J.* **94**, 4270–4276 (2008).
16. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
A recent publication demonstrating how a relaxation oscillator can be built by connecting a transcriptional activator with a transcriptional repressor. Fine-tuning of oscillatory dynamics was achieved with external application of inducers.
17. Swinburne, I. A., Miguez, D. G., Landgraf, D. & Silver, P. A. Intron length increases oscillatory periods of gene expression in animal cells. *Genes Dev.* **22**, 2342–2346 (2008).
18. Tigges, M., Marquez-Lago, T. T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312 (2009).
19. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
An engineered cell–cell communication using AHL was coupled with several variants of a synthetic signal processing network to produce a multicellular system that is capable of forming a range of spatial patterns.
20. Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
Novel RNA structures were constructed by fusing various sensor regions (aptamers) to catalytic regions (ribozymes). These RNA devices implement several important two-input logic functions that regulate protein expression at the translational level based on small molecule inputs.
21. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotech.* **25**, 795–801 (2007).
Implements RNA interference-based logic circuits that regulate protein expression based on up to five simultaneous endogenous molecular inputs. Any logic formula or statement can be represented with the scheme presented in this paper.
22. Haseltine, E. L. & Arnold, F. H. Synthetic gene circuits: design with directed evolution. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 1–19 (2007).
23. Nomura, Y. & Yokobayashi, Y. Dual selection of a genetic switch by a single selection marker. *Biosystems* **90**, 115–120 (2007).
24. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99**, 16587–16591 (2002).
Shows the use of directed evolution to optimize the behaviour of an engineered transcriptional cascade to meet desired input–output specifications.
25. Basu, S. & Weiss, R. in *First Workshop on Non-Silicon Computation (NSC-1)* 54–61 (Cambridge, 2002).
26. Shetty, R. P., Endy, D. & Knight, T. F. Jr. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5 (2008).
27. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nature Biotech.* **26**, 787–793 (2008).
Careful and thorough characterization of cell–cell communication BioBricks from the Registry of Standard Biological Parts, showing the type of information that is useful for describing the behaviour of parts.
28. Murphy, K. F., Balazsi, G. & Collins, J. J. Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl Acad. Sci. USA* **104**, 12726–12731 (2007).
29. Bronson, J. E., Mazur, W. W. & Cornish, V. W. Transcription factor logic using chemical complementation. *Mol. Biosyst.* **4**, 56–58 (2008).
30. Ham, T. S., Lee, S. K., Keasling, J. D. & Arkin, A. P. A tightly regulated inducible expression system utilizing the *fim* inversion recombination switch. *Biotechnol. Bioeng.* **94**, 1–4 (2006).
31. Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc. Natl Acad. Sci. USA* **104**, 14283–14288 (2007).
32. Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nature Biotech.* **24**, 1558–1564 (2006).
33. Yoshida, W. & Yokobayashi, Y. Photonic Boolean logic gates based on DNA aptamers. *Chem. Commun. (Camb.)*, **2**, 195–197 (2007).
34. Nomura, Y. & Yokobayashi, Y. Reengineering a natural riboswitch by dual genetic selection. *J. Am. Chem. Soc.* **129**, 13814–13815 (2007).
35. Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotech.* **22**, 841–847 (2004).
36. Deans, T. L., Cantor, C. R. & Collins, J. J. A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* **130**, 363–372 (2007).
37. Rackham, O. & Chin, J. W. Cellular logic with orthogonal ribosomes. *J. Am. Chem. Soc.* **127**, 17584–17585 (2005).
38. Rajendran, M. & Ellington, A. D. Selection of fluorescent aptamer beacons that light up in the presence of zinc. *Anal. Bioanal. Chem.* **390**, 1067–1075 (2008).
39. Wollacott, A. M., Zanghellini, A., Murphy, P. & Baker, D. Prediction of structures of multidomain proteins from structures of the individual domains. *Protein Sci.* **16**, 165–175 (2007).
40. Andre, I., Bradley, P., Wang, C. & Baker, D. Prediction of the structure of symmetrical protein assemblies. *Proc. Natl Acad. Sci. USA* **104**, 17656–17661 (2007).
41. Yarov-Yarovoy, V., Schonbrun, J. & Baker, D. Multipass membrane protein structure prediction using Rosetta. *Proteins* **62**, 1010–1025 (2006).
42. Dantas, G. *et al.* Mis-translation of a computationally designed protein yields an exceptionally stable homodimer: implications for protein engineering and evolution. *J. Mol. Biol.* **362**, 1004–1024 (2006).
43. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein–protein interface. *J. Mol. Biol.* **361**, 195–208 (2006).
44. Meiler, J. & Baker, D. ROSETTALIGAND: protein–small molecule docking with full side-chain flexibility. *Proteins* **65**, 538–548 (2006).
45. Sood, V. D. & Baker, D. Recapitulation and design of protein binding peptide structures and sequences. *J. Mol. Biol.* **357**, 917–927 (2006).
46. Filipovska, A. & Rackham, O. Building a parallel metabolism within the cell. *ACS Chem. Biol.* **3**, 51–63 (2008).
47. Weiss, R. & Knight, T. F. Jr. in *DNA6: Sixth International Meeting on DNA Based Computers* (eds Condon, A. & Rozenberg, G.) 1–16 (LNCS 2054, Springer, Berlin, 2001).
48. Bulter, T. *et al.* Design of artificial cell–cell communication using gene and metabolic networks. *Proc. Natl Acad. Sci. USA* **101**, 2299–2304 (2004).
49. Chen, M. T. & Weiss, R. Artificial cell–cell communication in yeast *Saccharomyces cerevisiae* using signaling elements from *Arabidopsis thaliana*. *Nature Biotech.* **23**, 1551–1555 (2005).
50. Wang, W. D., Chen, Z. T., Kang, B. G. & Li, R. Construction of an artificial intercellular communication network using the nitric oxide signaling elements in mammalian cells. *Exp. Cell Res.* **314**, 699–706 (2008).
51. You, L., Cox, R. S. 3rd, Weiss, R. & Arnold, F. H. Programmed population control by cell–cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
52. Shou, W., Ram, S. & Vilar, J. M. Synthetic cooperation in engineered yeast populations. *Proc. Natl Acad. Sci. USA* **104**, 1877–1882 (2007).
53. Brenner, K., Karig, D. K., Weiss, R. & Arnold, F. H. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc. Natl Acad. Sci. USA* **104**, 17300–17304 (2007).
54. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton University Press, Princeton, 1974).
55. Murray, J. D. & ebrary Inc. *An Introduction*. 3rd edn (Springer, New York; London, 2002).
56. Renshaw, E. *Modelling Biological Populations in Space and Time* (Cambridge Univ. Press, Cambridge, 1993).
57. Balagadde, F. K. *et al.* A synthetic *Escherichia coli* predator–prey ecosystem. *Mol. Syst. Biol.* **4**, 187 (2008).
58. Weber, W., Daoud-El Baba, M. & Fussenegger, M. Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proc. Natl Acad. Sci. USA* **104**, 10435–10440 (2007).
Implements cell–cell communication between mammalian cells and between cells of different organisms, and uses this to construct a range of multicellular systems with different 'rules' for cell survival and proliferation.
59. Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3**, 133 (2007).
60. Ajikumar, P. K. *et al.* Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.* **5**, 167–190 (2008).
61. Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotech.* **21**, 796–802 (2003).
62. Hale, V., Keasling, J. D., Renninger, N. & Diagana, T. T. Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs. *Am. J. Trop. Med. Hyg.* **77**, 198–202 (2007).
63. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
Shows that yeast can be engineered with a plant metabolic pathway to efficiently produce a drug precursor that is difficult and expensive to obtain using traditional methods.
64. Murtas, G. Question 7: construction of a semi-synthetic minimal cell: a model for early living cells. *Orig. Life Evol. Biosph.* **37**, 419–422 (2007).
65. Bagh, S. *et al.* Plasmid-borne prokaryotic gene expression: sources of variability and quantitative system characterization. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **77**, 021919 (2008).
66. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
Introduces the notion of intrinsic and extrinsic noise as important constituents of protein expression and genetic regulation.
67. Blake, W. J., Kærn, M. Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
68. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & Van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427**, 737–740 (2004).
69. Becskei, A., Seraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535 (2001).
70. Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M. & Serrano, L. Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol. Syst. Biol.* **2**, 41 (2006).
71. Geva-Zatorsky, N. *et al.* Oscillations and variability in the p53 system. *Mol. Syst. Biol.* **2**, 2006.0033 (2006).
72. Nelson, D. E. *et al.* Oscillations in NF-κB signaling control the dynamics of gene expression. *Science* **306**, 704–708 (2004).
73. Nelson, D. E., See, V., Nelson, G. & White, M. R. Oscillations in transcription factor dynamics: a new way to control gene expression. *Biochem. Soc. Trans.* **32**, 1090–1092 (2004).
74. Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. & Schaffer, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122**, 169–182 (2005).
75. Vilar, J. M., Kueh, H. Y., Barkai, N. & Leibler, S. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl Acad. Sci. USA* **99**, 5988–5992 (2002).
76. Kininis, M. & Kraus, W. L. A global view of transcriptional regulation by nuclear receptors: gene expression, factor localization, and DNA sequence analysis. *Nucl. Recept. Signal.* **6**, e005 (2008).
77. Costa, F. F. Non-coding RNAs, epigenetics and complexity. *Gene* **410**, 9–17 (2008).
78. Tang, W. Y. & Ho, S. M. Epigenetic reprogramming and imprinting in origins of disease. *Rev. Endocr. Metab. Disord.* **8**, 173–182 (2007).
79. Zheng, Y. G., Wu, J., Chen, Z. & Goodman, M. Chemical regulation of epigenetic modifications: opportunities for new cancer therapy. *Med. Res. Rev.* **28**, 645–687 (2008).
80. Tuttle, L. M., Salis, H., Tomshine, J. & Kaznessis, Y. N. Model-driven designs of an oscillating gene network. *Biophys. J.* **89**, 3873–3883 (2005).

81. Francois, P. & Hakim, V. Design of genetic networks with specified functions by evolution *in silico*. *Proc. Natl Acad. Sci. USA* **101**, 580–585 (2004).
One of the first papers to use a computational algorithm for designing genetic circuitry to achieve desired cellular behaviours.
82. Salis, H. & Kaznessis, Y. N. Computer-aided design of modular protein devices: Boolean AND gene activation. *Phys. Biol.* **3**, 295–310 (2006).
83. Sotiropoulos, V. & Kaznessis, Y. N. Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. *BMC Syst. Biol.* **1**, 7 (2007).
84. Tomshine, J. & Kaznessis, Y. N. Optimization of a stochastically simulated gene network model via simulated annealing. *Biophys. J.* **91**, 3196–3205 (2006).
85. Dasika, M. S. & Maranas, C. D. OptCircuit: an optimization based method for computational design of genetic circuits. *BMC Syst. Biol.* **2**, 24 (2008).
86. Feng, X. J. *et al.* Optimizing genetic circuits by global sensitivity analysis. *Biophys. J.* **87**, 2195–2202 (2004).
87. Batt, G., Yordanov, B., Weiss, R. & Belta, C. Robustness analysis and tuning of synthetic gene networks. *Bioinformatics* **23**, 2415–2422 (2007).
88. Beal, J. & Bachrach, J. in *Spatial Computing Workshop, at 2nd IEEE International Conference on Self-Adaptive and Self-Organizing Systems* (Venice, Italy, 2008).
89. Pedersen, M. & Phillips, A. Towards programming languages for genetic engineering of living cells. *J. R. Soc. Interface* 15 Apr 2009 (doi:10.1098/rsif.2008.0516.focus).
90. Abelson, H. *et al.* Amorphous computing. *Comm. ACM* **43**, 74–82 (2001).
91. Nagpal, R., Kondacs, A. & Chang, C. in *AAAI Spring Symposium on Computational Synthesis: From Basic Building Blocks to High Level Functionality* (AAAI, Stanford, 2003).
92. Rubanyi, G. M. The future of human gene therapy. *Mol. Aspects Med.* **22**, 113–142 (2001).
93. J. Lohmueller, N. N. *et al.* Progress toward construction and modelling of a tri-stable toggle switch in *E. Coli*. *IET Synth. Biol.* **1**, 25–28 (2007).
94. Haynes, K. A. *et al.* Engineering bacteria to solve the burnt pancake problem. *J. Biol. Eng.* **2**, 8 (2008).
95. Aleksic, J. *et al.* Development of a novel biosesnsor for the detection of arsenic in drinking water. *IET Synth. Biol.* **1**, 87–90 (2007).
96. Flynn, B. P. *et al.* Hydrogen detector testing and design using synthetic biology and fluorescence photo analysis. *IET Synth. Biol.* **1**, 83–86 (2007).
97. Hung, A. *et al.* The 'Cell-See-Us' cellular thermometer. *IET Synth. Biol.* **1**, 79–82 (2007).
98. Quintero, A. *et al.* A microbial biosensor device for iron detection under UV irradiation. *IET Synth. Biol.* **1**, 71–73 (2007).
99. Glass, J. I. *et al.* Essential genes of a minimal bacterium. *Proc. Natl Acad. Sci. USA* **103**, 425–430 (2006).
100. Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2**, 45 (2006).
A thorough perspective on the requirements for creating chemical systems that can replicate and evolve. Describes accomplishments to date and suggests future goals and challenges.
101. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
102. Murtas, G., Kuruma, Y., Bianchini, P., Diaspro, A. & Luisi, P. L. Protein synthesis in liposomes with a minimal set of enzymes. *Biochem. Biophys. Res. Commun.* **363**, 12–17 (2007).
103. Rasmussen, S. *et al.* Evolution: transitions from nonliving to living matter. *Science* **303**, 963–965 (2004).
104. Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl Acad. Sci. USA* **101**, 17669–17674 (2004).
105. Noireaux, V., Bar-Ziv, R., Godefroy, J., Salman, H. & Libchaber, A. Toward an artificial cell based on gene expression in vesicles. *Phys. Biol.* **2**, P1–P8 (2005).
106. Mansy, S. S. *et al.* Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**, 122–125 (2008).
107. Fiordemondo, D. & Stano, P. Lecithin-based water-in-oil compartments as dividing bioreactors. *ChemBioChem* **8**, 1965–1973 (2007).
108. Kuruma, Y. Question 7: biosynthesis of phosphatidic acid in liposome compartments — toward the self-reproduction of minimal cells. *Orig. Life Evol. Biosph.* **37**, 409–413 (2007).
109. Luisi, P. L. Chemical aspects of synthetic biology. *Chem. Biodivers.* **4**, 603–621 (2007).
110. Stano, P. Question 7: new aspects of interactions among vesicles. *Orig. Life Evol. Biosph.* **37**, 439–444 (2007).
111. Stano, P. & Luisi, P. L. Basic questions about the origins of life: proceedings of the Erice international school of complexity (fourth course). *Orig. Life Evol. Biosph.* **37**, 303–307 (2007).
112. Ogawa, A. & Maeda, M. An artificial aptazyme-based riboswitch and its cascading system in *E. coli*. *ChemBioChem* **9**, 206–209 (2008).
113. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
Another seminal paper for synthetic biology, experimentally demonstrating the construction of a synthetic genetic oscillator. This paper also fuelled interest in improving our understanding of gene expression noise.
114. Khlebnikov, A., Skaug, T. & Keasling, J. D. Modulation of gene expression from the arabinose-inducible *araBAD* promoter. *J. Ind. Microbiol. Biotechnol.* **29**, 34–37 (2002).
115. Yang, W. & Paschen, W. Conditional gene silencing in mammalian cells mediated by a stress-inducible promoter. *Biochem. Biophys. Res. Commun.* **365**, 521–527 (2008).
116. Bayer, T. S. & Smolke, C. D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotech.* **23**, 337–343 (2005).
117. Win, M. N., Klein, J. S. & Smolke, C. D. Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. *Nucleic Acids Res.* **34**, 5670–5682 (2006).
118. Win, M. N. & Smolke, C. D. RNA as a versatile and powerful platform for engineering genetic regulatory tools. *Biotechnol. Genet. Eng. Rev.* **24**, 311–346 (2007).
119. Saito, H. & Inoue, T. RNA and RNP as new molecular parts in synthetic biology. *J. Biotechnol.* **132**, 1–7 (2007).
120. Wickiser, J. K., Cheah, M. T., Breaker, R. R. & Crothers, D. M. The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44**, 13404–13414 (2005).
121. Voigt, C. A. Community collection. *Registry of Standard Biological Parts* [online] http://partsregistry.org/Ribosome_Binding_Sites/Prokaryotic/Constitutive/Community_Collection (2008).
122. Tikole, S. & Sankararamakrishnan, R. Prediction of translation initiation sites in human mRNA sequences with AUG start codon in weak Kozak context: A neural network approach. *Biochem. Biophys. Res. Commun.* **369**, 1166–1168 (2008).
123. Dai, C. *et al.* Improved fusion protein expression of EGFP via the mutation of both Kozak and the initial ATG codon. *Cell. Mol. Biol. Lett.* **12**, 362–369 (2007).
124. Rechsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21**, 267–271 (1996).
125. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
126. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).

Acknowledgements
We apologize to those colleagues whose work we do not directly cite owing to space limitations. We especially thank T. Dan-Cohen for critically reading the manuscript and providing much needed feedback and edits. We thank C. DeHart for her ideas and feedback. We would also like to thank past and present members of the Weiss laboratory for engaging daily discussions and for fostering a creative research environment. The Weiss laboratory is supported by the National Institutes of Health, the National Science Foundation, the Army Research Office and the Office of Naval Research.

DATABASES

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
araC | *GAL1* | *lacI*
UniProtKB: http://www.uniprot.org
CCDB | FimE | Msg5 | Ste50

FURTHER INFORMATION

Ron Weiss' homepage: http://weisswebserver.ee.princeton.edu
Bartel laboratory introduction to 21U-RNAs: http://web.wi.mit.edu/bartel/pub/softwareDocs/TouRnaMotif/Introduction.html
BioJade: http://web.mit.edu/jagoler/www/biojade
Clotho Development: http://biocad-server.eecs.berkeley.edu/wiki/index.php/Clotho_Development
Deepak laboratory downloads: http://staff.washington.edu/deepakc/downloads
Deepak laboratory syntax guide: http://staff.washington.edu/deepakc/PartSyntax.pdf
ESSA: http://www.inra.fr/internet/Departements/MIA/T/essa/Doc/essa_home.html
Gene Design: http://baderlab.bme.jhu.edu/gd
Genetdes: http://soft.synth-bio.org/genetdes.html
GenoCAD: http://www.genocad.org/genocad
iGEM: http://www.igem.org
MiRscan: http://genes.mit.edu/mirscan
OpenCell: http://www.cellml.org/tools/opencell
OpenWetWare: http://openwetware.org/index.php?title=Computational_Tools&oldid=260443
OpenWetWare BioBrick standard: http://openwetware.org/wiki/Biobrick_standard
PROTDES: http://soft.synth-bio.org/protdes.html
Registry of Standard Biological Parts: http://partsregistry.org
RNAdraw: http://www.rnadraw.com
RNAstructure: http://rna.urmc.rochester.edu/rnastructure.html
RnaViz: http://rnaviz.sourceforge.net
RNA world: http://www.imb-jena.de/RNA.html
Rosetta @ home: http://boinc.bakerlab.org/rosetta
Rosetta Commons: http://www.rosettacommons.org
RoVerGeNe: http://iasi.bu.edu/~batt/rovergene/rovergene.htm
Rutgers Case Group: http://casegroup.rutgers.edu
SynBioSS: http://synbioss.sourceforge.net
Tinkercell: http://www.tinkercell.com
UNAFold software: http://dinamelt.bioinfo.rpi.edu/unafold
Vienna RNA package: http://www.tbi.univie.ac.at/~ivo/RNA

ALL LINKS ARE ACTIVE IN THE ONLINE PDF