SYNTHETIC BIOLOGY

Steven A. Benner and A. Michael Sismour

Abstract | Synthetic biologists come in two broad classes. One uses unnatural molecules to reproduce emergent behaviours from natural biology, with the goal of creating artificial life. The other seeks interchangeable parts from natural biology to assemble into systems that function unnaturally. Either way, a synthetic goal forces scientists to cross uncharted ground to encounter and solve problems that are not easily encountered through analysis. This drives the emergence of new paradigms in ways that analysis cannot easily do. Synthetic biology has generated diagnostic tools that improve the care of patients with infectious diseases, as well as devices that oscillate, creep and play tic-tac-toe.

Those familiar with academia know that disputes over trademarks can be more intense (and, in a prurient sense, more interesting) than disputes over substance. Synthetic biology has such a dispute in the making.

The title 'synthetic biology' appeared in the literature in 1980, when it was used by Barbara Hobom to describe bacteria that had been genetically engineered using recombinant DNA technology[1]. These bacteria are living systems (therefore biological) that have been altered by human intervention (that is, synthetically). In this respect, synthetic biology was largely synonymous with 'bioengineering'.

In 2000, the term 'synthetic biology' was again introduced by Eric Kool and other speakers at the annual meeting of the American Chemical Society in San Francisco[2]. Here, the term was used to describe the synthesis of unnatural organic molecules that function in living systems. More broadly in this sense, the term has been used with reference to efforts to 'redesign life'[3–5]. This use of the term is an extension of the concept of 'biomimetic chemistry', in which organic synthesis is used to create artificial molecules that recapitulate the behaviour of parts of biology, typically enzymes[6]. Synthetic biology has a broader scope, however, in that it attempts to recreate in unnatural chemical systems the emergent properties of living systems[7], including inheritance, genetics and evolution[3–5,8]. Synthetic biologists seek to assemble components that are not natural (therefore synthetic) to generate chemical systems that support Darwinian evolution (therefore biological). By carrying out the assembly in a synthetic way, these scientists hope to understand non-synthetic biology, that is, 'natural' biology. This motivation is similar in biomimetic chemistry, where synthetic enzyme models are important for understanding natural enzymes.

More recently, an engineering community has given further meaning to the title. This community seeks to extract from living systems interchangeable parts that might be tested, validated as construction units, and reassembled to create devices that might (or might not) have analogues in living systems[9]. The parts come from natural living systems (that is, they are biological); their assembly is, however, unnatural. Therefore, one engineering goal might be to assemble biological components (such as proteins that bind DNA and the DNA sequences that they bind) to create, for example, outputs analogous to those of a computer.

A common ground between the 'synthetic biology' and engineering communities lies in the global strategy by which scientists come to understand their subject matter, make discoveries and overturn paradigms. Synthesis offers opportunities for achieving these goals that observation and analysis do not. The use of synthesis in a way that complements analysis will be a main theme of this review (BOX 1).

Synthetic biology already has many accomplishments to its credit. The effort to generate synthetic genetic systems has yielded diagnostic tools, such as Bayer's branched DNA assay (described in a later section), which annually helps improve the care of some 400,000 patients infected with HIV and hepatitis viruses[10–12]. These and other artificial genetic systems

Department of Chemistry, University of Florida, Gainesville, FL 32611-7200, USA.
Correspondence to S.A.B.
e-mail: benner@chem.ufl.edu
doi:10.1038/nrg1637

Box 1 | **What synthesis can do that observation and analysis cannot**

The scientific method captures the full range of human activities. Life scientists can apply it to observe the foraging strategies of moose[87], or to use X-ray crystallography to determine the molecular structure of the ribosome[88], although the intellectual procedures involved in each case have little in common.

However, one theme is universal: it is easy for human scientists to convince themselves that data contain patterns that they do not, conclude that patterns support models when they need not, and believe that models are truth, which they are not. These observations are not pejorative. They reflect the same processes in the human mind that enable it to be effective and creative. Therefore, although the 'scientific method' that is taught in school emphasizes unfiltered observations, analyzing data with an open mind and conducting value-neutral experiments, the outcome of science does not depend on how well it meets this largely fictional idea, but rather how well scientists manage the values and filters that come naturally with human thought.

Synthesis offers one way to manage these. Synthesis defines an ambitious 'put-a-man-on-the-moon' goal. By doing so, it forces scientists and engineers to cross uncharted terrain in pursuit of the goal. This requires the solution of unscripted problems that are not normally encountered through either observation or analysis. Furthermore, the problems cannot be ignored if they contradict a paradigm. With analysis, if the data contradict the theory, the data are (as often as not) discarded to protect the theory. If one does this when putting an orbiter around Mars, however, the orbiter crashes[89]. For this reason, synthesis drives the evolution of paradigms, however this elusive term is defined[90,91].

This is well illustrated in chemistry, which has long had powerful synthetic tools. For example, the late Robert Woodward credited the discovery of the rules underlying the orbital symmetry of molecules, for which Roald Hoffmann and Kenichi Fukui won the Nobel Prize in Chemistry in 1981, to problems encountered during the synthesis of vitamin B12 (REF. 92). Because vitamin B12 is a large molecule with many sterogenic (chiral) centers, the synthesis of B12 was, for chemistry at that time, the equivalent of a 'man-on-the-moon' goal. Accomplishing this goal led not only to a synthetic route to a complex molecule, but more importantly to a greater understanding of chemical bonding; something that was not accessible by simply observing the structure of B12. It also drove the development of modern analytical tools, such as high-performance liquid chromatography.

Similarly powerful synthetic tools are not available for many other fields. Planetary scientists and stellar physicists cannot, today, synthesize new planets or new stars to test theories and models about these systems. Biology, by contrast, has developed those tools over the past quarter century. During this time, in Barbara Hobom's sense of the term, biologists have been synthesizing parts of living systems to test their ideas. The combination of chemistry, biology and engineering is now at the point where the 'man-on-the-moon' goal is approachable: creating artificial Darwinian systems. One of the metrics of the success of synthetic biology will be how well the effort to assemble existing biological parts into machines, and how well the effort to create artificial systems that reproduce the emergent properties of living systems drives new discoveries and new theories.

now support primitive genetic processes, including replication with the possibility of mutation[13,14], selection[15] and evolution. Synthetic biology has also generated some interesting toys from biomolecular parts, including systems that oscillate[16] and that carry out simple computations[17].

For engineering purposes, parts are most suitable when they contribute independently to the whole. This 'independence property' allows one to predict the behaviour of an assembly. Therefore, it makes sense to structure this review to follow the search for independently interchangeable parts.

This search turns out to be interesting. In molecular science, it is well known that the simplest building units (the atomic parts) do not always contribute independently to the behaviour of a molecular assembly (the whole). In the macroscopic physical world, building units often do, especially if they are designed to do so (as in modular software assembly, for example). Ultimately, synthetic biology succeeds or fails as an engineering discipline depending on where independence approximations become useful in the continuum between the atomic and macroscopic worlds.

As a science, synthetic biology can be evaluated in different ways. By measuring the insights, discoveries and paradigm shifts that are driven by synthetic biology, we ask here whether the synthetic approach has contributed in a way that is not easily possible by analysis alone.

Seeking interchangeable parts: DNA

As described by Watson and Crick 52 years ago, DNA has a modular structure. In a reductionist sense, DNA can be described as two antiparallel strands. Each strand is assembled from four different nucleotide building blocks, which are themselves assembled from sugars, phosphates, and nucleobases. These are, in turn, assembled from carbon, nitrogen, oxygen, phosphorus and hydrogen atoms.

In the Watson–Crick model, nucleotide pairs contribute independently to the stability of a duplex. In reality, this is a good approximation. DNA duplexes can be designed with considerable success by applying just two rules: A pairs with T, and G pairs with C. A second-order model does very well by adding only the effect of adjacent base pairs into the calculation[18]. Although some diversity in nucleic acid structure and function is not captured by such simple rules (for example, that of Z-DNA[19], G QUARTETS[20], and catalytic RNA[21]), most molecular biologists only use this diversity occasionally.

The elegance of the Watson–Crick model has caused most molecular biologists to overlook the chemical peculiarity of such rules. No other molecular system can be described so simply. For example, the behaviour of a protein is generally not a transparent function, linear or otherwise, of the behaviours of its constituent amino acids, even as an approximation. The power of the Watson–Crick rules was nevertheless sufficient to lead to complacency by most of those who learned the double helix structure; molecular recognition in DNA was a 'solved problem'.

The role of the DNA backbone in molecular recognition. This complacency was only dislodged through synthesis of nucleic acids. Starting in the 1980s, some synthetic biologists began to wonder whether DNA and RNA were the only molecular structures that could support genetics on Earth or elsewhere[3,22,23]. Other biologists, seeking technological goals, attempted to replace modules in the DNA structure to create DNA analogues that would, for example, passively enter cells, but could still support the 'A pairs with T, G pairs with C' rule, with the aim of disrupting the performance of intracellular nucleic acids in a sequence-specific 'antisense' way[24].

This antisense idea was simple in cartoon form. The phosphate backbone was thought to be largely responsible for the unsuitability of DNA as a drug: the repeating backbone phosphates prevented nucleic acids

Z-DNA
DNA that exists in a left handed helix rather than the typical right-handed helix of A- and B-form DNA.

G QUARTET
A secondary structure observed in G-rich DNA in which four consecutive guanosine residues bind to each other through interactions in the major groove to form very stable tertiary structures.

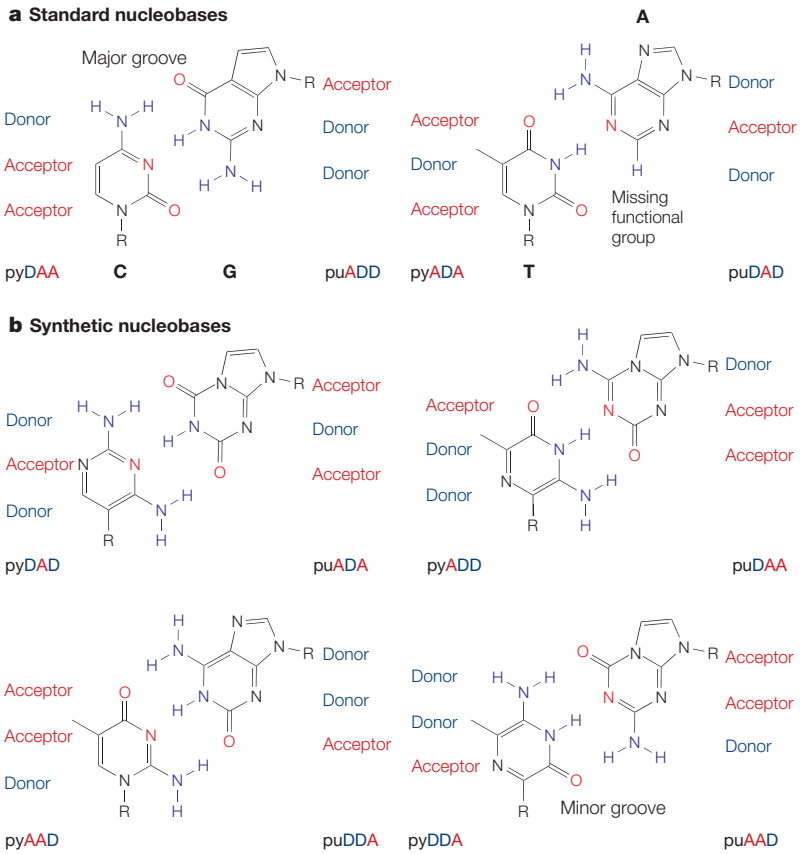


Figure 1 | **Examples of alternative nucleobases.** Parts of the nucleobases of DNA can be used as interchangeable building modules. The blue units are the hydrogen bonding donor (D) collections of atoms. The red units are the hydrogen bonding acceptor (A) collections of atoms. **a** | The four standard nucleobases are shown. **b** | Shuffling the hydrogen bond donor and acceptor modules generates eight additional nucleotides, which constitute a synthetic genetic system. These synthetic bases have been used in an artificial genetic system that can support Darwinian evolution. A, adenine; C, cytosine; G, guanine; Pu, purine; Py, pyrimidine; T, thymine.

from partitioning into lipid phases, an event believed to be essential for molecules to enter cells passively. The phosphate–ribose backbone is also the recognition site for nucleases. This knowledge, and the fact that the Watson–Crick model proposed no particular role for the phosphates in molecular recognition, encouraged the inference that the backbone could be changed without affecting pairing rules.

The effort to synthesize non-ionic backbones changed the established view of nucleic acid structure. Nearly 100 linkers were synthesized to replace the 2′-deoxyribose sugar, starting with the first by the Pitha[25] and Benner[26] laboratories. Nearly all analogues that lacked the REPEATING CHARGE showed worse rule-based molecular recognition. Even with the most successful uncharged analogues (such as the polyamide-linked nucleic-acid analogues (PNA) created by Nielsen and his group[27]) molecules longer than 15 or 20 building units generally failed to support rule-based duplex formation. In other uncharged systems, the breakdown occurs earlier[28].

This discovery was unfortunate for the antisense industry, but it had a marked effect on our understanding of DNA. The repeating charge in the DNA backbone could no longer be viewed as a dispensable inconvenience. The same is true for the ribose backbone of RNA: although several backbones (such as THREOSE DNA or LOCKED NUCLEIC ACIDS) work as well or better than ribose[24,29,30], most of the replacements work less well. The backbone is not simply scaffolding to hold the nucleobases in place; it has an important role in the molecular recognition that is central to genetics.

Evolution of genetic molecules. The above example illustrates how synthesis drives discovery and paradigm change. The failure to obtain non-ionic DNA analogues that retain rule-based pairing led scientists to think about the chemical structures that might be needed to support Darwinian evolution.

In particular, a genetic molecule must be able to suffer change (mutation) without markedly changing its overall physical properties. Again, this feature is infrequent in chemical systems (in proteins, for example). But because charge dominates the physical properties of a molecule, a repeating charge should allow appendages (the nucleobases, in the case of DNA and RNA) to be replaced without changing the dominant behaviour of a genetic system[31]. This has led to the suggestion that a repeating charge might be a universal feature of genetic molecules that work in water[31].

Furthermore, the discovery that ribose was one of the better backbone sugars for supporting molecular recognition[24,32] had implications for the origin of life on Earth. In the mid 1990s, Miller had commented that because of the ease with which ribose decomposes as a sugar on heating[33], ribose could not have supported the first genetic system on Earth. The results from synthesis, which indicated that ribose is especially good for genetics, drove efforts to find prebiotic routes to ribose that would overcome its intrinsic instability[34,35].

Creating synthetic genetic systems. Synthesis focusing on the nucleobases also generated discoveries. The Watson–Crick pairing rules arise from two rules of chemical complementarity. The first, size complementarity, pairs large purines with small pyrimidines. The second, hydrogen-bonding complementarity, pairs hydrogen-bond donors from one nucleobase with hydrogen-bond acceptors from the other.

If nucleobase pairing were indeed so simple, it should be possible to move atoms around within the nucleobases (on paper) to synthesize unnatural nucleobases that would still pair following rules of size and hydrogen bonding complementarity, but differently from the natural nucleobases. Indeed, by shuffling the hydrogen-bond donating and accepting groups, one can easily generate eight additional synthetic nucleobases, forming four additional base pairs (FIG. 1).

In this case, synthesis showed that nucleobase pairing is as simple as the Watson–Crick model implies. A synthetic genetic alphabet with up to 12 independently replicatable nucleobase pairs can be supported by an extended set of Watson–Crick rules[36]. Furthermore, a small amount of protein engineering converts natural polymerases into polymerases that accept components

REPEATING CHARGE
Where the same ionic charge occurs at the same position of each monomeric subunit of a polymer.

THREOSE DNA
DNA molecules based on the sugar threose rather than ribose.

LOCKED NUCLEIC ACIDS
Nucleic acids containing a CH_2 linkage between the 2′OH and 4′ carbon of the ribose sugar.

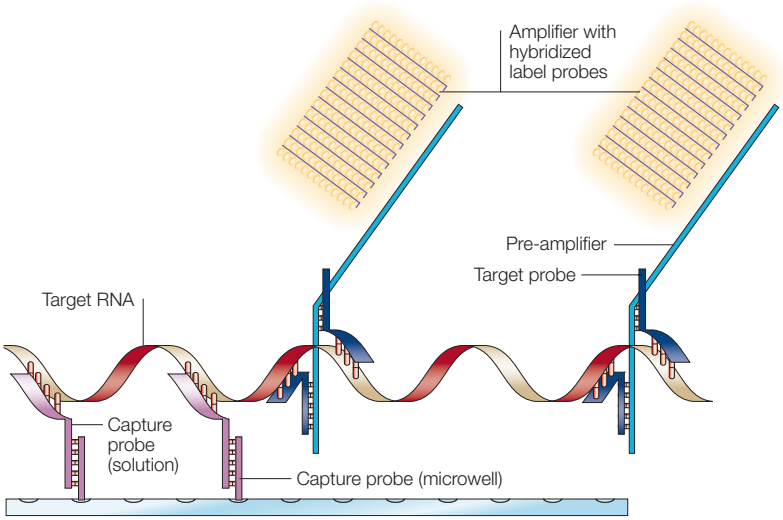


Figure 2 | **Branched DNA assay developed by scientists at Chiron and Bayer Diagnostics.** The target RNA molecule to be detected (the analyte) is attached to the plastic of a microwell (bottom) by the hybridization of the analyte to a series of capture probes. This complex then captures, through hybridization, a target probe, which in turn hybridizes to a pre-amplifier molecule, thereby 'sandwiching' the analyte between the capture probe and the pre-amplifier. The pre-amplifier captures a branched DNA dendrimer (amplifier) that contains several signalling molecules on each branch. As a consequence of the branching, a single analyte assembles a large number of signalling molecules in the microwell. These assays use the expanded genetic alphabet shown in FIG. 1. When standard nucleotides were used to assemble the signalling nanostructure, significant noise was seen, because non-target DNA that was present in the biological sample was captured by the probes in the microwell even in the absence of analyte. Incorporating components of the artificial genetic alphabet in the dendrimer reduced the noise. As a consequence, the assay now helps manage the care of some 400,000 patients annually, detecting as few as eight molecules of the analyte DNA in a sample.

of an expanded genetic alphabet in a polymerase chain reaction[14]. This created, for the first time, a synthetic genetic system that can be repeatedly copied, with the level of mutation needed to support adaptation and evolution.

By searching for synthetic systems that could recreate such emergent properties, synthetic biologists have discovered a great deal. For example, it was proposed that DNA polymerases scan the minor groove of a DNA duplex to look for unshared pairs of electrons as a recognition feature[37]. It was likewise proposed that this scanning of the minor groove was essential for the high fidelity of DNA replication. Efforts to obtain polymerases to support the evolution of the artificial genetic system led to the discovery that minor-groove scanning is not an essential feature of all polymerases.

Today, the effort to make a synthetic chemical system that is capable of Darwinian evolution is an important focus of the National Science Foundation's Chemical Bonding Program. Here, the details of the chemical structures of nucleobases that are essential to support genetics have been determined, with the goal of repairing specific chemical problems that limit the use of specific components of an expanded genetic alphabet. For example, several components of an artificial genetic system suffer from EPIMERIZATION; this has been rectified by adding nitro substituents to the nucleobases[38]. Another component of the artificial system, iso-guanosine, has a minor TAUTOMERIC form that cross bonds with thymidine, creating a significant number of mutations in polymerase chain reactions. This defect was solved by replacing a nitrogen in the structure by a carbon atom[39].

Because it provides rule-based molecular recognition that is orthogonal to the recognition provided by natural DNA, this synthetic genetic system is found today in the clinic. As part of the Bayer VERSANT branched DNA diagnostic assay[40], synthetic biology helps to manage the care of approximately 400,000 patients infected with HIV and hepatitis viruses each year[10,11] (FIG. 2).

Seeking interchangeable parts: proteins

The amino acid as a building module in proteins. The synthetic biology of nucleic acids is successful because the repeating charge in the backbone enables the nucleotide parts to be exchanged independently (although we acknowledge the fact that some RNA structures, such as G-rich sequences, are themselves problematic to engineer). Proteins, unfortunately, do not have a repeating charge; engineering them has therefore been more difficult.

Proposals to engineer proteins, for which the interchangeable unit is the amino acid, is as old as recombinant DNA technology[41,42]. This idea was discussed in an engineering context in 1983 by Kevin Ulmer, then director of exploratory research at Genex[43]. In Ulmer's vision, synthetic biologists would first alter the behaviours of proteins by replacing amino acid BUILDING MODULES in the natural proteins. The replacements would come from the standard set of 20 natural amino acids, and would be chosen using primitive design principles to meet specific goals defined by the properties desired in the synthetic protein. Such primitive principles might, for example, place charged residues at the top and bottom of α helices, or strategically alter amino-acid size complementarity in the active site of an enzyme.

Design based on such primitive rules was expected to frequently fail. Failure, however, would drive the development of better design rules[44]. This would generate a cycle, involving the setting of a goal, the replacement of amino acids to create proteins to meet the goal using the improved design rules, followed by success and failure, refinement of the design rules, and the setting of new goals. This process might go on for decades, and perhaps even generate some of the emergent properties that characterize biological systems. This vision remains largely unrealized. The 20 years of experience since Ulmer presented his vision has shown that the behaviour of a protein is not a simple combination of independent contributions from the constituent amino acids[45].

The failure of the independence approximation was, in large part, expected[46]. First, amino acids in a folded polypeptide sequence strongly interact with others, even amino acids distant in the polypeptide chain[47]. More seriously, even the simplest of molecular interactions are poorly understood. Today, chemical theory still cannot retrodict the freezing point of water[48], the solubility of simple salts in water[49], or the packing of

EPIMERIZATION
Organic molecules can have configurational isomers that have the same number of atoms, and that are bonded in the same way, but that differ in the orientation of the atoms in three dimensions. Epimerization is the spontaneous interconversion of such isomers.

TAUTOMERIC
Isomeric forms of organic molecules in which the same number of atoms are bonded in the same way, except for hydrogen atoms, which are positioned differently.

BUILDING MODULE
A discrete unit used in the construction of a larger structure.

crystals of simple organic molecules[50]. Protein folding is, in one view, an aggregate of these particular processes. A theory that cannot manage the particulars is not expected to manage the whole. Nevertheless, the synthetic effort will be crucial for demonstrating and overcoming these limitations of theory.

Serious efforts are under way to improve the computational tools needed to design and engineer proteins[51–53]. Some have attempted to improve design principles by examining ROTAMERS of amino acids, where different arrangements of side-chain atoms are used as the building modules, rather than the amino acids themselves[54]. Small protein folds have focused the simplification of design issues[55]. These include elegant examples from the laboratories of Imperiali, Allemann and Mayo.

Today, the technology of amino-acid replacement is done using a combination of calculation, design, screening, selection, and luck[56]. Even so, the outcome has been positive[57]; many useful enzymes have emerged by means of amino acid replacement, including polymerases used for DNA sequencing[58], reverse transcriptases that use PCR to amplify synthetic genetic systems[14], and enzymes in commercial laundry detergents[59]. But these are far from a synthetic biology that captures the emergent properties of living systems.

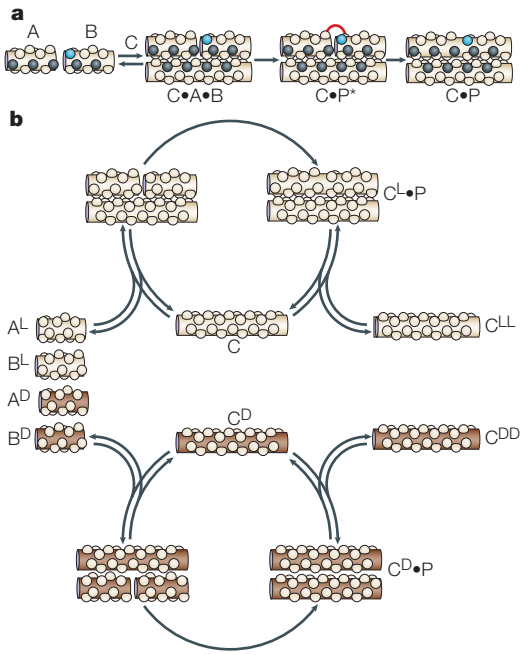


Figure 3 | **A self-templating system built from peptide units. a** | A *de novo* designed peptide ligase. α-helical peptides A and B bind to the electrostatically complementary α-helical peptide C to form the C•A•B ternary complex, which is composed of two coiled coils. Peptide A has a modified amino terminus that reacts with the chemically modified carboxyl terminus of peptide B on formation of the ternary complex. The reaction of peptides A and B is a ligation that forms the product C•P (C•P* represents the chemical reaction between A and B to produce P). **b** | Peptide replicator schematic based on the reaction illustrated in part **a**, showing the reaction of peptide A with peptide B on formation of a ternary complex with peptide C. Peptides A^L, B^L, and C^{LL} are composed of L-amino acids, whereas peptides A^D, B^D, and C^{DD} are composed of D-amino acids. Peptides C^{LL} and C^{DD} are produced autocatalytically in a template-directed fashion through the reaction of precursors A^L with B^L, and A^D with B^D, respectively. Therefore, this replicator is stereochemically selective, only producing products (C^{LL} and C^{DD}) that are isomerically pure. Part **b** modified, with permission, from REF. 65 © American Chemical Society (2001) and from *Nature* REF. 66 © (2001) Macmillan Magazines Ltd.

The protein-folding unit as the building module. Proteins are built from secondary structure units, including the α helix and the β strand[60]. This gave rise to the idea that such secondary structural elements might serve as interchangeable parts to support protein design[61].

Kaiser and his many students have been especially successful in using the amphiphilic helix as the interchangeable building unit. In an amphiphilic secondary structure, hydrophobic and hydrophilic amino-acid side chains are arranged in the sequence so that a hydrophilic side of the unit can face water, while a hydrophobic side can be buried in the protein fold. Such amphiphilic structures are expected to pack spontaneously in water.

Using this strategy, DeGrado *et al.* designed an artificial polypeptide that reproduced some of the folding and biological properties of mellitin, a protein from the sting of a bee, without reproducing its exact sequence[62]. The designed peptide was amphiphilic, and the model proposed that its hydrophobic side buries itself in the hydrophobic membrane of a cell. Amphiphilic helices as units have frequently been exploited since then[63]. Analogous approaches have used the β strand as the architectural module[64].

Several laboratories have worked to create emergent biological properties, including templated replication, by using secondary structural elements as interchangeable building modules. For example, the Ghadiri laboratory designed a peptide by assembling α helical coiled coils to obtain a peptide ligase[65] and a peptide replicator[66] (FIG. 3).

Much of this work falls squarely within the definition of biomimetic chemistry, in that it reproduces isolated behaviours of natural biological systems. This includes using amphiphilic helices as design elements to create a synthetic enzyme that catalyzes the decarboxylation of oxaloacetate[67]. Further work with modular design has improved this artificial enzyme[68]. A wide range of catalytic activities are now being sought in this fashion based on the modular assembly of secondary structure building units. These include artificial enzymes that catalyze ester hydrolysis[69], oxidize phenol using molecular oxygen[70] and catalyze aldol reactions[71].

Despite these successes, helices and strands remain challenging as building modules to support synthetic biology in proteins. Even when amphiphilic elements assemble in water to form a fold, the interior packing is often dynamic (a 'molten globule'). Subsequent design, trial and luck are required to refine the dynamic core to make it rigid. More trivially, many protein design projects have failed simply because proteins have a tendency to precipitate. The backbones of proteins do not carry a repeating charge (like nucleic acids do);

ROTAMERS
(Rotational isomers). Isomeric forms of organic molecules that have the same number of atoms, that are bonded in the same way, but in which the placement of atoms in three dimensional space differs by their rotation around single bonds.

instead, the repeating unit is a dipole (the amide linkage), which is well suited for self-aggregation. Much of protein design, therefore, is an effort to design peptides that remain dissolved in water.

It has been suggested that natural selection might have used secondary structural elements as interchangeable parts in constructing new proteins. Vestiges of such swapping might be detectable in very ancient steps for the construction of proteins[72]. But the recent widespread recruitment of function (proteins that share 50% sequence identity can catalyze very different reactions[73]) seems to have arisen primarily through point mutation, insertion and deletion.

The natural protein as the engineering unit. As we move from the atomic to the macroscopic world, the next step considers folded proteins as interchangeable parts. The biosphere contains proteins that have a wide diversity of physical and catalytic properties. For several decades, metabolic engineers have asked whether these can be reassembled as building modules to create new pathways that capture at least some of the emergent properties of biological systems. This might be done without needing to solve the difficult problem of constructing artificial proteins from scratch.

The simplest idea behind metabolic engineering has been to simply redirect the metabolism of a cell by altering its genetic makeup[74]. This has long been achieved by non-rational means; fermentation strains obtained by classical screening and selection are widely used in industry (in the synthesis of citric acid, for example)[75]. The advent of recombinant DNA technology has led to enhanced efforts to pick and choose enzymes from a range of organisms and then to assemble these enzymes into a single organism to produce products that might not be native to the organism. For example, Fukui *et al.*[76] constructed a strain of the bacterium *Ralstonia eutropha* that was able to produce poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from fructose (FIG. 4a). The strain was engineered by introducing genes for various steps in the pathway from two other microorganisms in addition to *R. eutropha*.

An analogous example comes from the Department of Synthetic Biology at the division of Physical Biosciences at the Lawrence Berkeley National Laboratory[77]. Here, researchers developed a strain of *Escherichia coli* that could synthesize amorphadiene, an isoprenoid precursor for the antimalarial drug artemisinin. The pathway combined the acetoacetyl-CoA thiolase (encoded by the *atoB* gene) from *E. coli*, an isopentenylpyrophosphate isomerase from *Haematococus pluvialis*, and several enzymes from *Saccharomyces cerevisiae* (FIG. 4b). The Gates Foundation is now funding the scaling up of this process as part of its mission to generate inexpensive drugs for the third world.

As Khosla and Keasling noted[78], metabolic engineering generally requires more than simply throwing enzymes together in a cell. Achieving a synthetic goal (here, a strain that produces a particular product) requires the management of complex metabolic and regulatory processes. In pursuit of this goal, one cannot help but learn about metabolism and its emergent behaviours, including the regulation of metabolism and the extent to which enzymes drawn from various sources can be combined independently. So, synthesis drives discovery and learning.

Using genes and genetic elements. One way to search for interchangeable parts is to identify the parts that are used in natural evolution. The archetypal example of modularity in evolution is found in genetic regulatory and signalling pathways. Here, combinations of proteins often function as molecular switches, that is, they are stimulated by upstream events (ligand binding, a chemical reaction, or the movement of components into new locations) and generate an output, which can then serve as an input for another switch. Recruitment and the exchange of individual proteins in such pathways throughout eukaryotic evolution has generated many logical networks that control cellular behaviour[79].

Given this natural precedent, one goal for synthetic biology is to take the proteins themselves as building modules and synthesize artificial regulatory circuits that have preselected inputs and outputs. As before, the motivation is not so much to construct 'toys', but rather to use the challenge of a synthetic goal to discover principles that connect the chemistry of signal transduction to emergent regulatory properties in complex biology.

Several examples show how this might be done. In signal transduction pathways that involve receptor tyrosine kinases, tyrosine-containing motifs are autophosphorylated. These then dock to the Src homology 2 (SH2) and phosphotyrosine-binding (PTB) domains of adaptor proteins, which, in turn, activate signalling pathways that generate specific cellular responses. The signal created by these events depends on the details of the adaptor proteins. For example, the SH2 domain of the growth factor receptor-bound protein 2 (GRB2) adaptor is flanked by two Src homology 3 (SH3) domains. These bind proteins, such as the guanine nucleotide exchange factor son of sevenless homolog 2 (SOS2) and GRB2-associated binding protein 1 (GAB1) docking protein, which then help to activate the Ras and phosphatidylinositol 3'-kinase pathways, respectively. GRB2 therefore couples the phosphorylated motifs to signalling pathways that lead to cell survival, growth and proliferation. Different combinations lead to the opposite outcome. For example, autophosphorylated death receptors bind adaptors, such as fas (TNFRSF6)-associated via death domain (FADD); this is followed by an analogous signalling process that promotes cell death through apoptosis.

In an intriguing study from Pawson's group[80], growth-promoting and death-promoting modules were rearranged. In this work, the phosphotyrosine recognition domains of GRB2 — either SH2 domain or ShcA phosphotyrosine-binding domain —were fused to the death effector domain of FADD. When the construct was transformed into fibroblasts, the synthetic adaptors re-routed the signals that normally lead to growth so that they instead resulted in cell death. Here, the

synthetic goal confirmed the power (and, to some extent, the underlying correctness) of a modular model of regulation by this system.

It is clear that this type of restructuring of existing regulatory components will be used in the future to rewire signalling more generally. Here, the analogy with the construction of electrical circuitry has been used by Hasty *et al.*[81], who recently reviewed some of the spectacular accomplishments in this area (summarized in BOX 2).

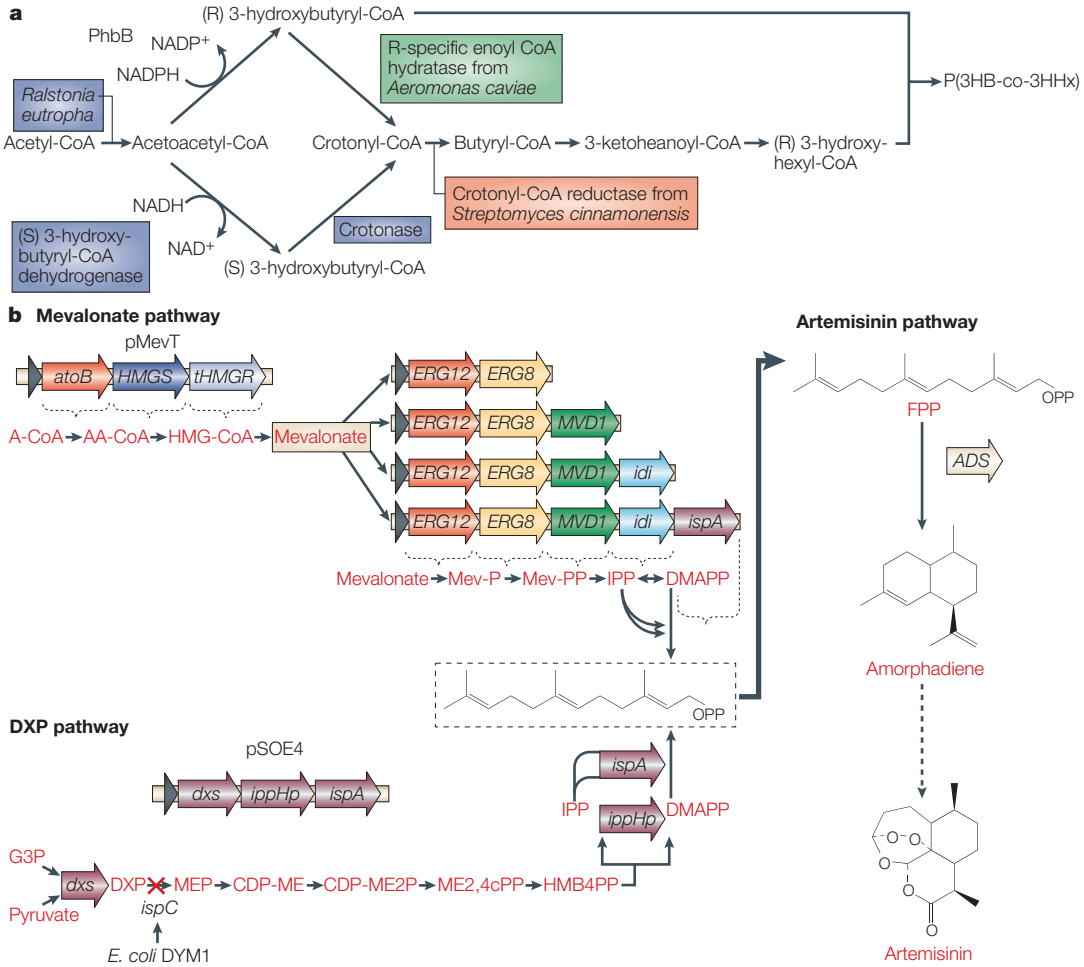
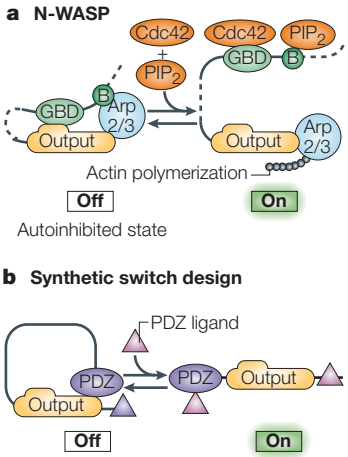


Figure 4 | **Using proteins as interchangeable parts in synthetic biology. a** | The combination of enzymes from three sources in a *Ralstonia eutropha* host generated a strain that produced large amounts of a poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) polymer from fructose. All enzymes from *Ralstonia eutropha* are shown in black, whereas those from *Areomonas caviae* and *Streptomyces cinnamonensis* are shown in green and red, respectively. Modified, with permission, from REF. 76 © American Chemical Society (2002). **b** | The combination of enzymes from three sources in an *Escherichia coli* host generated a strain of the bacterium that produced a precursor for artemisinin, an antimalarial drug. The challenge of this experiment lay in the need to curtail the pathway to recognize and avoid metabolite toxicity, while optimizing the yield of the desired product. The general methodology for the pathway design was to use an engineered mevalonate pathway that is absent in *E. coli*, rather than the DXP (1-deoxy-D-xylulose 5-phosphate) pathway that is native to the organism. The synthetic operons used are depicted, and the engineered pathway metabolites are shown in red. In the engineered mevalonate pathway, the fan of genes from *ERG12* to *ispA* exist on multiple plasmids to tune the pathway for optimization of the product while avoiding metabolite toxicity. As depicted at the bottom of the figure, the *E. coli* strain DYM1, a strain deficient in isoprenoid synthesis, was used, because the DXP pathway was found to limit product yield, probably owing to an unrecognized link between the pathway and physiological control elements in the organism. Enzymes used (isolated from *Saccharomyces cerevisiae* unless otherwise noted): *ADS*, amorphadiene synthase; *atoB*, acetoacetyl-CoA thiolase (*E. coli*); *dxs*, 1-deoxy-D-xylulose 5-phosphate synthase; *ERG12*, mevalonate kinase; *ERG8*, phosphomevalonate kinase; *HMGS*, HMG-CoA synthase; *idi*, IPP isomerase (*E. coli*); *ippHp*, IPP isomerase (*H. pluvialis*); *ispA*, FPP synthase (*E. coli*); *ispC*, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; *MVD1*, mevalonate pyrophosphate decarboxylase; *tHMGR*, truncated HMG-CoA reductase. Pathway intermediates: AA-CoA, acetoacetly-CoA; A-CoA, acetyl-CoA; CDP-Me, 4-diphosphocytidyl-2-C-methyl-D-erythritol; CDP-ME2P, 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate; DMAPP, dimethylallyl pyrophosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; FPP, farnesyl pyrophosphate; G3P, glyceraldehyde 3-phosphate; HMB4PP, 1-hydroxy-2-methyl-2-(E)-butenyl 4-pyrophosphate; HMG-CoA, hydroxymethylglutaryl-CoA; IPP, isopentenyl pyrophosphate; Mav-P, mevalonate 5-phosphate; ME-2,4cPP, 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; Mev-PP, mevalonate pyrophosphate. Adapted, with permission, from *Nature Biotechnology* REF. 77 © (2003) Macmillan Magazines Ltd.

Box 2 | **Engineering regulatory circuits**



The engineering of molecular circuits has recently been reviewed by Hasty *et al.*[81]; this article should be consulted for more details of the cases briefly reviewed below, and for further examples.

In one experiment in yeast, Lim and his co-workers rewired Mapk signalling[93] and an actin regulatory switch known as N-WASP [94,95]. N-WASP carries an output region that (in isolation) stimulates the polymerization of actin by binding the actin-related protein (Arp) 2/3 complex (see part a of the figure). This activity is autoinhibited, however, by means of two modules from N-WASP, a highly basic (B) motif and a guanosine 5'- triphosphatase (GTPase)-binding domain (GBD). These bind to two activating inputs, the phosphoinositide Pip2 and the activated GTPase Cdc42, where the respective binding disrupts autoinhibition. Because the two activating inputs function cooperatively, N-WASP behaves as an 'AND' gate; the output is positive (actin is polymerized) only if both Pip2 and Cdc42 inputs are present.

Lim and co-workers then attempted to reprogram the input control of N-WASP. They tethered an unrelated modular domain–ligand pair (a PDZ domain and its cognate C-terminal peptide ligand) to the end of the N-WASP output domain (see part b of figure). They expected this to create an unnatural autoinhibitory interaction that could be relieved by the competitive binding of an external PDZ ligand. Under basal conditions, this synthetic switch was not activating. The unnatural autoinhibitory interaction was then removed by adding the PDZ ligand. The maximal activity was close to that of the isolated output domain. The result was the same switch, but one that was turned on by a different activator.

The synthetic biologist then attempted to construct AND gates by tethering two unnatural modular domain-ligand pairs to the N-WASP output domain, replacing the two natural modules (B and GBD). A total of 34 constructs were synthesized and examined for their gating behaviour.

As expected for any real chemical system, the output depended on the concentrations of the input molecules, and was not predictable in an engineering sense. Some switches showed little basal repression. Others could not be activated. A few showed antagonistic gating, with one input activating and the other repressing. A few showed 'OR' gating, where one or the other input generated a positive output.

Nevertheless, about half of the synthetic constructs showed positive AND gating, where both inputs were required to generate a positive output. This input-output relationship was analogous to that seen with native N-WASP, but was obtained from a synthetic system, by using unnatural activators. Figure modified, with permission, from REF. 95 **© (2004) Elsevier Science.**

Synthetic biological 'toys' can also be created. For example, an artificial oscillatory network (a 'repressilator') was constructed in *E. coli* by Elowitz and Leibler[16] (FIG. 5) These synthetic biologists placed into an *E. coli* host three transcriptional repressor systems that are not together part of any natural biological system, and coupled these to the synthesis of green fluorescent protein. The fluorescent output of the cell oscillated as a function of time, providing a visual signal of the state of the oscillator in the cell. The oscillation period of hours was longer than the time typical to complete a cycle of cell division. This indicates that the state of the oscillator was transmitted from generation to generation, and became a property of a colony of cells.

The repressilator is, of course, a synthetic equivalent of oscillation systems that are found throughout biology[82]. The goal of producing a synthetic oscillator might therefore lead to a deeper understanding of natural oscillators. For example, the repressilator showed noisy behaviour. This might arise from stochastic fluctuations of its component molecules, which are present in very few copies in individual cells. It might also reflect an interaction between the synthetic parts, or between the synthetic parts and the natural parts of the chemically complex biological host. For this type of network design to lead to an improved understanding of naturally occurring networks, we need to go back to analysis. A detailed analytical study of the synthetic system is needed, just as it was for the synthetic genetic systems or for the synthetic metabolic pathways described above.

Other toy systems have emerged in the past year. Some of these involve biomolecules other than proteins. For example, a molecular automaton that plays tic-tac-toe interactively against a human opponent has been built using deoxyribozymes as the modules[83]. The system is based on a network of 23 molecular-scale LOGIC GATES and one constitutively active deoxyribozyme. These are arrayed in nine wells (3 x 3), which correspond to the game board. To make a move, the molecular automaton analyzes the input oligonucleotide that is associated with a particular move by the human opponent and indicates a move by fluorescence signalling in a response well.

LOGIC GATE
An arrangement of switches used to calculate operations in Boolean algebra; statements are connected into more complicated compound statements by the use of AND, NOT and OR relationships.

Where are the hazards of synthetic biology?

A provocative title such as 'synthetic biology' suggests a potential for hazard. Accordingly, the past year has seen a call for an 'Asilomar for synthetic biology'[84], a reference to a conference in Monterey in 1975 that considered the public hazards of the recombinant DNA technology that supports the synthetic biology discussed above.

Much of what is currently called synthetic biology is congruent with the recombinant DNA technology discussed in Asilomar 30 years ago. This includes bacteria that express heterologous genes, proteins in which amino acids have been replaced, and cells with altered regulatory pathways. Placing a new name on an old technology does not create a new hazard.

Those seeking to create artificial chemical systems to support Darwinian processes are, however, creating something new. We must consider the possibility that these artificial systems might cause damage if, for example, they escape from the laboratory. A general principle in biology is relevant for assessing this potential for hazard. The more different an artificial living system is from natural biological systems, the less likely it is that the artificial system will survive in the natural world. A living organism survives when it has access to the resources that it needs, and is more fit than competing organisms in recovering these. Therefore, a completely synthetic life form that has eight nucleotides (FIG. 1) in its genetic alphabet would find survival very difficult if it were to escape from the laboratory. What would it eat? Where would it get its unnatural nucleosides?

This general principle also applies to less exotic examples of engineered life. The 30 years of experience with genetically altered organisms since Asilomar have indicated that virtually any human-engineered organism is less fit than its natural counterpart in a natural environment. If they survive at all in the environment, they do so either under the nurturing of an attentive human, or by ejecting their engineered features. Therefore, an *E. coli* engineered to play tic-tac-toe might survive in the human intestine, but would probably do so by jettisoning its game-playing skills.

Losing genetic information is much easier than obtaining new information. By contrast with the power of Darwinian processes implied by the Jurassic Park principle ('life finds a way'), Darwinian processes are highly conservative when it comes to creating new functions. When tackling new problems, Darwinian systems take small steps from what they already have; they are not innovators on a large scale. Extinctions are one consequence of this, occurring when the environmental challenges change too fast for Darwinian processes to keep pace.

This is seen, for example, with natural life. The Ebola virus would spread more rapidly if it became airborne and less virulent, as would anthrax bacteria that were transmissible from an infected human host to an uninfected human. The fact that these features have not emerged in these infectious agents indicates the difficulty that Darwinian evolution has in generating such novel properties in existing organisms. In fact, the most hazardous type of bioengineering is the type that is not engineering at all, but reproduces exactly an already existing virulent agent. The synthesis of smallpox virus[85] is perhaps the riskiest recent example of synthetic biology.

The discussion so far presumes an absence of malice. Suppose one actually wanted to do damage? Would one genetically engineer *E. coli* carrying ricin to create a threat? Or place fuel and fertilizer in a rented truck and detonate it outside a Federal Building? We know the answer to this question for one individual; Timothy McVeigh used the latter method to damage the federal building in Oklahoma City a decade ago. We do not know it for all individuals.

At the 2004 International Meeting on Synthetic Biology in Boston, one discussion centered on the 'hacker culture', referring to those who create computer viruses. As is the case whenever resources are diverted from productive to non-productive activity, computer viruses and spam cause human deaths. What if a group of people set out to create an airborne Ebola virus? And what if a synthetic biologist, having set the man-on-the-moon goal of creating such a virus, to learn more about how viruses become airborne, told (through the published literature) the biohacker how to do it?

However, the potential benefits of synthetic biology must be juxtaposed with this hazard. History provides only a partial guide. For example, in 1975, the City of Cambridge banned recombinant DNA research in an effort to manage what it perceived as a danger in this technology. In the same decade, an ill-defined syndrome noted in patients having 'acquired immune deficiency' was emerging around

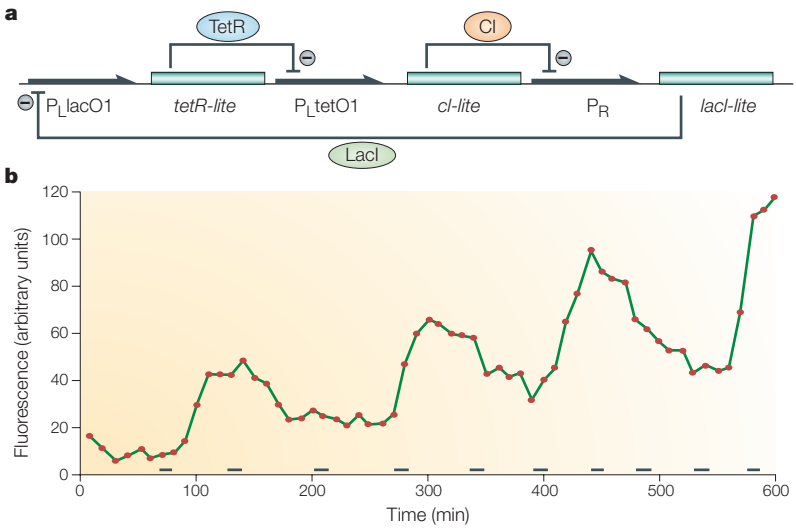


Figure 5 | **The design and application of the repressilator. a** | Schematic showing the regulation pattern that forms the basis of a repressilator. Three gene–promoter pairs are arranged so that the product derived from the expression of the gene following a promoter is a repressor for the next promoter in the cycle. Black connecting lines show that promoter $P_L$lacO1 controls the transcription of the gene *tetR-lite*, the tetracycline repressor protein TetR represses $P_L$tetO1, which is the next promoter in the sequence. $P_L$tetO1 in turn controls the transcription of *cI-lite*, and the protein CI represses the promoter P_R. Finally, P_R controls the expression of *lacI-lite*, and the lactose repressor protein LacI represses $P_L$lacO1, completing the cycle. The suffix '*-lite*' refers to the presence of tags that increase the degradation rate of the proteins. **b** | The luminescence pattern of a reporter plasmid that carries *GFP* under the transcriptional control of the $P_L$tetO1 promoter, when the reporter construct is transferred to an *Escherichia coli* in the presence of the repressilator. As the experimental trace shows, the oscillation of the TetR repressor expressed from the repressilator results in the time dependent oscillation of *GFP* expression. Bars at the bottom of the diagram show the timing of cell division events. The period of the oscillations is longer than the cell division time, and the cycle of oscillations continues in the subsequent generations. Adapted, with permission, from *Nature* REF. 16 © (2000) Macmillan Magazines Ltd.

Box 3 | **Other directions in synthetic biology**

As synthetic methods emerge, the various activities that can lay claim to the title of 'synthetic biology' has expanded beyond what can be covered in a single article. Listed here are just three other themes in the discipline, each with a leading reference.

One example is the 'top-down' approach towards synthetic biology that is being pioneered by Venter and his colleagues[96]. These authors have stripped genes from a living bacterium to determine the minimal set of genes that is required to support a cellular life form. Determining the minimal biological functions required to sustain life will be important for understanding primitive organisms in both a functional and evolutionary context. In addition, synthetic biologists who are practicing a 'bottom-up' approach will find this information crucial for designing synthetic biological systems that include all the necessary components — either in the form of genes or their synthetic equivalents — that are required to sustain life or a desired biological property.

Various types of artificial life that live *in silico*[97] have been suggested as being a form of 'synthetic biology'. This approach involves using simulations to evolve computational analogues of the emergent behaviours of living systems. Many of these artificial life forms compete for resources (such as computer processor cycles) within a computer, and therefore evolve.

Yet another type of synthetic biology uses individual cells as interchangeable parts, and attempts to engineer cell–cell communication. This has recently been reviewed in REF. 98**.**

the planet as an important health crisis. Without the technology that had been banned by the City of Cambridge, it would have been difficult to learn what the human immunodeficiency virus was, let alone have compounds in hand today that manage the infection. Today, as SARS, bird influenza, and other infectious disease emerge from animal populations, recombinant DNA technology is what distinguishes our ability to manage this threat today and what was possible a century ago.

Conclusions

This might be the last time that a short review on synthetic biology will attempt to span work that ranges from organic chemistry to intercellular interactions. Even in 2005, much has been done to do justice to the individual topics discussed here. Several of these are mentioned in BOX 3.

Truly interchangeable parts have, so far, been obtained only by applying synthetic biology to nucleic acids. Here, the independence approximation seems to be adequate because of the repeating charge on the backbone, which remains constant while the information-containing nucleobases are changed. The repeating charge dominates the physical properties of the molecule overall, allowing the pieces to be replaced independently.

Because of the robustness of the interchangeable parts in expanded genetic information systems, we suspect that an artificial chemical system that supports Darwinian evolution — the bridge between non-life and life[86] — will first be obtained with these. Here, the man-on-the-moon goal cannot help but deepen our understanding of the relationship between chemistry and life.

To obtain the same with proteins will require the development of basic chemical theory, including a better understanding of the behaviour of water, the dissolution of salts in water, and the packing of organic molecules. A well-targeted set of experiments, coupled with the appropriate theory, would help this system develop more rapidly. As with the synthetic biology of nucleic acids, chemistry will drive the development of the field.

Folded protein modules behave with sufficient independence to be useful building modules in synthetic biology. Here, we expect a substantial amount of exciting work over the next few years, as synthetic biologists struggle with the chemical reality, determine which modules behave independently, and define synthetic goals accordingly. Here, synthesis will demonstrate its power as a complement to analysis in the development of theory and the modification of paradigms associated with genetic regulation. The setting of ambitious synthetic goals cannot help but deepen our understanding of the intimate relationship between chemistry and life at the regulatory level, and better understand the emergent properties of complex biological systems.

1. Hobom, B. Surgery of genes. At the doorstep of synthetic biology. *Medizin. Klinik* **75**, 14–21 (1980).
2. Rawls, R. 'Synthetic Biology' makes its debut. *Chem. Eng. News* 49–53 (24 April 2000).
3. Benner, S. A. Redesigning life. Organic chemistry and the evolution of protein. *Chimia* **41**, 142–148 (1987).
4. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387–390 (2001).
5. Benner, S. A. Act natural. *Nature* **421**, 118 (2003).
   This paper outlines the importance of pursuing the science of synthetic biology.
6. Breslow, R. Centenary Lecture. Biomimetic chemistry. *Chem. Soc. Rev.* **1**, 553–580 (1972).
7. Salt, G. W. Use of the term 'emergent properties'. Comment. *Am. Nat.* **113**, 145–148 (1979).
8. Benner, S. A. (ed.) *Redesigning the Molecules of Life* 115–175 (Springer, Heidelberg, 1988).
9. Gibbs, W. W. Synthetic life. *Sci. Am.* **290**, 74–81 (2004).
10. Elbeik, T. *et al*. Simultaneous runs of the Bayer VERSANT HIV-1 version 3.0 and HCV bDNA version 3.0 quantitative assays on the system 340 platform provide reliable quantitation and improved work flow. *J. Clin. Microbiol.* **42**, 3120–3127 (2004).
11. Elbeik, T. *et al.* Multicenter evaluation of the performance characteristics of the Bayer VERSANT HCV RNA 3. 0 assay (bDNA). *J. Clin. Microbiol.* **42**, 563–569 (2004).
12. Benner, S. A. Redesigning genetics. *Science* **306**, 625–626 (2004).
13. Latham, J. A., Johnson, R. & Toole, J. J. The application of a modified nucleotide in aptamer selection. Novel thrombin aptamers containing 5-(1-pentynyl)-2'-deoxyuridine. *Nucleic Acids Res.* **22**, 2817–2822 (1994).
14. Sismour, A. M. *et al.* PCR amplification of DNA containing non-standard base pairs by variants of reverse transcriptase from Human Immunodeficiency Virus-1. *Nucleic Acids. Res.* **32**, 728–735 (2004).
    This paper describes the first example of an artificial genetic system that can be replicated by enzymes, and highlights some important concepts in the design of genetic systems.
15. Battersby, T. R. *et al.* Quantitative analysis of receptors for adenosine nucleotides obtained via *in vitro* selection from a library incorporating a cationic nucleotide analog. *J. Am. Chem. Soc.* **121**, 9781–9789 (1999).
16. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
17. Ruben, A. J. & Landweber, L. F. The past, present and future of molecular computing. *Nature Rev. Mol. Cell Biol.* **1**, 69–72 (2000).
18. SantaLucia, J. & Hicks, D. The thermodynamics of DNA structural motifs. *Ann. Rev. Biophys. Biomol. Struct.* **33**, 415–440 (2004).
19. Rich, A. & Zhang, S. G. Z-DNA: the long road to biological function. *Nature Rev. Genet.* **4**, 566–572 (2003).
20. Sen, D. & Gilbert, W. Novel DNA superstructures formed by telomere-like oligomers. *Biochemistry* **31**, 65–70 (1992).
21. Kazakov, S. & Altman, S. A. Trinucleotide can promote metal ion-dependent specific cleavage of RNA. *Proc. Natl Acad. Sci. USA* **89**, 7939–7943 (1992).
22. Benner, S. A. Understanding nucleic acids using synthetic chemistry. *Acc. Chem. Res.* **37**, 784–797 (2004).
23. Ball, P. Synthetic biology: starting from scratch. *Nature* **431**, 624–626 (2004).
24. Freier, S. M. & Altmann, K. H. The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.* **25**, 4429–4443 (1997).
25. Pitha, J., Pitha P. M. & Ts'o, P. O. Poly(1-vinyluracil): the preparation and interactions with adenosine derivatives. *Biochim. Biophys. Acta* **204**, 39–48 (1970).
26. Schneider, K. C. & Benner, S. A. Oligonucleotides containing flexible nucleoside analogs. *J. Am. Chem. Soc.* **112**, 453–455 (1990).
27. Nielsen, P. E. PNA technology. *Mol. Biotechnol.* **26,** 233–248 (2004).
28. Benner, S. A. & Hutter, D. Phosphates, DNA, and the search for nonterrean life: A second generation model for genetic molecules. *Bioorg. Chem.* **30**, 62–80 (2002).

29. Wilds, C. J., Wawrzak, Z., Krishnamurthy, R., Eschenmoser, A. & Egli, M. Crystal structure of a B-form DNA duplex containing (L)-α-threofuranosyl (3'→2') nucleosides: A four-carbon sugar is easily accommodated into the backbone of DNA. *J. Am. Chem. Soc.* **124**, 13716–13721 (2002).
30. Wengel, J. Synthesis of 3′-C- and 4′-C-branched oligodeoxynucleotides and the development of locked nucleic acid (LNA). *Acc. Chem. Res.* **32**, 301–310 (1999).
31. Hutter, D., Blaettler, M. O. & Benner, S. A. From phosphate to bis(methylene) sulfone: Non-ionic backbone linkers in DNA. *Helv. Chim. Acta*. **85**, 2777–2806 (2002).
32. Declercq, R., Van Aerschot, A., Read, R. J., Herdewijn, P. & Van Meervelt, L. Crystal structure of double helical hexitol nucleic acids. *J. Am. Chem. Soc.* **124**, 928–933 (2002).
33. Larralde, R., Robertson, M. P. & Miller, S. L. Rates of decomposition of ribose and other sugars. Implications for chemical evolution. *Proc. Natl Acad. Sci. USA* **92**, 8158–8160 (1995).
34. Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* **284**, 2118–2124 (1999).
35. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate minerals stabilize ribose. *Science* **303**, 196 (2004).
36. Geyer, C. R., Battersby, T. R. & Benner, S. A. Nucleobase pairing in expanded Watson-Crick-like genetic information systems. *Structure* **11**, 1485–1498 (2003).
37. Meyer, A. S., Blandino, M. & Spratt, T. E. *Escherichia coli* DNA polymerase I (Klenow fragment) uses a hydrogen-bonding fork from Arg(668) to the primer terminus and incoming deoxynucleotide triphosphate to catalyze DNA replication. *J. Biol. Chem.* **279**, 33043–33046 (2004).
38. Hutter, D. & Benner, S. A. Expanding the genetic alphabet. Non-epimerizing nucleoside with the pyDDA hydrogen bonding pattern. *J. Org. Chem.* **68,** 9839–9842 (2003).
39. Martinot, T. A. & Benner, S. A. Artificial genetic systems: exploiting the 'aromaticity' formalism to improve the tautomeric ratio for isoguanosine derivatives. *J. Org. Chem*. **69**, 3972–3975 (2004).
40. Collins, M. L. *et al.* A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res*. **25**, 2979–2984 (1997).
41. Schmeissner, U. & Miller, J. H. Genetic studies of Lac repressor. *Experientia* **32**, 811–811 (1976).
42. Smith, M. Synthetic DNA and biology (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* **33**, 1214–1221 (1994).
43. Ulmer, K. M. Protein engineering. *Science* **219**, 666–671 (1983).
44. Arnold, F. H. *et al.* Engineering nonaqueous solvent-compatible enzymes. *ACS Symp. Ser.* **516**, 109–113 (1993).
45. Mildvan, A. S. Inverse thinking about double mutants of enzymes. *Biochemistry* **43**, 14517–14520 (2004).
46. Knowles, J. R. Tinkering with enzymes: what are we learning? *Science* **236**, 1252–1258 (1987).
47. Fukami-Kobayashi, K., Schreiber, D. R. & Benner, S. A. Detecting compensatory covariation signals in protein evolution using reconstructed ancestral sequences. *J. Mol. Biol.* **319**, 729–743 (2002).
48. Ding, K., Chandler, D., Smithline, S. J. & Haymet, A. D. Density-functional theory for the freezing of water. *Phys. Rev. Lett*. **59**, 1698–1701 (1987).
49. Zepik, H. & Benner, S. A. Catalysts, anticatalysts, and receptors for unactivated phosphate diesters in water. *J. Org. Chem.* **64**, 8080–8083 (1999).
50. Dunitz, J. D. & Bernstein, J. Disappearing polymorphs. *Acc. Chem. Res.* **28**, 193–200 (1995).
51. Mayo, K. H. Recent advances in the design and construction of synthetic peptides: for the love of basics or just for the technology of it. *Trends Biotechnol.* **18**, 212–217 (2000).
52. Venkatraman, J., Shankaramma, S. C. & Balaram, P. Design of folded peptides. *Chem. Rev.* **101**, 3131–3152 (2001).
53. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. Computational design of a biologically active enzyme. *Science* **304** 1967–1971 (2004).
This paper illustrates the progress made in the *de novo* design of biological molecules.
54. Shah, P. S., Hom, G. K. & Mayo, S. L. Preprocessing of rotamers for protein design calculations. *J. Comput. Chem.* **25**, 1797–1800 (2004).
55. Ali, M. H., Peisach, E., Allen, K. N. & Imperiali, B. X-ray structure analysis of a designed olilgomeric miniprotein reveals a discrete quaternary architecture. *Proc. Natl Acad. Sci. USA* **101**, 12183–12188 (2004).
56. Arnold, F. H. Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253–257 (2001).
57. First, E. A. & Fersht, A. R. Analysis of the role of the KMSKS loop in the catalytic mechanism of the tyrosyl-tRNA synthetase using multimutant cycles. *Biochemistry* **34**, 5030–5043 (1995).
58. Tabor, S. & Richardson, C. C. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl Acad. Sci. USA* **92**, 6339–6343 (1995).
59. Igarashi, K., Hagihara, H. & Ito, S. Protein engineering of detergent α-amylases. *Trends Glycosci. Glycotechnol.* **15**, 101–114 (2003).
60. Pauling, L. & Corey, R. B. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc. Natl Acad. Sci. USA,* **37**, 235–240 (1951).
61. Velicelebi, G., Patthi, S. & Kaiser, E. T. Design and biological activity of analogs of growth hormone releasing factor with potential amphiphilic helical carboxyl termini. *Proc. Natl Acad. Sci. USA* **83**, 5397–5399 (1986).
62. Degrado, W. F., Kezdy, F. J. & Kaiser, E. T. Design, synthesis and characterization of a cyto-toxic peptide with melittin-like activity. *J. Am. Chem. Soc.* **103**, 679–681 (1981).
63. Ghirlanda, G., Lear, J. D., Ogihara, N. L., Eisenberg, D. & DeGrado, W. F. A hierarchic approach to the design of hexameric helical barrels. *J. Mol. Bio.* **319**, 243–253 (2002).
64. Quinn, T. P., Tweedy, N. B., Williams, R. W., Richardson, J. S. & Richardson, D. C. Betadoublet: *de novo* design, synthesis, and characterization of a β-sandwich protein. *Proc. Natl Acad. Sci. USA* **91**, 8747–8751 (1994).
65. Kennan, A. J., Haridas, V., Severin, K., Lee, D. H. & Ghadiri, M. R. A *de novo* designed peptide ligase: A mechanistic investigation. *J. Am. Chem. Soc.* **123**, 1797–1803 (2001).
66. Saghatelian, A., Yokobayashi, Y., Soltani, K. & Ghadiri, M. R. A chiroselective peptide replicator. *Nature* **409**, 797–801 (2001).
The chiroselective peptide replicator developed in this article illustrates the modular design aspects of synthetic biology and the ability to replicate information in a peptide based system.
67. Johnsson, K., Allemann, R. K., Widmer, H. & Benner, S. A. Synthesis, structure and activity of artificial, rationally designed catalytic polypeptides. *Nature* **365**, 530–532 (1993).
68. Taylor, S. E., Rutherford, T. J. & Allemann, R. K. Design of a folded, conformationally stable oxaloacetate decarboxylase. *J. Chem. Soc. Perkin Trans. 2*, 751–755 (2002).
69. Nicoll, A. J. & Allemann, R. K. Nucleophilic and general acid catalysis at physiological pH by a designed miniature esterase. *Org. Biomol. Chem.* **2**, 2175–2180 (2004).
70. Kaplan, J. & DeGrado, W. F. *De novo* design of catalytic proteins. *Proc. Natl Acad. Sci. USA* **101**, 11566–11570 (2004).
71. Tanaka, F. & Barbas, C. F. A modular assembly strategy for improving the substrate specificity of small catalytic peptides. *J. Am. Chem. Soc.* **124**, 3510–3511 (2002).
72. Roy, S. W. & Gilbert, W. Complex early genes. *Proc. Natl Acad. Sci. USA* **102**, 1986–1991 (2005).
73. Benner, S. A. & Ellington, A. D. Interpreting the behavior of enzymes. Purpose or pedigree? *CRC Crit. Rev. Biochem.* **23**, 369–426 (1988).
74. Jacobsen, J. R. & Khosla, C. New directions in metabolic engineering. *Curr. Opin. Chem. Biol*. **2**, 133–137 (1998).
This paper presents some of the recent work in metabolic engineering that is typically used in synthetic biology.
75. Das, A. Strain selection in citric acid fermentation. A review. *Curr. Sci.* **41**, 593–596 (1972).
76. Fukui, T., Abe, H. & Doi, Y. Engineering of *Ralstonia eutropha* for production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from fructose and solid-state properties of the copolymer. *Biomacromolecules* **3**, 618–624 (2002).
77. 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).
The development of a bacterial strain that can synthesize pharmaceuticals illustrates the power of synthetic biology as a tool for technology.
78. Khosla, C. & Keasling, J. D. Metabolic engineering for drug discovery and development. *Nature Rev. Drug Discov.* **2**, 1019–1025 (2003).
This paper highlights the complexity that is required to engineer a metabolic pathway successfully.
79. Lim, W. A. The modular logic of signaling proteins: Building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol*. **12**, 61–68 (2002).
This paper lays the foundation for designing complex logic networks.
80. Howard, P. L., Chia, M. C., Del Rizzo, S., Liu, F. F. & Pawson, T. Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. *Proc. Natl Acad. Sci. USA* **100**, 11267–11272 (2003).
81. Hasty, J., McMillen, D. & Collins, J. J.. Engineered gene circuits. *Nature* **420**, 224–230 (2004).
82. Eisen, H., Brachet, P., Pereira da Silva L. & Jacob, F. Regulation of repressor expression in λ. *Proc. Natl Acad. Sci. USA* **66**, 855–862 (1970).
83. Stojanovic, M. N. & Stefanovic, D. A deoxyribozyme-based molecular automaton. *Nature Biotech.* **21,** 1069–1074 (2003).
This paper illustrates not only the ability to develop interactive logic networks, but also some of the new kinds of technology that synthetic biology is producing.
84. Ferber, D. Synthetic biology. Time for a synthetic biology Asilomar? *Science* **303**, 159 (2004).
85. Cello, J., Paul, A. V. & Wimmer, E. J. Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. *Science* **297**, 1016–1018 (2002).
The chemical synthesis of poliovirus cDNA described in this article illustrates the ability to synthesize life; the paper also discusses the dangers of synthetic biology and organism engineering.
86. Rasmussen, S., Chen L., Nilsson, M. & Abe, S. Bridging nonliving and living matter. *Artif. Life* **9**, 269–316 (2003).
87. Belovsky, G. E. *et al.* Ten suggestions to strengthen the science of ecology. *Bioscience* **54**, 345–351 (2004).
88. Steitz, T. A. & Moore, P. B. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.* **28**, 411–418 (2003).
89. Bearden, D. A. A complexity-based risk assessment of low-cost planetary missions: when is a mission too fast and too cheap? *Acta Astronaut.* **52**, 371–379 (2003).
90. Kuhn, T. S. *The Structure of Scientific Revolutions* (Univ. Chicago Press, 1962).
91. Masterman, M. in *Criticism and the Growth of Knowledge* (eds Lakatos, I. & Musgrave, A.) 59–89 (Cambridge Univ. Press, 1970).
92. Woodward, R. B. Recent advances in the chemistry of natural products. *Pure Appl. Chem*. **17**, 519–547 (1968).
93. Park, S. H., Zarrinpar, A. & Lim, W. A. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299**, 1061–1064 (2003).
94. Prehoda, K. E., Scott, J. A., Mullins, R. D. & Lim, W. A. Integration of multiple signals through cooperative regulation of the N-WASP-Arp2/3 complex. *Science* **290**, 801–806 (2000).
95. Dueber, J. E., Yeh, B. J., Bhattacharyya, R. P. & Lim. W. A. Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. *Curr. Opin. Struct. Biol.* **14**, 690–699 (2004).
96. Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal mycoplasma genome. *Science* **286**, 2165–2169 (1999).
97. Adami, C. *Introduction to Artificial Life* (Springer, New York, 1998).
98. Yokobayashi, Y., Collins, C. H., Leadbetter, J. R., Arnold, F. H. & Weiss, R. Evolutionary design of genetic circuits and cell-cell communications. *Adv. Complex Systems* **6**, 37–45 (2003).

Competing interests statement
The authors declare no competing financial interests.

Online links

DATABASES
The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
Fadd | *Gab1* | *Grb2* | *Sos2*

FURTHER INFORMATION
Benner laboratory: http://www.chem.ufl.edu/groups/benner
Kool laboratory: http://www.stanford.edu/group/kool
National Science Foundation's Program in Chemical Bonding: http://www.nsf.gov/news/news_summ.jsp?cntn_id=100424&org=NSF&from=news
The Bill and Melinda Gates Foundation: http://www.gatesfoundation.org
Access to this interactive links box is free online.