nature biotechnology

Unbottling the genes

The ability to plug and play synthetic genes into minimized genomes promises to transform biological engineering.

As the first decade of the 'century of biology' draws to a close, our attempts to engineer novel traits into living cells remain remarkably primitive. Most genetic modification is constrained to tinkering with a handful of genes in a handful of laboratory-adapted natural organisms. We can get cells to express new genes, as long as it's only one or a few genes, and as long as we don't want to control expression too precisely. Try to go much beyond that and our efforts flounder. If biological engineering were aviation, it would be at the birdman stage: some observation and some understanding, but largely naive mimicry. For the field to really take flight, it needs the machinery of synthetic biology.

There are many views on what synthetic biology is, and what it should be, but one aspect that differentiates this field from previous genetic/metabolic engineering is that everything proceeds from the computer: the necessary starting materials are digital code and four bottles of chemicals (A, G, T and C). A DNA synthesizer converts these precursors into oligonucleotide (oligo) sequences *in vitro*. The oligos are assembled into larger pieces (genes, gene circuits and even artificial chromosomes) and, after error checking, plied into use.

The simplest application of this approach, gene synthesis, already thrives commercially. More and more laboratories are requesting genes from oligo companies rather than using laborious recombinant DNA cloning techniques and PCR. Some gene synthesis providers recode sequences to improve protein properties, such as solubility, toxicity, efficiency of translation and ease of purification.

But the approach offers many more possibilities than simply cutting costs and time; it can create artificial products beyond the reach of nature's own evolved art. For example, nuclease-resistant sequences can be fabricated from nucleoside triphosphate analogs (e.g., for aptamers, RNAi or antisense for RNA, and gene therapies for DNA). Similarly, sequences can be redesigned with expanded codon usage that can be read by systems (e.g., *Escherichia coli* or yeast) engineered with corresponding orthogonal suppressor tRNA and aminoacyl-tRNA synthetase pairs to generate proteins containing unnatural amino acids with useful properties.

In more complex manifestations, synthetic biology aims to design gene circuits analogous to electrical components and circuits. Over the past decade, such efforts have led to all manner of synthetic gene switches, oscillators, digital logic evaluators, filters, sensors and communicators (see p. 1139). Even though these circuits remain relatively crude, falling short of the most complex gene manipulation achieved using traditional metabolic engineering (e.g., Jay Keasling's *tour de force* engineering of the pathway for precursors of the malaria drug artemisinin), they ultimately promise exquisite control of outputs. Rather than constructing the biological equivalent of a radio that receives one station and outputs at a set volume, engineers will be able to spin the dial and choose a listening comfort zone. Responsive elements will allow biological devices to adapt to their environment, a useful characteristic in, say, synthetic islet cells or gene therapy systems. Cell systems could even be designed to learn.

For synthetic biology to truly transcend the current limitations, however, it will be necessary to move away from laboratory-adapted versions of 'natural' organisms—organisms that bring with them the genetic and metabolic 'baggage' of millions of years of evolution. This baggage has been essential in the face of environmental instability and insult, reproductive fitness, invaders and predators, and the other rigors of survival. But it is redundant in the context of the constant, cosseted, aseptic and substrate-rich environments of man-made production systems. Natural organisms' metabolic flexibility almost invariably limits their metabolic flux in culture and thus industrial productivity. The ideal industrial bug is not a utility player but an extreme specialist honed to metabolic perfection.

This is where another concept in the field—the minimal genome capable of supporting a self-replicating organism—becomes important. Theoretically, an organism with a genome stripped of superfluous functions that drain away carbon, nitrogen or energy could serve as a 'shell' or 'chassis' into which interchangeable cassettes of genes encoding traits of interest could be placed.

Chassis organisms can be generated by serially deleting parts of an existing organism's genome or identifying nonessential genes and then synthesizing/assembling a minimized artificial chromosome from scratch (see p. 1121). The final step of 'rebooting' the synthetic minimal genome has not yet been attained, but we may not be so far from the goal of creating a 'chassis' organism—the blank canvas onto which the bold and efficient metabolic brushstrokes of synthetic biology can be made.

By applying the principles of engineering to living systems and allowing us to move away from mimicry and optimization of natural cells, synthetic biology thus opens up the possibility of design in completely artificial systems. One day, these systems may provide insights into existing living organisms—life as it already is—enhancing our understanding of basic biology and disease. But that goal still seems some way off.

In contrast, it is not too hard to imagine a future where, with relatively little effort, we can create alternative life forms—minimal-genome chassis organisms with interchangeable standardized gene circuits—that will enable genetic engineers to rapidly move from one industrial project to another. The technology is disruptive, with the potential to transform biological engineering, which until now has been limited to tinkering with natural organisms, and relies on a good deal of serendipity for success.

At the turn of the last century, the Wright brothers achieved manned flight not by mimicking natural systems, but by applying the principles of engineering and aerodynamics. Similarly, synthetic biology allows us to dispense with biological mimicry and design life forms uniquely tailored to our needs. In doing so, it will offer not only fundamental insights into questions of life and vitality but also the type of exquisite precision and efficiency in creating complex traits that genetic engineers could previously only dream of.

What's in a name?

Defining an emerging field can be challenging. *Nature Biotechnology* asked 20 experts for their views on the term 'synthetic biology'.

Similar to other new and trendy fields, synthetic biology has been defined so loosely that it can seem like all things to all people. Traditional genetic or metabolic engineering has been rebranded as synthetic biology, often to take advantage of the hype cycle that fuels investor interest. Below, 20 experts give their own definitions. The diversity of responses indicates that consensus as to the meaning of synthetic biology still lies some way off.

Adam Arkin, professor, Department of Bioengineering, University of California, Berkeley, California.

Synthetic biology aims to make the engineering of new function in biology faster, cost effective, scalable, predictable, transparent and safe. It focuses on improvement of standard genetic engineering technology; development of standards for genetic assembly and rapid characterization; creation of families of genetic 'parts' that behave reliably in designated hosts and have no undesigned interactions; and generation of safe, robust host cells. That is, it aims to remove the burden of synthesis and endless rounds of optimization of functional performance and thereby facilitate the design of increasingly complex systems. Although chemical production is the most powerful current application, synthetic biology seeks to address a much broader class of problems, including programmable materials, therapeutic organisms and systems that support agricultural and environmental services. Many of these systems will be engineered for operation beyond the bioreactor, requiring sophisticated sensing, computing and actuating systems to perform effectively and safely in complex environments.

Frances Arnold, professor, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California.

Synthetic biologists construct new biological entities—molecules, pathways, regulatory networks, organisms and ecosystems—by programming them, or reprogramming them, at the level of the DNA code. The new name 'synthetic biology' reflects an explosion in our ability to genetically engineer increasingly complex systems and the desire of scientists and engineers from fields outside molecular biology and genetics to participate in the fun, contributing to the technology and its applications.

David Berry, partner, Flagship Ventures, Cambridge, Massachusetts.

The term synthetic biology should really be synthetic biotechnology. The reason for that is, effectively, it is using tools of modern biology, including DNA sequencing, DNA synthesis, cell analytics, etc., to design biological tools to accomplish tasks. The goal is to leverage exponential information-generation with the precision of biology to create these tools. The use of them can be broad, including sense-response proteins or cells, engineered biocatalysts, or cells that undergo conversions. On this last point, the difference between metabolic engineering and synthetic biotechnology is that only with the latter can you design cells that accomplish a task that is independent from what the cell normally does—that is, causing a heterotrophic organism to be autotrophic, not improving a yeast's ability to make ethanol.

Joachim Boldt, assistant professor, and **Oliver Müller**, junior research group leader, Department of Medical Ethics and the History of Medicine, Freiburg University, Germany.

Synthetic chemistry has shown the way: from systematic analysis of chemical processes to synthesis of novel products. Synthetic biology does the same, but in the realm of the living. It leads us from analyzing complex molecular processes inside the cell to generating novel cellular functions and novel single-cell organisms. As such, synthetic biology comprises our full-blown ability to technically manipulate genetic, metabolic and signaling processes inside and in between cells. It is turning us into creators of the most basic parts of living nature. Synthetic biology opens up the possibility to augment nature with neo-microbes by an effort of engineering, thus aiming at controlling the uncontrollable. Philosophically speaking, the project of synthetic biology crystallizes in one single question: can we or should we, undoubtedly being part of nature, understand ourselves as co-creators of the evolution?

George Church, professor, Department of Genetics, Harvard Medical School, Boston, Massachusetts.

Genetic engineering focuses on individual genes (typically cloning and overexpression). The logical extension of that to system-wide change is genome engineering. Intermediate between these is metabolic engineering, which involves optimizing several genes at once. Synthetic biology is 'meta' to all of these in establishing standards for modules, intentionally interoperable in their assembly and functioning. Hierarchical properties permit computer-aided design at different levels of abstraction, from the sub-molecular level to supra-ecosystem levels.

Andrew D. Ellington, professor, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas.

These words [synthetic biology] don't have much meaning. The definition of a new field is either based on a discovery or redefinition, and—because I can't point to a single great discovery in this field—synthetic biology is really more about a redefinition of biotechnology. It encompasses the rather old notion that you can engineer living systems, but updates that notion with the universal realization that the ability to synthesize lots of DNA and do mathematical modeling is a very powerful combination. But I'd say synthetic biology's key utility is to excite engineers, undergraduates and funding agencies. Its key disadvantage is to create hysteria in the defense community.

Drew Endy, assistant professor, Department of Bioengineering, Stanford University, Stanford, California.

We human beings belong to the clan of the opposable thumbs; we are very good at discovering and making new things by building. Synthetic biology, by exploring how to remake or assemble the molecules of life, provides a complementary scientific approach for learning how life works. Synthetic biology also celebrates getting much better at constructing new living things by recognizing that a good biological engineer will not just deliver on any one biotechnology application but will also contribute to the development of tools, so that all who might follow will find a safer and easier path.

Martin Fussenegger, professor, Swiss Federal Institute of Technology, Zurich, Switzerland.

Since its inception some 40 years ago, molecular biology has largely remained a descriptive discipline using a rather childish strategy to unravel the inventory of biological parts that are essential for life on this planet: disassemble to understand. Life becomes a lot more thrilling when we are assembling parts to make functional systems. With the post-genomic era having provided encyclopedic information on gene-function correlations, and systems biology now delivering comprehensive details on the dynamics of biochemical reaction networks, molecular biology has come of age and life scientists are now adult: ready to reassemble these cataloged items in a systematic and rational manner to create and engineer functional biological designer devices and systems with novel and useful functions. A new type of constructive systems biology—synthetic biology—is born.

E. Richard Gold, professor, Faculty of Law, McGill University, Montreal, Quebec, Canada.

Synthetic biology comprises the research necessary to develop a living organism that can be described without reference to an existing organism. Drawing on my patent law background, what seems critical to me is that any resulting organism can be described in words without having to refer, directly or indirectly, to any living organism. The end result will therefore be something completely new rather than a modification or change to an existing organism. By defining the field in terms of a result, it leaves the specific disciplines that are included open. For example, I would imagine that the fields include not only biology, but computer science and even social sciences to the extent that these help overcome important roadblocks to researchers' ability to do their work.

Jim Greenwood, president and CEO, Biotechnology Industry Organization, Washington, DC.

Synthetic biology is an interdisciplinary approach that applies engineering principles to biology. It builds on both improvements in the speed and cost of chemical synthesis of naturally occurring DNA and growing knowledge of genomics to enable researchers to design and synthesize modified microorganisms, such as bacteria, that can produce useful products in the pharmaceutical industry, personal care, specialty chemicals and biofuels. Whereas systems biology studies complex natural biological systems using modeling and simulation comparison to experiment, synthetic biology studies how to build artificial biological systems and synthesize industrial products. The focus is often on taking parts of natural biological systems, characterizing and simplifying them, and using them as components of an engineered biological system.

Sang Yup Lee, distinguished professor and LG Chem Chair professor, Korea Advanced Institute of Science and Technology, Daejeon, Korea.

Originally, synthetic biology sought to redesign and rebuild biological parts and systems without specific biotechnological objectives, whereas metabolic engineering aimed at purposeful modification of metabolic and other cellular networks to achieve desired goals, such as overproduction of bioproducts. Recently, it has become more difficult to distinguish the two disciplines as each is employing the other's approaches. Metabolic engineering is adopting synthetic biology's strategies of gene synthesis, very fine control of gene expression, etc., while synthetic biology is taking metabolic engineering's objective-driven strategies of engineering circuits and consideration of whole-cell metabolism. And both are moving towards integration with systems biology. We do not need to argue about what synthetic biology is and what is different about it as we want to honor people's diverse thinking.

Wendell Lim, professor, Department of Cellular & Molecular Pharmacology, University of California, San Francisco, California.

Synthetic biology is the application of engineering principles towards the construction of novel biological systems. At its heart, all synthetic biology shares a constructivist philosophy of trying to figure out how simpler parts can be combined to build systems with much more sophisticated behaviors, whether the goal is to build something useful or to increase our basic knowledge. Although most synthetic biology uses modules of biological origin as its toolkit, I am agnostic about whether this must be part of the definition of synthetic biology. For example, if someone figured out how to use abiotic components to build a material with the very 'biological' behavior of self-repair, I would consider that synthetic biology. In many cases we learn more about the 'rules of living systems' if we mimic them with a range of completely different parts.

Jeremy Minshull, CEO, DNA2.0, Menlo Park, California.

Synthetic biology began in earnest when phosphoramidite chemistry first allowed us to design and synthesize DNA sequences *de novo*. Now, we can make genes easily and are on the brink of synthesizing functional genomes, but we are only starting to learn how to design the sequences we really want. Scientific progress is incremental, but people holding purse strings, public or private, are most excited by paradigm shifts and the prospect of quick payoffs. Synthetic biology, then, is a useful term to attract funding for the ongoing (~30-year-old) biological revolution, powered by advances in molecular biology techniques coupled with increases in computing power. It means whatever the listener wishes to hear.

Thomas H. Murray, president, The Hastings Center, Garrison, New York.

Multiple streams of scientific inquiry and engineering practice, some decades old, converge

under the marketing banner 'synthetic biology'. The ways we think and feel about biology are evolving along with the technologies used to manipulate it. Synthetic biology embodies: a faith that biological systems can be brought to heel, and made predictable and controllable; a stance toward the intricacy of biological organisms aptly described by Tom Knight [MIT] as an "alternative to understanding complexity is to get rid of it"; a confidence that biological entities can be hacked apart and reassembled to satisfy human curiosity and to serve important, legitimate human purposes; a *hope* that error and malevolence can be deterred, contained or outmaneuvered through the vigilance of governments and, especially, the collective efforts of well-intentioned scientists, engineers and garage biologists. Will what we might dub the 'Legoization' of biology fully justify the faith, stance, confidence and hope invested in it? The answer to this question will help to shape the future of humankind and the world we inhabit.

George Poste, chief scientist, Complex Adaptive Systems Initiative, Arizona State University, Phoenix, Arizona.

The boundary between synthetic biology and systems biology should reside in a single criterion: has the engineered process, product or organism been fabricated from natural materials (systems biology) or from components not adopted in natural evolution (synthetic biology)? Non-natural substrates include novel nucleotides and amino acids, proteins with unique tertiary structures, hybrid organic–inorganic molecular assemblies, biomimetic nano- and meso-scale materials and devices, and genetic sequences that did not arise through natural evolution. The construction of complex multigenic assemblies from known genetic sequences to synthesize biofuels or natural biomolecules that cannot be readily produced by chemical synthesis represents advanced genetic engineering and not synthetic biology. If such manipulations were classified as synthetic biology, the entire history of biotechnology and heterologous gene transfer would warrant redefinition as synthetic biology. The prospect of novel organisms created by synthetic biology has provoked scrutiny about potential health and environmental risks and dual-use abuse. Inaccurate definitions of the field, driven by efforts to attract publicity or funding, run the risk of attracting regulatory oversight to advanced biotechnology activities that do not pose the complex public policy issues raised by synthetic biology.

Kristala L.J. Prather, assistant professor, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts.

If you ask five people to define synthetic biology, you will get six answers. I'd say it is the (re-)design, construction and analysis of biological systems or sub-systems. It is an effort to apply engineering principles in the context of biology and includes a focus on the development of well-characterized parts from which higher-order devices and systems can be reliably and robustly assembled.

Hana El-Samad, assistant professor, Department of Biochemistry and Biophysics, California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, California.

Synthetic and systems biology are the ultimate synergetic partners for ushering in an era of rapid and provably systematic biological discovery. There are two ingredients necessary to unravel a biological system: the ability to generate perturbations that are maximally informative, and the ability to accurately measure the impact of such perturbations and organize the information they yield into a framework that can be easily queried and methodically analyzed. Synthetic biology could provide the first ingredient by generating genetically encoded 'perturbation' generators that are well-designed and characterized, while being tunable and portable. Conversely, systems biology should provide the technological innovations necessary to measure quantitatively the dynamical outcomes of these perturbations in any system of interest. It should also provide the computational innovations that are appropriate for a brand of system identification tailored to biological questions, in addition to analysis tools that can transition between different biological scales. This last feature is absolutely necessary—whereas the immediate goal might be investigation of a given biological mechanism, the ultimate goal should be the identification of the overarching organizational principles of cells and organisms. Systems and synthetic biology share this common vested interest, and a close-knit collaboration will reap many benefits for both fields.

Christina Smolke, assistant professor, Department of Bioengineering, Stanford University, Stanford, California.

Synthetic biology involves the development and application of engineering principles to make the design and construction of complex synthetic biological systems easier and more reliable. It is the focus on the development of new engineering principles and formalism for the substrate of biology that sets it apart from the more mature fields upon which it builds, such as genetic engineering. Synthetic biology represents an approach to biological design and genetic programming that can be used in a variety of different application areas in biological engineering, such as metabolic engineering or genetic/cellular therapies. However, one can conduct projects in these application areas (that is, metabolic engineering) without them falling into the category of synthetic biology, depending on the approach and tools implemented in the design, construction and characterization processes.

Ron Weiss, associate professor, Department of Biological Engineering and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Synthetic biology is the engineering discipline for building novel and sophisticated living systems. In this discipline, we view cells as 'programmable matter', and strive to design and control complex intracellular and extracellular activities that allow us to achieve precisely defined engineering or scientific goals. To be successful, we will need to incorporate engineering principles and methodologies that have worked well in other established fields (e.g., modularity, system fabrication using libraries of well-characterized and interchangeable parts, rapid prototyping, predictive models and robust designs). But at the same time, we must also be cognizant of the interesting and challenging features of the biological substrate that make it different from all other existing engineering disciplines (e.g., self-replication, self-repair, mutation and evolution, high degree of noise, incomplete information and the importance of cellular context).

Making green

Biofuels top the list of products for many biotech companies using advanced biological engineering. Cormac Sheridan examines the diverse commercial paths being taken to reach this goal.

The $600 million research alliance on algal biofuels, which Synthetic Genomics entered into with ExxonMobil in July, represents the most persuasive evidence yet that the challenge of developing economically viable and environmentally sustainable biofuels offers a perfect proving ground for synthetic biology. The scale of the problem and the technological means of addressing it are well aligned. Effecting even a partial transition from fossil fuels to biomass-derived alternatives could substantially reduce greenhouse gas emissions in transportation—and in chemicals production, which is heavily dependent on petrochemicals—and it would spur the creation of a new industrial biotech sector.

Even though the commercial opportunity is vast, the constraints within which biofuel producers must operate are extremely tight. "The price of fuel is less than the price of bottled water, so you have to be incredibly efficient," says Jack Newman, cofounder and senior vice president of research at Amyris Biotechnologies, of Emeryville, California, one of the first movers in the field. Genetically reprogramming microorganisms on a scale that was unprecedented even five years ago holds out the prospect of achieving some of those efficiencies.

Of course, sophisticated biological manipulation alone will not deliver next-generation biofuels—the technology is only one part of a much wider effort. But many companies claim it will be difficult to contemplate any meaningful progress in the area without it.

The grand challenge

Particularly in the US, geopolitical as well as environmental considerations are behind the drive to reduce fossil fuel consumption. The US Congress, under the 2007 National Renewable Fuel Standard program, has mandated production of 36 billion gallons of all biofuels by 2022 (ref. 1). Corn-based ethanol production in the US grew rapidly during the present decade, from 1.6 billion gallons in 2000 to 9 billion gallons in 2008, according to the Washington, DC-based lobby group the Renewable Fuels Association[2]. The environmental sustainability of corn-based ethanol remains controversial, however, both from a land-use perspective and from the critical standpoint of reducing overall greenhouse gas emissions[3].

Cellulosic ethanol, produced from plant waste and other nonfood sources, offers a more attractive greenhouse gas emissions profile than its corn-derived counterpart, and the first commercial quantities of cellulosic ethanol are due to come online in 2010. One recent report indicates that some producers have either scaled back or delayed their production plans, however, because of the current economic downturn[4].

Although large volumes of cellulosic ethanol will be used in the coming decade and beyond, its long-term technical feasibility has been questioned because of its low energy density, its miscibility with water and its corrosive properties[5]. It lacks the 'drop-in-ability' of next-generation alternatives, such as butanol and alkanes, which can be handled in the existing fuel storage and distribution infrastructure[6].

Getting bacteria, yeast and algae to produce these kinds of molecules—in very large volumes, at very low cost—is the grand challenge that a slew of young biotech firms has taken on (**Table 1**). Metabolic engineering, or extensively reprogramming the physiology of the producing organisms, is the goal. Synthetic biology offers companies the means to achieve it.

"Five years ago, forget it. It would have been a dream," says Pat Gruber, CEO of Englewood, Colorado-based Gevo. His company is working on the production in yeast of isobutanol and butanol, building blocks that can be easily modified chemically to yield fuel molecules. Gruber is, however, wary of the language used to denote the techniques that Gevo and other firms are employing. "When people say 'synthetic biology,' that to me has almost no meaning. It's just genetic engineering," he says. "People are enamored with the name and the concept. It's like a [form of] branding."

Others also see synthetic biology as a continuation of classical genetic engineering, albeit with a greater degree of intensity. "When I think of synthetic biology, I think of all the things I used to do as a biologist, slowly, methodically, and not always with a lot of success," says Stephen del Cardayre, vice president of research and development at LS9 of South San Francisco, California. "Today, most of these methods can be automated such that robots can carry out thousands of experiments effectively, efficiently and with excellent success. This allows us to test many, many more hypotheses in parallel—quickly and cost effectively."

Improving on nature

But the present effort goes beyond a massive industrial scale-up of genetic engineering. The ready availability of custom-designed synthetic DNA molecules marks a significant point of departure from earlier eras of genetic manipulation. Amyris Biotechnologies, for example, is engineering yeast cells to make a class of branched hydrocarbons called isoprenoids using biosynthetic genes that are ultimately derived from plants, but it rarely works with actual plant DNA. "You might generate a lead that way, but by the time that piece of DNA sees a microbe it's been through DNA synthesis," says Newman. And more often than not, the immediate 'source' of a gene is a database rather than a living plant.

Using synthetic DNA is not simply a matter of speed or convenience. The redundancy of the genetic code enables scientists to improve the expression of foreign genes in different hosts by designing synthetic DNA molecules that encode the correct protein but have a nucleotide sequence that takes account of the host organism's codon usage bias. As Newman observes, DNA contains a lot more information than just an amino acid sequence. "There's a whole rule set about translatability we haven't even formulated." Design procedures that are 'translation aware' are beginning to emerge, however.

Verdezyne (formerly Coda Genomics), of Carlsbad, California, has developed a computational biology platform for designing self-assembling synthetic genes that not only addresses codon usage bias but also takes into account a lesser-known phenomenon, codon pair bias[7]. The over-representation of certain codon pairs in an RNA sequence appears to act as a brake on translation. Altering them can boost gene expression. Including more of them in a sequence can slow it down. "It's not just the codon abundance but the codon context that affects translation," says Verdezyne CSO Stephen Picataggio.

Jonathan Wolfson, cofounder of the company Solazyme, shows green algae that his company is engineering to create renewable biofuels.

Table 1 Selected advanced biofuel companies

Company	Process	Producing Organism
Ethanol		
Algenol Biofuels, Naples, Florida	Photosynthesis in contained bioreactor	Cyanobacteria
BioGasol, Ballerup, Denmark	Glucose & xylose fermentation	Yeast, anaerobic, thermophilic bacteria
Codexis[1], Redwood City, California	High performance enzymes via gene shuffling	Not available
Coskata, Warrenville, Illinois	High temperature biomass gasification and fermentation using carbon monoxide and hydrogen	Not disclosed
Dupont Danisco Cellulosic Ethanol, Itaca, New York	Combined cellulosic conversion and fermentation	*Zymomonas mobilis*
Gevo, Englewood, Colorado	Production of higher alcohols via amino acid biosynthetic pathway	Yeast
Green Biologics, Abingdon, UK	Modified classical acetone, butanol ethanol (ABE) fermentation	Clostridium species, Geobacillus species
Joule Biotechnologies, Cambridge, Massachusetts	Helioculture modified photosynthetic process in closed bioreactor	Modified photosynthetic organisms
Lanza Tech, Auckland, New Zealand	Fermentation process using carbon monoxide and hydrogen from syngas and fluegas	Not disclosed
LS9, S San Francisco	Biodiesel fermentation via fatty acid metabolism	*E. coli*
Mascoma, Lebanon, New Hampshire	Combined lignocelulose conversion with fermentation	Yeast, *Clostridium thermocellum*
Qteros, Marlborough, Massachusetts	One-step bacterial lignocellulose conversion and fermentation	*Clostridim phytofermentans*
TMO Renewables, Guildford, UK	Combined cellulosic conversion and fermentation	Geobacillus TM242
Verdezyne, Carlsbad, California	Improved yeast fermentation based on microbial glycolytic pathway and xylose isomerase	Yeast
Verenium, Cambridge, Mass	Combined cellulosic conversion and fermentation	Ethanologenic bacteria
Zeachem, Lakewood, Colorado	Hybrid biochemical and thermochemical process involving acetic acid fermentation	Naturally occurring acetate producing bacteria
Diesel		
Amyris Biotechnologies, Emeryville, California	Isoprenoid biosynthesis via mevalonate pathway	Yeast
Aurora Biofuels, Alameda, California	Photosynthesis in open pond system	Naturally occurring algae
OPX Biotechnologies, Boulder, Colorado	Undisclosed	*E. coli*
Algal oils		
Sapphire Energy, San Diego, California	Photosynthesis	Photosynthetic algae
Solazyme, S. San Francisco	Photosynthesis	Photosynthetic algae
Solix Biofuels, Fort Collins, Colorado	Photosynthesis in close system photobioreactor	Photosynthetic algae
Synthetic Genomics, La Jolla, California	Combined phtosynthetic production and secretion	Photosynthetic algae

[1]Codexis is a technology provider to cellulosic ethanol producer logen, of Ottawa, Canada.

His company has successfully applied this approach to the expression of the bacterial enzyme xylose isomerase in yeast, enabling it to ferment the five-carbon sugar, which is a significant cellulose constituent. "People have been trying to do this for 35 years already. I tried it as a postdoc," Picataggio says. "The problem has been the enzyme misfolds in the yeast cytoplasm."

DNA 2.0, of Menlo Park, California, is also working on the role of codon usage in synthetic gene expression, and it recently published data suggesting that codons used to encode a subset of amino acids were strongly correlated with expression of two genes in *Escherichia coli*[8].

Nature, however, remains the starting point for any gene (or protein) engineering effort, even though the final molecule may undergo many alterations. "We don't think that we understand enzymes nearly well enough to sit down and design the perfect one," says Lori Giver, vice president of systems biology at Codexis, of Redwood City, California, which uses DNA shuffling to evolve high-performance enzymes.

The fork in the road

Although all the firms working on advanced biofuels share similar goals, the specifics of their technologies and their business strategies differ. Given the field's early stage of development—and the commercial rewards at stake—some firms are reluctant at this stage to divulge fully their technology strategies. But each company has a fundamental decision to make: whether to engineer a biofuel-producing capability into a well-known, robust industrial organism or to engineer industrial fitness and other necessary attributes into an organism that is a natural producer of the molecule of interest.

Extensive genetic tools and components are available for engineering yeast and *E. coli*. These also have long histories as fermentation organisms. The same cannot be said for photosynthetic algae or bacteria, although such organisms are attractive because they require no raw materials, other than sunlight, carbon dioxide and water, to make hydrocarbons. Joule Biotechnologies, of Cambridge, Massachusetts, is one firm taking the photosynthetic route to biofuels. Photosynthesis offers the overarching advantage of bypassing the need to break down the lignin and cellulose present in plant biomass. "The notion of really doing it directly is where you get quite a lot of your efficiency," says company cofounder David Berry, of Cambridge-based Flagship Ventures (which has also funded LS9 and Lebanon, New Hampshire-based Mascoma). Joule has developed proprietary methods for manipulating its target organisms. "The way I like to describe it is recapitulating about 30 years of *E. coli* engineering in about eighteen months," says Berry. The company has not identified the organisms it is working on, however. "They are naturally photosynthetic organisms we've engineered in a number of new ways," Berry says.

Synthetic Genomics, of La Jolla, California, and ExxonMobil, of Irving, Texas, are also keeping specific details of their research program

under wraps for now, although the scale of the alliance suggests that it will have a very broad scope. "We're going to be testing probably every approach that's out there," says Synthetic Genomics CEO and cofounder J. Craig Venter. Synthetic Genomics scientists have engineered algal strains that can transport lipids out of the cell, which offers the possibility of setting up a continuous biomanufacturing process rather than an intermittent cycle of growing and harvesting. "The conventional wisdom did not have algae that secreted hydrocarbons in a pure form into the media—so I think our breakthrough on that front changes the entire equation," Venter said last July. Although its approach has yet to be scaled up, the early indications are promising. "The existing starting yields that we have are on the order of ten times more efficient than acreage for production of corn and we hope to substantially build on that through this program," Venter went on.

Around a month after the ExxonMobil announcement, Venter and colleagues at the J. Craig Venter Institute (JCVI), of Rockville, Maryland, reported that they had succeeded in transferring an entire bacterial chromosome—that of *Mycoplasma mycoides* ssp. *mycoides*—into a yeast cell and then subjecting it to modification using the yeast genetic system[9]. If applicable to other organisms, such as biofuel producers, it could provide a general method for engineering the genomes of organisms that are otherwise difficult to manipulate. "It will be a key enabling technology for the whole field," Venter says. "It creates the ability to do rapid changes that were not remotely possible before."

The road map to success

The traits that biofuel companies want to engineer into producing organisms extend far beyond those directly associated with the production of a specific fuel molecule. The specific changes that individual companies make—and the methods they use to make them—obviously vary. Michael Lynch, founder and CSO of OPX Biotechnologies, of Boulder, Colorado, likens the problem to driving from New York to Los Angeles without a road map. Companies that build up a large-scale, automated genetic engineering platform are, he says, building a faster car or plane. "We focus a lot on how you map out the space between here and there." OPX takes a "population-based approach" to mapping the links between genotypes and phenotypes. "We perturb the network and measure the results in a massively parallel way," Lynch says.

All strain improvement efforts have to operate within the boundaries imposed by microbial physiology. "Each time you do some permutation it impacts the energy balance," Gevo's Gruber says. If a biosynthetic pathway is to be massively overexpressed, it requires a steady supply of the appropriate cofactors, for example. Pathways that consume energy may need to be switched off. "You have to prune the metabolic pathways you don't want," says Jim Flatt, president of Mascoma. Mascoma is developing what it terms consolidated bioprocessing technology, which aims to combine the hydrolysis of lignocellulose and the fermentation of the resulting sugar molecules to ethanol in a single process. It involves engineering cells—bacterial or yeast—to express and secrete a suite of cellulase enzymes to break down cellulose and hemicellulose into their constituent sugars and to convert those sugars to ethanol. Hydrolysis of cellulose and hemicellulose requires about 20 distinct enzymes—that are normally provided by commercial suppliers such as Novozymes, of Bagsvaerd, Denmark, or the Genencor unit of Copenhagen, Denmark-based Danisco. "It's basically a chamber orchestra of activity you need here—it's not just breaking down starch, which is an easy process," says Flatt. The hydrolysis process can also result in the production of byproducts, including acids, ketones and aldehydes, that can inhibit the growth of cells as well as the secreted enzymes. Mascoma is working on directed evolution strategies to adapt cells so that they not only tolerate but thrive on these molecules, Flatt says.

LS9's del Cardayre emphasizes the advantages of harnessing biosynthetic pathways involved in the production of primary metabolites, as a large fraction of the cell's normal metabolic flux will be directed toward that pathway in any case. His firm is focused on the production of alkanes from fatty acid intermediates in *E. coli*. "Just before the fatty acid intermediates are incorporated into the cell membrane, we steal them from that pathway and divert them into a fuel biosynthetic pathway we've engineered into the cell," he says. Gevo is also focused on a primary metabolic pathway. It is building on technology in-licensed from James Liao of the University of California, Los Angeles, who demonstrated how to divert intermediates from *E. coli*'s amino acid biosynthetic pathway toward the production of branched chain alcohols[10].

Industrial scale biofuels

Scaling up any of these processes represents a daunting challenge. What works on a lab bench will not necessarily work in an industrial fermentation vessel. Gevo, which is pursuing a retrofit strategy based on migrating corn ethanol plants over to butanol production, recently opened its first demonstration-scale facility. Amyris opened a demonstration-scale production facility in Brazil midyear and aims to produce commercial quantities of its renewable diesel from sugarcane feedstock in 2011. LS9 says it will reach demonstration-scale production of its UltraClean Diesel fuel in 2010. The availability of cost-effective enzymes for breaking down cellulose will be critical for the success of the field, del Cardayre says. "We are rooting for and working with those companies developing technologies for converting biomass cost-effectively into sugar."

Although it is evident that synthetic biology is central to the development of advanced biofuels, it is not yet clear whether a fully synthetic genome will ever be deployed in a live production environment. "A fully synthetic microorganism may not have the robustness which is needed for large-scale industrial bioprocesses," Picataggio says. Venter says the first generation of producing organisms he is working on could potentially be a naturally occurring strain. Further generations will be what he calls "synthetic genomic constructs," with partially synthetic genomes.

In any case, economics, not technology, will be the ultimate arbiter of success. The list of successful industrial fermentations "is not long," Gruber notes. The most noteworthy include the production of commodities such as ethanol, lysine, citric acid, lactic acid, polyhydroxyalkanoate, 1,3-propanediol and erythritol. To be competitive, he says, a fermentation needs to produce around 100 grams per liter of end product; its productivity should exceed two grams per liter per hour; and its anaerobic yield should stand at ~95% of the theoretical yield. "If you meet those requirements, you will be in economically efficient space," he says.

The synthetic biology pioneers will have to clear these hurdles just as their industrial microbiology predecessors did previously. Although buzzwords and dazzling science have hyped expectations, considerable challenges lie ahead before new genome engineering applications can be turned into green gold. As Gruber puts it: "I've never seen anything commercially successful in our industrial biotech space that started with, 'Gee whiz, this is a cool invention.'"

Cormac Sheridan, Dublin

1. US Environmental Protection Agency. EPA proposes new regulations for the renewable fuel standard program for 2010 and beyond. (EPA-420-F-09-023) <http://www.epa.gov/otaq/renewablefuels/420f09023.htm#3> (2009).
2. Renewable Fuels Association. Ethanol industry statistics. <http://www.ethanolrfa.org/industry/statistics/#A> (2005–2009).
3. Searchinger, T. *et al. Science* **319**, 1238–1240 (2008).
4. Kwok, R. *Nature* **461**, 582–583 (2009).
5. Keasling, J.D. & Chou, H. *Nat. Biotechnol.* **26**, 298–299 (2008).
6. Lee, S.K. *et al. Curr. Opin. Biotechnol.* **19**, 556–563 (2008).
7. Larsen, L.S.Z. *et al. Int. J. Bioinform. Res. Appl.* **4**, 324–336 (2008).
8. Welch, M. *et al. PLoS One* **4**, e7002 (2009).
9. Lartigue, C. *et al. Science* **325**, 1693–1696 (2009).
10. Atsumi, S. *et al. Nature* **451**, 86–89 (2008).

Biotech in the basement

Do it yourself 'biohackers' want to break down institutional barriers and bring science to the people. But good intentions are up against the hard realities of doing science. Joe Alper reports, with additional reporting by Laura DeFrancesco.

In September, a New York City-based group of biohackers held a DNA extraction party as part of the city's ConfluxCity day—an annual street fair for the investigation of urban life or psychogeography, in the parlance of festival organizers. Several people went home with a test tube of their own DNA, a modest success, which is emblematic of how things are going for the garage biotech movement. Whereas eight cities now have active do-it-yourself bio (DIYbio) groups and over a thousand individuals have joined various Listserves for biohackers, the number of actual participants in the movement, if you can call it that, is quite small. One might even argue that as some of the most vocal proponents—some might call them publicity hounds—are not scientists but artists and social commentators, garage biotech might actually be something more akin to performance art or guerrilla theater.

But this is not to say that the movement lacks seriousness of purpose. Inspired by the International Genetically Engineered Machine (iGEM) project at Massachusetts Institute of Technology (MIT) in Cambridge, amateur scientists have attempted some fairly sophisticated projects in home laboratories, and a handful of small companies and projects have emerged to serve the population.

The next HP?

Biohackers like to point out the parallels between the biotech and information technology (IT) industries. Both have a common birthplace in the San Francisco Bay area and derived their early funding from many of the same venture capitalists. But whereas the entrepreneur-driven IT industry was born out of the work of hobbyists working in their garages—think Hewlett and Packard, Jobs and Wozniak—the biotech industry was and still is a product of well-funded, professional researchers working in big, well-equipped labs.

Not that there aren't biotech hobbyists trying to follow in the footsteps of their IT counterparts. In fact, put 'do-it-yourself biotech' into Google and you'll be rewarded with dozens of newspaper and magazines article heralding the rise of a garage biotech movement. There's a Garage Biotech blog (http://blog.openwetware.org/freegenes/category/garage-biotech/), bulletin board (http://www.biopunk.org/) and online community (http://www.diybio.org/) for biotech do-it-yourselfers.

Several garage (or, in one case, bedroom) biotech stories have attracted the media spotlight. Using a PCR machine that was purchased on eBay for a mere $59, Kay Aull, a former researcher at the now defunct Cambridge, Mass.-based Codon Devices, genotyped herself to see if she carried the gene for hemochromatosis, which afflicts her father. Computer programmer Meredith Patterson, after creating glow-in-the-dark yogurt in her San Francisco apartment, is working on a biosensor for melamine, the toxic contaminant of the Chinese infant formula that sickened 300,000 infants in 2008.

Redefining hacking. The term hacking in popular usage refers to computer criminals, but to DIYbio enthusiasts, hacking is taking things apart and putting them back together in a good way.

DIYbio hosts several ongoing projects on its website, among them an openware hardware management package, called SKDB, and SmartLab, which aims to build inexpensive hardware for lab settings. SmartLab is working on such things as data logging instruments and video streaming for recording lab activities and capturing "did-I-just-pipette-that-into-the-wrong-tube?" moments. And since being excluded from iGEM's annual competition, which now requires university sponsorship in order to compete, DIYbio may start its own, according to DIYbio Boston founder MacKenzie Cowell (**Box 1**).

Making hardware

At least two equipment suppliers have popped up in recent years as an outgrowth of the biohacker movement. Pearl Biotech, a small instrument supply company with origins in a Sacramento, California garage, markets to the do-it-yourself crowd with a gel box for under $200, complete with power supply and transilluminator. The company, which grew out of the 'Open Gel Box' project, has also set up a site for tracking the cost of DNA synthesis, called the "1 cent per base pair" project (http://www.1centbp.com/) that provides information on DNA synthesis companies providing the most competitive rates.

Ginko Bioworks, located in a former ship container near Boston's harbor, is a bona fide start-up that is developing a set of tools for the uninitiated. It is now marketing a cloning kit through New England Biolabs in Ipswich, Mass., which is tailor-made to facilitate interaction among end users; the kit has a set of linkers taken from the Biobrick registry of standardized parts, which makes kit-generated components compatible with any other. Started by a group of five MIT students and faculty, the company bootstrapped itself into existence and created a fully functional laboratory with some seed money and used equipment. They now support themselves through (an undisclosed amount of) kit revenue and a contract with the Scottish life science consultancy ITI Life Sciences in Dundee, worth £1.25 million ($2.09 million).

Reshma Setty, one of Ginko's founders, feels accomplished, having gotten off the ground without going to venture capitalists for money. "There's a presumption that you need $5–$10 million of VC funding. We deliberately started differently," she says. She thinks that there's room for a tier of companies in this area that can provide services and tools with much smaller amounts of funding than traditional VC-funded models.

Thinking big

At least one project supported by the do-it-yourself bio movement is moving out of the confines of the home laboratory into the global biosphere. The BioWeatherMap Project hopes to send people around the globe out into their environments to sample the resident microorganisms, which will then be used to create a global microbiome. With low-cost cotton swabs and costs of analysis so far donated by like-minded scientists, the project got off the ground in June at the X,Y,Z

Box 1 Enthusiasts versus professionals?

How garage biotech differs from biotech practiced in institutions is subject to debate among university scientists. To Stanford University's Drew Endy, this is an "artificial divide." The notion that separating science from the university is a means of promoting exploration may overstate the reality as well as raise expectations of what can be accomplished. He sees a large amount of excitement over the technology. "The technology is cooler than PCs," says Endy. And without venues, either within universities or without, to service that excitement, a vacuum is created, he says, which inevitably gets filled.

However, missing in the world of amateur biotechers is oversight, which Endy sees as less of a problem than some of his colleagues. "All the three-letter agencies are abreast of what's going on and are trying to figure out the right strategy to prevent an accident or deliberate [harmful] act. We live in a world where there are attacks," Endy says. But Jim Thomas, of the nonprofit action group on Erosion, Technology and Concentration, based in Ottowa, Ontario, Canada, feels that those undertaking biological engineering research should be doing so in the contained conditions of authorized laboratories. According to Thomas, the biosafety of synthetic organisms has yet to be assessed in a serious way. "As far as we know, regulators are not considering this or else [are] assuming they can be assessed as if they were transgenic species," he says.

Safety concerns are among the reasons that iGEM has decided to restrict the competitive part of the festivities to university-based students. "How biosafety is handled depends on having recognized controls in place. This could be recapitulated at the local level and surrogates provided for what resides in institutions, but the reality is that it's not in place," says Endy, one of iGEM's founding faculty when he was at MIT. DIYbio's Cowell understands the concern but disagrees with the solution. "A better outcome might have been for iGEM to help interested amateurs team up with local iGEM teams, or to work in the same lab but on a separate team. Instead, iGEM wants the amateur community to figure it out on its own, and then—maybe—they'll let us in."

Jim Collins, Professor of Biomedical Engineering at Boston University, thinks the movement is generally a bad idea. He finds that it's not appropriately regulated, and [doing synthetic biology] is sufficiently challenging that he doubts that anything of value will come of it. "At best, they will make a mess; at worst, they will get sick or make someone sick," he worries.

and U Workshop in Los Angeles, curated by the League of Imaginary Scientists. Funding for this remains an issue, however. Jason Bobe, Directory of Community at the Personal Genomes Project (which aims to attract public volunteers willing to have their genome sequenced in return for open disclosure), says that they are looking to corporations and foundations to help underwrite a national BioWeatherMap day to coincide with the week of DNA 2010, an annual event held in April to commemorate the completion of the human genome sequence. As part of the festivities, the BioWeatherMap project hopes to put swab kits into the hands of every high school biology teacher across the nation and to inspire a new generation of students to get excited about biology, genomics and ecology, according to Bobe.

But a few one-off projects, largely inactive websites and some press coverage in places such as *Wired*, *Le Monde* and even the staid *Economist* does not a movement make. And in fact, some of the more legitimate practitioners of grassroots biotech cringe when asked about their so-called movement. "The hype was funny, and it's far from over," says Tito Jankowski, a San Francisco Bay area do-it-yourselfer, who is a biomedical engineer by training.

Another biotech hobbyist notes that while there may be over 1,000 people on an enthusiast's mailing list, only a dozen or so people actually are doing experiments in makeshift labs. This relates to the issue of how many 'amateur' biologists can afford a thermocycler, centrifuge or –80 °C freezer, let alone procure the supplies needed for an at-home biotech lab. Indeed, calls to several companies that supply reagents to mainstream molecular biology laboratories failed to find one willing to deal with individuals. "This is a joke, right?" sums up the tone of the response.

Joe Alper, Louisville, Colorado

The changing economics of DNA synthesis

Robert Carlson

How are the economics of synthetic biology likely to develop in the coming years?

Biological technologies come in many different guises. For millennia, humans have used selection and breeding to direct the evolution of organisms in a sort of top-down approach, a powerful but unpredictable means to achieve a desired behavior. At the opposite extreme, genes and genomes can now be written from chemical precursors, a more precise but sometimes less effective means of producing a particular biological behavior—the design rules for bottom-up engineering of biology in the vast majority of cases are still poorly understood. In between, practicing metabolic engineers use any and all tools at hand to herd and cajole organisms into producing products with market value in the many hundreds of billions of dollars.

At the core of all these approaches to biological engineering is the creation of a particular genomic sequence that produces behaviors according to human desire or need. In addition to nearly a century of evolution and selection based on early knowledge of genetics, we are already four decades into the direct manipulation of genomes through recombinant DNA technology. Synthetic oligonucleotides (oligos) have been available by mail order for the past 20 years, and synthetic genes have been built commercially from those oligos for the last ten. In that time, the number of bases a single individual can synthesize in a day using commercial instruments has increased by five orders of magnitude, whereas the per base cost of synthetic genes has dropped by nearly three orders of magnitude (**Fig. 1**).

I argue here that in the coming years, synthetic DNA manufacturers will come under increasing pressure to reduce costs and decrease turnaround times. At the same time as these DNA synthesis companies face the commoditization of their product, the commercial sector that focuses on assembling and engineering genomes (or at least chromosomes) to create novel products for use in the medical, energy or industrial sectors is likely to become increasingly profitable. Demand for synthetic DNA will consequently spread around the globe as organizations of all sizes exploit biological technologies for many different aims. That some of those aims may be less appealing than others is already prompting calls to regulate synthesis in one way or another. But the global proliferation of demand is likely to limit the effectiveness of regulations implemented on the grounds of improving safety and security.

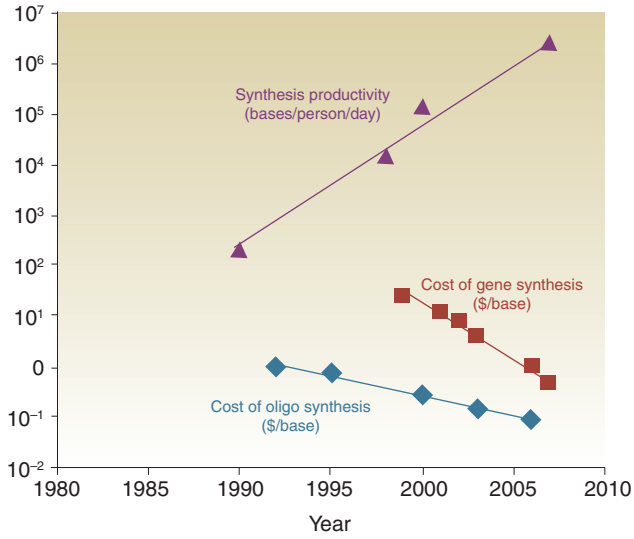


Figure 1 Productivity of oligo synthesis and cost of oligos and genes.

A nascent field

The commercial availability of synthetic DNA has clearly found a use in constructing ever longer genes and now genomes (**Fig. 2**). Recent work at the J. Craig Venter Institute (JCVI; Rockville, MD, USA) has resulted in relatively painless assembly of a 580-kb microbial genome from 101 pieces of starting material each 5–6 kb in length[1]. That this technique makes use of native recombination mechanisms in yeast suggests that it can be implemented in just about any laboratory that takes the time to learn the recipe. Combined with an existing widespread industry that regularly supplies synthetic DNA fragments of 5–10 kb in length, assembly in yeast will put the ability to build a wide range of DNA genomes in the hands of scientists, entrepreneurs and other interested parties worldwide.

Single-step DNA assembly in yeast should prove useful as a tool to rapidly assemble metabolic pathways from many short DNA sequences. Shao and Zhao[2] demonstrated precisely this sort of application in early 2009 by assembling functional metabolic pathways

Robert Carlson is at Biodesic, Seattle, Washington, USA.
e-mail: rob@biodesic.com

from three, five and eight separated genes in a single-step process nearly identical to that implemented at the JCVI. If used with libraries of gene variants, where the sequence of every gene in a synthetic pathway might be varied at the same time, assembly in yeast could be used to rapidly and simultaneously test thousands of different mutations in different genes. This should provide a powerful tool to supplement existing metabolic engineering techniques and potentially to speed identification of useful gene and pathway variants.

New design and error correction technologies will continue reducing the cost of synthetic genes, pathways and genomes. Assembling long stretches of DNA is presently accomplished by annealing overlapping oligos, followed by ligation and PCR amplification and then an error removal step. The process typically results in a population of sequences, many of which contain at least one error consisting of base substitutions, deletions or insertions, due largely to mistakes in the source oligos and annealing errors. Proofreading of multiple individual assembly products is usually required to identify a single molecule of the desired sequence. The many kinds of assembly steps, and the many kinds of errors present in the pool before sequencing, represent a variety of opportunities to introduce improvements in gene assembly.

For instance, Integrated DNA Technologies (Coralville, IA, USA) is regularly producing and selling high-quality oligos >200 bp in length, the use of which in assembling genes should result in much lower error rates (John Havens, personal communication). At the Synthetic Biology 4.0 meeting last year in Hong Kong, Alex Borokov described a design strategy that enables using lower quality, less expensive oligonucleotides to assemble genes in one step[3]. And whereas most gene assembly techniques rely on the removal of mistakes from a sequence pool, Novici Biotech (Vacaville, CA, USA) has just released the ErrASE system, which enables true error correction of genes assembled even from unpurified oligos (Hal Padgett, personal communication; readers should note Novici is a client of my company, Biodesic). Dan Gibson has recently demonstrated that yeast can even assemble overlapping oligos *in vivo* into sequences at least one kilobase long[4]. These many examples of new technologies suggest that continued innovation will reduce the cost of gene and genome assembly, particularly as proofreading by means of DNA sequencing is a substantial fraction of the overall cost.

Practical limits of current technology

Yet it already appears that the technical ability to build large genetic circuits and genomes outstrips our ability to understand and design systems of that size. Whereas the group at the JCVI has demonstrated how to assemble nearly a megabase of DNA—a small genome's worth of genes—UC Berkeley's Jay Keasling and his colleagues[5] are working at the cutting edge of metabolic engineering while manipulating a network of just 12 genes. The cost of the latter project, based on the size of the grant that funded it, is several tens of millions of dollars to pay for infrastructure and labor. Keasling is very upfront about the extent of true design versus tinkering that the team members have been able to employ, and in his talks refers to the not inconsequential number of serendipitous "miracles" that have kept the project on schedule to commercially produce precursors to the malaria drug artemisinin next year[5].

Consequently, the 12 genes manipulated in the artemisinic acid pathway may be a practical upper limit on engineering capabilities in the near term. In the slightly longer term, the more interesting numbers for synthesis may, therefore, be only 10–50 genes and 10,000–50,000 bases. Genetic circuits of this size may represent a limit in complexity for systems with economic value for many years to come. More important questions for would-be genome engineers are, How will the costs fall for constructs of this size? When will DNA of that length be available in days or hours instead of weeks? How soon before one can buy or build a desktop box that prints synthetic DNA of this length?

Improvements in DNA synthesis and gene assembly technologies will be driven by demand. Academic researchers, particularly those attempting to model, build and test complex networks of genes, are likely to consume as much synthetic DNA as their budgets allow. The ability to experiment will be fueled in part by the availability of synthetic DNA at reasonable prices. Industrial consumers are likely to be driven more by product development projects in which timely progress is a key to producing a return on investment, which suggests an increasing market for rapid turnaround.

In the commercial world, biological technologies have been deployed largely in the service of developing drugs and transgenic plants. Those sectors are dominated by relatively large companies that earn profit margins that are the envy of businesses in other manufacturing and service sectors. With the substantial sales that support these margins, pharmaceutical and biotech industry associations advertise that their members spend a larger fraction of revenues on research than other industries. For companies in these sectors, outsourced DNA synthesis and gene assembly are technological and economic levers to reduce labor and infrastructure costs.

Changing economics

But the bio-economy is changing rapidly. Revenues from industrial applications of

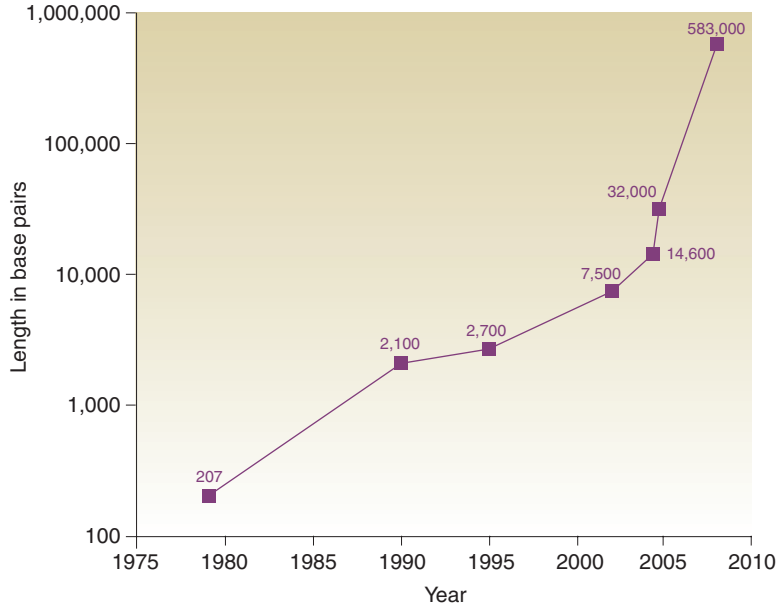


Figure 2 Longest published synthetic DNA[1,12–17] (http://www.synthesis.cc/2009/06/data-and-references-for-longest-published-sdna.html).

biotech appear to have overtaken US revenues from biotech drugs (biologics) and transgenic crops[6], each in the neighborhood of $70 billion, whereas the combination of biofuels, enzymes and materials has reached about $85 billion[7,8].

The relative size of revenue numbers is important for the future of biological engineering for several reasons. Because of lower regulatory burdens, industrial applications are likely to be faster moving and may rapidly become more competitive in their respective markets than biologics or transgenic crops. Green chemistry based on biological processing is already competing with more traditional synthetic chemistry based on petroleum feedstocks in markets worth many hundreds of billions of dollars worldwide[7,8]. It seems likely that in markets that are simultaneously less regulated and closer to the consumer, smaller organizations—operating on smaller budgets and shorter timelines—will have plenty of room to enter markets with products derived from biological processing.

This leads to the question of just which parts of the infrastructure for biological engineering hold the greatest economic value. Is it in the design (bits), or in the objects (atoms)? The synthetic biology market—the ecology of companies that produce and consume products and services related to building genes and genomes—still isn't very big. A generous estimate would put the market for synthetic genes in the neighborhood of $100 million in 2009 (ref. 9). Thus, the revenues for any given synthesis firm are (optimistically) probably no more than a few tens of millions of dollars. Compare the value of the gene synthesis market with the many tens of billions of dollars in revenues from industrial biotech in the United States, and an interesting perspective presents itself.

It is possible that the value of gene synthesis as a service is transitory. Although the assembly of large DNA circuits is presently a technological challenge, and is therefore valuable, the relative value of that assembled DNA is quite small. Of much greater value are the molecules or behaviors specified by those sequences: networks that enable computation or fabrication, enzymes that facilitate processing plants into fuels or fine chemicals, proteins and other molecules that serve as therapeutics and antibiotics.

DNA is cheap, and getting cheaper. Given that the maximum possible profit margin on assembling synthetic genes is falling exponentially (roughly the difference between the cost of genes and the cost of oligos; **Fig. 1**), it would seem that finding value in those particular atoms will become ever more difficult. The design of genetic circuits (resulting in bits) definitely costs more in labor than obtaining the physical sequence by express delivery (resulting in atoms). Moreover, because the electronic specification of DNA sufficient to reproduce the molecule, maintaining proprietary control over sequences of value during the design phase will become ever more important to commercial viability.

In the race to generate new products with new value, firms that develop and sell engineering infrastructure like DNA synthesis and gene assembly will be under pressure to reduce costs and decrease turnaround times. The question is whether customers for DNA of a specific sequence will continue to order it from centralized facilities, or whether economic, technical and regulatory factors might contribute to a decentralization of synthesis. New technologies could enable desktop instruments that provide rapid and secure gene synthesis. Similar technological transitions have resulted in profound transformations of the infrastructure we use for computing, printing and communicating, all of which can now fit in a pocket. At Biodesic, my engineering and design company, our experience is that electrical engineering projects that once required five to ten people working for several years, at a cost of several million dollars, can now be accomplished by one person in less than six months using open source tools. To be sure, there is no guarantee that biological technologies will follow the same route. But neither is there any a priori reason to think that route implausible or unlikely.

We will certainly be hearing more about regulating access to synthesis over the coming months and years.

Market and regulatory forces

To be completely clear, it is not my argument that there will be a gene or genome synthesizer in every small business or even every home, but rather that relatively soon the technology is likely to work well enough to be marketed and used that way. In practice, DNA will be assembled in whatever locale and at whatever scale is demanded by the market and allowed by regulation. And we will certainly be hearing more about regulating access to synthesis over the coming months and years. I suspect, however, that regulations intended to shape the DNA synthesis market will be short-lived and ineffective.

Implementing restrictions on biological technologies that might improve safety and security should always remain among our options. But implementing regulations without a careful examination of possible consequences is unwise. Instituting security measures, and maintaining auditable records of both security and access, will incur costs for producers, users, governments and society as a whole. Understanding the potential costs of restricting access to synthesis first requires examining proposed mechanisms of regulation in greater depth.

To facilitate the control of access to DNA synthesis, Garfinkel *et al.*[10] propose the option of establishing a registry of DNA synthesizers, service providers and certified users. Such requirements would allow DNA synthesis only in what amounts to secure facilities, where security is defined by monitored operation of DNA-synthesis technology either through licensed or permitted ownership of instruments or through licensing of "legitimate users," or both. Sequences submitted to these secure facilities might be kept on file for some number of years to facilitate any forensic efforts. Screening software would examine submitted sequences to identify potential threats in the form of genes and pathways that code for toxins or genomes that code for pathogens.

With respect to the costs of this registry, Garfinkel *et al.*[10] note that "if a review mechanism were too burdensome, small startup firms might shift to in-house DNA synthesis instead." Thus, one of the immediate *social* costs of implementing a registry might be that some otherwise legitimate users opt out of participating due to the *monetary* costs of compliance, thereby limiting the utility of the registry. Given a choice—or if forced by regulatory action to make a choice—some designers of new DNA circuits will inevitably conduct business with synthesis providers who do not maintain an archive of design files. Those who choose to drop off the grid by synthesizing genes in-house could be monitored only if reagents and instruments were strictly controlled. As a result, one potential outcome of restricting access to synthesis might mirror the problem encountered by the US Drug Enforcement Agency when it cracked down on domestic methamphetamine production: information on activities the agency wished to monitor and suppress became much harder to obtain, whereas methamphetamine use continued to rise[7,11]. Similarly, regulatory actions that motivate users to pursue synthesis outside the registry may reduce knowledge

of what is being synthesized, and by whom. Thus, it is not at all clear that regulation will limit access to synthesis technology by users who may be considered a threat. Restricting access to DNA synthesis may motivate some consumers—including those most deserving of scrutiny—to seek access to producers who are either not bound by restrictions or who are willing to ignore them.

In this context, one must also keep in mind the already intrinsically international nature of the DNA-synthesis market[9]. Effective restrictions of access to synthesis must therefore be international in scope and must track the flow of valuable design information through electronic networks, often across borders. This raises the most important vulnerability in DNA synthesis registries and archives, one inherent in the inevitable and increasing reliance on information technology. Whether in the form of electronic signatures, databases of 'legitimate users', screening software or a design tool, this information can be viewed, copied and even altered. Moreover, it is subject to a growing number of security threats that cover the range from simple human mistakes, to fraud, to interception of information during transmission, to complex software attacks on other complex software. Under any international regulatory regime that required screening, individual firms would be faced with exposing their designs to multiple sets of eyes, which would threaten their economic security.

If the policy decision is made by either industry or government to adopt electronic records, another question that must be addressed is, Who will indemnify the security of sequence archives? That is, as the archives by definition hold information that customers deem to be economically valuable, the archives will increasingly be targets for industrial espionage. Who pays for securing the archives long-term? How much does insurance cost? Who is ultimately responsible in the event of a breach?

The future

In the preceding article, I have argued that DNA assembly technologies will become increasingly important tools in specifying synthetic genes and genomes. Those genomes will be used to produce products with revenues much higher than are presently associated with producing the genome itself. As the value of the products generated from synthetic genes and genomes increases, so will the demand to produce DNA faster and cheaper. In that environment, any barrier, whether economic, technological or regulatory, that slows down access to synthetic genes and genomes will be subject to pressure. We are likely to see continued innovation that enables ever more rapid genome construction.

COMPETING INTERESTS STATEMENT
The author declares competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturebiotechnology/.

ACKNOWLEDGMENTS
Portions of this text are drawn from *Biology is Technology*, scheduled for publication in 2010. I would like to thank J. Mulligan, J. Minshull, J. Koschwanez, L. Stewart, D. Endy and R. Wehbring for illuminating and entertaining discussions.

1. Gibson, D.G. *et al. Science* **319**, 1215–1220 (2008).
2. Shao, Z. & Zhao, H. *Nucleic Acids Res.* **37**, e16 (2009).
3. Borokov, A. SynBuild, an accurate high-throughput gene assembly platform that uses microarray synthesized oligos. Poster at Synthetic Biology 4.0, Hong Kong, China, October 10–12, 2008.
4. Gibson, D.G. *Nucleic Acids Res.* published online, doi:10.1093/nar/gkp687 (10 September 2009).
5. Keasling, J.D. Synthetic biology in pursuit of low-cost, effective, anti-malarial drugs. Presentation at the Institute for Systems Biology Seventh Annual International Symposium, Seattle, April 21, 2008.
6. Carlson, R. *Nat. Biotechnol.* **27**, 984 (2009).
7. Carlson, R. *Syst. Synth. Biol.* **1**, 109–117 (2007).
8. Carlson, R. *Biology is Technology* (Harvard University Press, Cambridge, in the press).
9. Newcomb, J., Carlson, R. & Aldrich, S. *Genome Synthesis and Design Futures: Implications for the US Economy* (Bio Economic Research Associates, Cambridge, Massachusetts, 2007).
10. Garfinkel, M.S., Endy, D., Epstein, G.L. & Friedman, R.M. *Synthetic Genomics: Options for Governance* (J. Craig Venter Institute, The Center for Strategic and International Studies, Massachusetts Institute of Technology, October 2007). http://www.jcvi.org/cms/fileadmin/site/research/projects/synthetic-genomics-report/synthetic-genomics-report.pdf
11. Carlson, R. *Biosecur. Bioterror.* **1**, 203–214 (2003).
12. Khorana, H.G. *Science* **203**, 614–625 (1979).
13. Mandecki, W., Hayden, M.A., Shallcross, M.A. & Stotland, E. *Gene* **94**, 103–107 (1990).
13. Stemmer, W.P.C., Crameria, A., Hab, K.D., Brennan, T.M. & Heynekerb, H.L. *Gene* **164**, 49–53 (1995).
15. Cello, J., Paul, A.V. & Wimmer, E. *Science* **297**, 1016–1018 (2002).
16. Tian, J. *et al. Nature* **432**, 1050–1054 (2004).
17. Kodumal, S.J. *et al. Proc. Natl. Acad. Sci. USA* **101**, 15573–15578 (2004).

Parts, property and sharing

Joachim Henkel & Stephen M Maurer

Synthetic biology should look to other industries' models for ownership and open sharing.

Synthetic biologists have spent the past decade trying to recast genetic engineering in the image of electronics. Today's microprocessors are universally assembled from libraries of reusable modules, which are composed in turn of standard parts. The premise behind synthetic biology is that this same approach can be used to design the most complex devices of all—living organisms. But the standard parts agenda is much more than a technological choice. As in Silicon Valley, standardization will also help determine the new industry's structure and economics. These social arrangements will, in turn, have a profound impact on the rate at which synthetic biology generates new products, the affordability of those products and (through affordability) the number of human beings whose lives are actually improved.

We discuss here how the parts agenda is likely to shape commercial synthetic biology, the pitfalls this new industry could encounter and what governments and firms can do to address them. The first set of issues stems from synthetic biology's reliance on large numbers of patented parts. As with earlier 'complex technologies', this suggests that intellectual property (IP) rights will often be hard to identify, fragmented across many owners and sometimes overly broad. All of these factors will make it harder for would-be innovators to obtain the licenses they need to go forward. The second set of issues arises from synthetic biology's defining emphasis on standardization. In the electronics and software industries, the need for common standards has repeatedly produced a 'tipping dynamic' in which one solution quickly comes to dominate the rest. In principle, the dominant parts can be owned by one firm (as is true of Windows, for example), fragmented across many owners (mobile telephony standards), or owned by no one (Linux). We argue that Linux-style openness in synthetic biology is desirable and, to a significant extent, feasible.

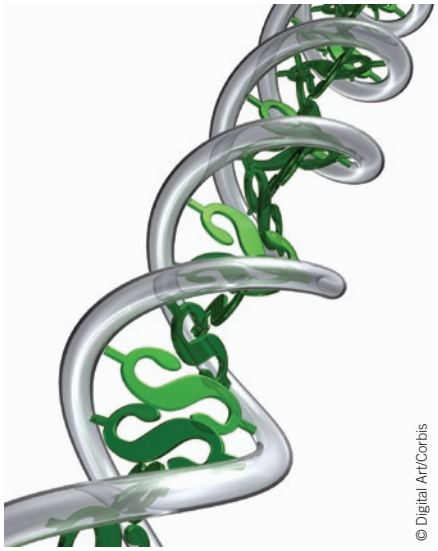
Joachim Henkel is at Munich University of Technology, Munich, Germany and Stephen M. Maurer is at the Goldman School of Public Policy and Boalt Law School, University of California, Berkeley, California, USA.
e-mail: smaurer@law.berkeley.edu

Complex technologies

Commercial applications of the life sciences (for example, biotech R&D) have traditionally involved 'discrete technologies' that generate new products seldom consisting of more than a few individual inventions. In contrast, synthetic biology—with its emphasis on assembling organisms from dozens and eventually hundreds of standard biological parts—is a 'complex technology' similar to those found in the electronics and software industries. This makes it natural to think that the new synthetic biology companies will often resemble Microsoft at least as much as Pfizer.

This complexity has important implications for the management of IP. For example, no mobile phone manufacturer owns all the patents that cover its products. This forces the industry to share technology through cross-licensing instead of using IP to exclude competitors, as commonly occurs, for example, in pharmaceuticals. We expect something similar to happen in synthetic biology. The more complex the systems designed by synthetic biologists become, the less likely it is that any company will own all of the IP rights needed for each R&D project.

Scholars have documented various problems where IP ownership is very fragmented[1,2]. First, firms can encounter an 'anticommons' scenario[3], in which follow-on research is hampered by the high cost and difficulty of negotiating contracts with very large numbers of IP owners. This is aggravated by each individual owner's incentive to overcharge for its IP. Second, many types of cross-licenses (for example, per-unit royalties) tend to generate higher prices for consumers. This is because higher royalties push up each company's costs and therefore prices. This can happen even where payments cancel out so that no firm earns a net royalty.



Synthetic biology's future will depend on who owns its most popular parts.

The existence of these problems suggests the importance of cutting the number of licensing transactions that firms face wherever possible. In principle, this could be done by making standard biological parts unpatentable. Legislatures and courts, however, are highly unlikely to do this. Furthermore, this would also reduce incentives to innovate[2,4]. Traditional private-sector solutions based on patent pools—perhaps with zero royalties—seem more promising[5–9]. Here, the main difficulties are getting contributors to agree on terms and writing agreements that do not exclude competitors in violation of the antitrust laws[10]. An ASCAP-style clearinghouse

for patents would go further by providing licenses to any company that requested one[11–13]. Alternatively, where royalties are much smaller than the expected transaction cost, companies may decide that it is simpler to share their IP in the style of open source collaborations[14,15]. We return to this point below.

But there are other issues beyond licensing. Complex technologies are also more prone to inadvertent IP infringement. This problem is particularly pronounced when patents are overly broad or so numerous that they create webs of overlapping rights or 'patent thickets' that are so dense that infringement becomes almost inevitable[2,5]. In the electronics industry, even large firms find it difficult to identify each and every patent that potentially covers their products. This is due both to complexity of the technology and to the fact that many patents are so vaguely written that they can no longer reliably fulfill their 'notice function'[16]. Genetic engineering already faces significant problems in finding out whether parts are patented or not[17], despite attempts, for example by Cambia's Patent Lens project, to increase transparency. It is reasonable to think that the same problem will similarly affect synthetic biology as designs become more complex. The problem is already evident in the Registry of Standard Biological Parts, where it is seldom clear which parts are or are not patented.

To make matters worse, the increasing risk of inadvertent infringement encourages 'patent trolls': that is, firms that acquire patents not because they want to make products but because they hope to extract extortionate payments from companies that do[18,19]. We expect this problem to become increasingly relevant also for synthetic biology, especially if patent trolls start to acquire patents from bankrupt biotech firms. Industry initiatives to buy up patents are a natural way to mitigate this threat. The Open Invention Network already does this for Linux-related patents in the software industry.

Network effects

IP ownership and royalties are not the only issues. Other criteria may prove even more important in selecting a part for a specific application. Characterizing a new part requires considerable time and effort, and so researchers have a strong preference for parts that have been used before. After all, the only way to learn about parts is to use them. Researchers estimate that the cost of using parts falls 20%–30% each time they are used, so the information obtained by using a part is significant[20].

All else equal, a firm will tend to become locked in to those parts it has used before. If, however, parts information is shared, a firm may find it advantageous to switch to a part that is already widely used across the industry. This preference for widely used parts is an instance of what economists call a 'network effect'. Where network effects are strong, individual lock-in tends to be replaced by global, industry-wide lock-in. Network effects are not new to biology. Indeed, researchers in various biology disciplines focus disproportionately on a half-dozen cell lines out of the many thousands that could be used in principle[21]. The fact that these lines are widely used makes it easier to find out how to maintain and culture them, compare experiments with earlier published work, and acquire them in the first place[21,22].

Economists already know a great deal about network effects from studying the electronics and software industry. Software markets in particular have demonstrated how network effects produce a runaway dynamic in which whichever product starts with the biggest user base attracts still more users until it eventually dominates the industry. Crucially, this dynamic does not depend on whether the dominant standard is owned by one company, several, or no one at all (that is, 'open'). At the same time, ownership matters very much to the price that consumers and follow-on innovators must pay to use or improve the product. This suggests that early interventions to create and promote open standards will often yield important benefits to society.

It is reasonable to think that a similar dynamic will operate in synthetic biology so that popular parts become steadily more entrenched over time. Crucially, such dominant parts could be open or proprietary. If a popular part is open and costs nothing to use, well and good. But if not, researchers will be willing to pay for a proprietary part that comes with a large experience base, so long as license fees are less than the cost of characterizing and learning to work with a substitute part. More generally, the same argument should apply not just to individual parts but also to families of parts that are routinely used together.

Making synthetic biology more open

We have already said that the tipping dynamic can produce dominant parts that are owned by one company, several companies or no one at all. Which regime should society hope for? For existing parts, the answer is simple: open parts are preferable, because they offer the lowest prices to consumers and follow-on innovators. But what about incentives for creating new parts? Such innovative activity is costly, and patents are known to create incentives for innovation in the biotechnology and pharmaceutical industry. So, what can be done to support sharing in synthetic biology, while maintaining incentives for innovators? We see four viable measures.

Wherever possible, use unpatented parts. Many parts are not, or are no longer, patented. Today, academic researchers often care little about the patent status of the parts they use. This is shortsighted because it may be expensive to replace patented parts if and when a project is later commercialized. Deliberately selecting open parts over 'closed' substitutes avoids this, and more generally increases the odds that open parts will become dominant. The problem for now is that researchers often find it difficult to tell which parts are patented and which are not. Extending platforms like the Registry of Standard Biological Parts to include ownership information would help boost open parts usage. Patent offices can also help by requiring applicants to do a better job of specifying claims. The increasing willingness of US, European and Japanese patent offices to deny patents to applicants who fail to disclose a specific gene sequence—that is, who only provide a functional definition without specifying the relevant structural elements—is a useful step in this direction.

Donate parts to the commons. Commercial software firms frequently donate code to public open source projects. They do this for a variety of reasons. These commonly include establishing a reputation, hoped-for reciprocity by others and the desire to build a user base[23]. These incentives should similarly apply to synthetic biology firms. Some firms and universities already do this for parts that are not central to their business (W. Weber, personal communication).

Link public funding to the obligation to share. Many firms in the nascent field of synthetic biology receive public funding. This potentially lets governments adjust the balance between IP protection and sharing without changing existing patent law. In synthetic biology, the main issue is whether the full 20 years' patent reward is needed to elicit investment, especially for companies that receive significant grant support. The problem, of course, will be figuring out how much patent duration these firms actually do need. We think that the best option is to ask firms to specify a desired patent duration as part of their grant applications. In this way,

competition for grants would provide a powerful incentive for companies to limit patent duration and maximize sharing.

Create open parts licenses. Commons models rely on firms' willingness to share information voluntarily. Open source licenses, such as the General Public License (GPL), provide an important additional incentive to share. They do so by requiring those who develop improvements to GPL code, or who merge GPL code with other code, to license the resulting software under the GPL. As a practical matter, this enormously increases the chances that developers will make their improvements public so that the original author can use them.

Commentators have talked about extending open source principles to biology since the late 1990s (refs. 15,24). Despite this, not much has happened. The best-known project, Cambia's 'Bioforge' initiative[25], seems to have elicited little shared research[26]. Within synthetic biology, recent efforts by the Biobricks Foundation to write an open parts license have similarly stopped short of conferring a GPL-style obligation on the recipients to share their improvements[27]. For this reason, researchers' incentives to donate parts are not significantly stronger than they would be in the commons schemes described above.

Ten years on, the absence of anything resembling an open parts regime in synthetic biology is striking. Most commentators (for example, ref. 28) explain it in two ways. First, they argue that biology research requires a much larger up-front investment than software. However, this could be addressed by writing licenses that let companies retain ownership of parts for a commercially reasonable period of time—say, several years—before sharing. The required period would almost always be far less than the 20 years specified by patent law. In fact, schemes that feature sharing after similarly short periods of exclusive ownership already exist and provide important incentives for the developers of the 'embedded Linux' software used in cell phones, machine controls and the like[23]. Second, commentators argue that existing open source licenses rely on copyright protection, which attaches to software automatically at no cost to the author. By contrast, standard biological parts are usually protected by patents, and these are expensive— ~$10,000 per application in the United States[29]. It is difficult to see how even the wealthiest open parts collaboration can obtain enough patents to protect its work.

However, copyrights and patents are not the only choices. Instead, all modern jurisdictions recognize trade secret laws that let collaborators make binding agreements as to when and how to share confidential information. Commentators have long speculated that an open parts collaboration could be built around such agreements. Furthermore, trade secret protection, like copyright, costs nothing to acquire. Instead, the main drawback would be that trade secret agreements—unlike most open source software agreements—require "extremely broad restrictions on dissemination" to nonmembers[3]. Even so, this seems like a small price to pay provided that anyone who wanted to join the collaboration was truly able to do so. Large pharmaceutical companies, which already have long experience keeping and managing trade secrets, should find such collaborations particularly straightforward.

Legally, it is easy to see what such an agreement would look like. Members who joined the collaboration would receive access to a confidential database of parts and parts information. In return, they would promise to share whatever data they acquired in the course of using and/or improving the collaboration's parts after some short period of time. This simple bargain would be the same whether the collaboration consisted of two firms or an entire industry. A potential downside of trade secret protection is that, unlike patents or copyright, it could suddenly disappear if the underlying secret became public. A related and potentially more severe problem arises when a third party independently discovers the secret and patents it. However, these issues do not seem fatal. Instead, trade secrecy exists in all industries, and firms have invented various strategies to manage them both individually and in joint ventures. An open parts collaboration could similarly mitigate risk by allowing members to seek patent rights on the express condition that these could only be asserted against nonmembers. Alternatively, a collaboration could give members the right to make any information they supplied public at any time[30]. This 'defensive publishing' would block third parties from obtaining patents as a matter of law[31]. A famous example of the latter strategy is the Merck Gene Index, a public domain database of expressed human gene sequences[32].

Would companies that use synthetic biology approaches be willing to share information in return for a right that might suddenly evaporate? This kind of open parts model is obviously very different from life science firms' usual strategies for managing IP. In the short run, therefore, the new model will probably encounter a certain amount of cultural resistance. Here, synthetic biology's status as a crossover discipline with deep roots in chemical engineering, electronics and software should predispose it toward sharing. More importantly, companies can be wonderfully receptive to new business models that help the bottom line. For every firm that earns a living by selling patented parts to others, we expect several who see themselves as net consumers with an interest in keeping parts prices as low as possible. This group notably includes the big pharmaceutical companies that have repeatedly used their deep pockets to bankroll projects (for example, The SNP Consortium) aimed at keeping the biology's basic building blocks as open as possible.

In the long run, then, the only real question is whether an open parts model makes economic sense. Will companies that use synthetic biology approaches really share information in return for trade secret protection that might suddenly evaporate? We are optimistic. In the real world, companies can and do routinely enter agreements to share and improve unpatented trade secrets. Extending this model from commercial joint venture agreements to open parts collaborations seems straightforward.

Conclusions

Synthetic biology is bound to change the rules of the game in genetic engineering. Its reliance on large numbers of parts turns the field into a complex technology, and the importance of shared learning implies network effects and makes winner-take-all outcomes likely. Both aspects are compounded by weaknesses of the IP system—in particular, its lack of transparency. Although these problems may seem modest today, they are likely to become much more serious once the synthetic biology industry starts to generate significant profits.

For these reasons—and even though the general usefulness of patents in the life sciences is beyond doubt—reasonable steps to grow the commons and support open sharing seem highly advisable. We have already argued that an embedded Linux–style open parts collaboration makes good legal and economic sense. Furthermore, the open parts idea enjoys widespread support, not just in the academic community but also, to a large extent, in industry. For every front-runner like Amyris (Emeryville, CA), there are several firms for whom sharing is the only way to catch up. Similarly, companies that sell synthetic genes and other support services know that cheap, abundant, high-quality parts are good for business. Open parts are the best way to deliver this result. Finally, government

has repeatedly intervened to promote open source–style sharing in software and, more recently, stem cell research. We think it will be similarly predisposed to support an open parts project. Yet no matter how synthetic biology is made more open, it needs to happen soon.

ACKNOWLEDGMENTS

We are grateful to S. Panke, B. Rutz and W. Weber for comments on earlier versions of the paper and for sharing insights about synthetic biology with us.

1. Van Overwalle, G., van Zimmeren, E., Verbeure, B. & Matthijs, G. *Nat. Rev. Genet.* **7**, 143–148 (2006).
2. Rai, A. & Boyle, J. *PLoS Biol.* **5**, e58 10.1371/journal.pbio.0050058 (2007).
3. Heller, M.A. & Eisenberg, R.S. *Science* **280**, 698–701 (1998).
4. Rutz, B. *EMBO Rep.* **10**, S14–S17 (2009).
5. Shapiro, C. in *Innovation Policy and the Economy* Vol. 1 (ed. Jaffe, A.B., Lerner, J. & Stern, S.) 1–32 (MIT Press, Cambridge, Massachusetts, USA, 2001).
6. Goldstein, J.A. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 50–60 (Cambridge University Press, Cambridge, UK, 2009).
7. Verbeure, B., van Zimmeren, E., Matthijs, G. & Van Overwalle, G. *Trends Biotechnol.* **24**, 115–120 (2006).
8. Horn, L.A. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 33–49 (Cambridge University Press, Cambridge, UK, 2009).
9. Verbeure, B. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 3–32 (Cambridge University Press, Cambridge, UK, 2009).
10. Ullrich, H. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 339–349 (Cambridge University Press, Cambridge, UK, 2009).
11. Van Zimmeren, E., Verbeure, B., Matthijs, G. & Van Overwalle, G. *Bull. World Health Organ.* **84**, 352–359 (2006).
12. Spence, M. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 161–168 (Cambridge University Press, Cambridge, UK, 2009).
13. Van Zimmeren, E. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 63–119 (Cambridge University Press, Cambridge, UK, 2009).
14. Maurer, S. & Scotchmer, S. in *Handbook on Information Systems.* (ed. Hendershott, T.) 285–322 (Elsevier, Amsterdam, The Netherlands, 2006).
15. Hope, J. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 171–193 (Cambridge University Press, Cambridge, UK, 2009).
16. Bessen, J. & Meurer, M.J. *Patent Failure* (Princeton Univ. Press, Princeton, New Jersey, USA, 2008).
17. Huys, I., Berthels, N., Matthijs, G. & Van Overwalle, G. *Nat. Biotechnol.* **27**, 903–909 (2009).
18. Lemley, M.A. & Shapiro, C. *Tex. Law Rev.* **85**, 1991–2048 (2007).
19. Henkel, J. & Reitzig, M. *Harv. Bus. Rev.* 129–133 (June 2008).
20. Henkel, J. & Maurer, S. *Mol. Syst. Biol.* **3**, 117 (2007).
21. Stern, S. *Biological Resource Centers: Knowledge Hubs for the New Economy* (Brookings, Washington, DC, 2006).
22. Henkel, J. & Maurer, S. *Am. Econ. Rev. Pap. Proc.* (in the press).
23. Henkel, J. *Res. Policy* **35**, 953–969 (2006).
24. Maurer, S. *UMKC Law Rev.* **76**, 405–435 (2007).
25. Berthels, N. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 194–203 (Cambridge University Press, Cambridge, UK, 2009).
26. Jefferson, R. *Innovation* **1**, 13–44 (2006).
27. BioBricks Foundation. The Biobricks Public Agreement (2009). <http://openwetware.org/wiki/The_BioBricks_Foundation:BPA>
28. Rai, A.K. in *Gene Patents and Collaborative Licensing Models* (ed. Van Overwalle, G.) 213–218 (Cambridge University Press, Cambridge, UK, 2009).
29. Lawrence, S. *Nat. Biotechnol.* **26**, 1326 (2008).
30. Maurer, S.M. *EMBO Rep.* **10**, 806–809 (2009).
31. Henkel, J. & Pangerl, S. Defensive publishing—An empirical study. <http://ssrn.com/abstract=981444>.
32. Merges, R.P. *Univ. Chic. Law Rev.* **71**, 183–203 (2004).

Building outside of the box: iGEM and the BioBricks Foundation

Christina D Smolke

Innovative community efforts in academia and non-profits to engage student researchers, encourage open sharing of DNA constructs and new methodology as well as build a Registry of Standardized Biological Parts have been central to the emergence of synthetic biology.

One aspect of synthetic biology is to develop tools that make the engineering of biology easier. Such engineering research can benefit from communities and venues that collectively engage and support work to develop, test and support open technology platforms. Two community-based efforts, the International Genetically Engineered Machines (iGEM) competition and the BioBricks Foundation (BBF), have enabled such communities and venues to form through unconventional approaches. With the field of biological engineering poised to achieve hitherto unprecedented levels of precision, efficiency and scale, I provide here a perspective on the role of these two organizations in shaping the ideology, values and culture of the synthetic biology community.

The genesis of iGEM

Last month marked the completion of the fifth annual iGEM competition. Over 1,100 people from 100 teams participated in the three-day event, the iGEM Jamboree held at the Massachusetts Institute of Technology (MIT; Cambridge, MA, USA), at which students presented their research projects to peers and policy experts and a mixed audience from academia, industry and social sciences.

The iGEM Jamboree is now the largest synthetic biology event in the world and, beyond its intrinsic value for participants, highlights for observers several amazing aspects of the field. First, students at the undergraduate and high school levels are incredibly excited about biotech; by participating in iGEM, teams of students work together with the goal of identifying and prototyping an engineered genetic program that addresses a real-world problem or opportunity. Second, young would-be genetic engineers are capable of getting new ideas to work; just some examples of successful projects include *Escherichia coli* that smell like bananas, that are newly responsive to light, that produce a full rainbow of pigments, that float or sink in response to transcription signals or that detect environmental pollutants. Third, an open technology platform based on standard biological parts—even if the parts collection itself remains incredibly immature—can be a powerful enabling tool. The iGEM students receive a kit of the best available genetic parts at the beginning of each competition, and then contribute their favorite new parts to the collection at the end, so that future students can build upon their work—thousands of parts are now available to iGEM students.

The iGEM competition grew out of month-long courses that were taught at MIT by Drew Endy, Tom Knight, Randy Rettberg, Pamela Silver and Gerry Sussman during MIT's extended January intersessions in 2003 and 2004. The objective of these courses was to learn from students how to become better engineers of biology. On the basis of conversations with Lynn Conway, a pioneer of early VLSI (very large-scale integrated) electronics during her time in the 1970s at the Xerox Palo Alto Research Center in California, the MIT instructors decided to initially focus on the idea of decoupling the design and construction of genetic circuits, and later to explore the use of abstraction as a tool for managing biological complexity.

Because of time and technology limitations, combined with the complexity of the systems designed by the students, the projects designed in these courses were not successfully constructed, much less characterized and debugged. Even so, by working directly with students in these early courses, the instructors learned about and developed solutions to three basic challenges limiting genetic engineering work. First, given a limited budget for *de novo* DNA synthesis, the instructors discovered the utility of having students share and reuse parts; this led directly to the world's first Registry of Standard Biological Parts. Second, given the relative immaturity of the gene synthesis industry at the time, many of the students' desired DNA parts could not be synthesized because of problems in cloning or expression; this led the instructors to help obtain, optimize and freely provide variable copy number vectors with enhanced transcriptional insulation for use in the commercial gene synthesis process. And third, given the complexity of system function desired by the students, too much time was being spent simply trying to understand how each system might work; this led to the formalization of a first functional abstraction hierarchy based on a common transcription signal carrier, now called polymerase per second, or PoPS.

Inspired in part by the success of other student-oriented engineering competitions, such as the FIRST Robotics Competition (an annual competition organized by the For Inspiration and Recognition of Science and Technology, FIRST, organization), the group made a decision to extend their early efforts into a multischool biological design competition with funding from the US National

Christina D. Smolke is in the Department of Bioengineering, Stanford University, Stanford, California, USA.
e-mail: csmolke@stanford.edu

The most recent iGEM in October at MIT Killian Court. Teams attended from over 100 universities and from 26 different countries.

Science Foundation. Related decisions were made to extend the event from a one-month design challenge, to a summer-long design, build and test experience.

A competition matures

The first 2004 synthetic biology competition had participation from five invited US universities—Boston University, Caltech, MIT, Princeton and the University of Texas, Austin. The name iGEM was decided upon soon thereafter and the competition has been held each summer since 2005, growing dramatically to its current size of over 100 universities and extending its geographical reach to 26 different countries.

As iGEM has grown, it has expanded its goals, refined its approach, and responded to the needs of a young and rapidly growing community. Early on the leadership within iGEM made a decision to focus the experience on standardized parts and open sharing, where teams were tasked with examining whether integrated biological systems could be efficiently built from standard biological parts. The Registry of Standard Biological Parts, envisioned as an online catalog that would organize and document parts encoding biological functions, thereby became a central resource for making available samples of DNA encoding parts to all participants.

The requirements associated with defining a 'standard' biological part were related to the first technical standard introduced by Tom Knight for a physical parts assembly method[1]. This technical standard specified an idempotent assembly method (assembly reactions that leave the key elements unchanged) for physically linking parts together and associated sequence requirements. These ideas were disruptive to prevailing practice in molecular biology at the time, and as a result met with significant resistance from many in the basic and applied biological research communities. As resources and broader support for such work were not available at the time within the biotech community—whether from federal agencies, foundations or industry—to help develop a registry at a professional scale, the leadership of iGEM asked the teams to build the registry over time through their accrued contributions.

This participant-based 'get' and 'give' approach to developing a collection of standard biological parts led to the first set of challenges faced by iGEM. First, getting people to adopt standards in a field that has been operating without them is difficult. Many laboratories build up their own assembly methods and constructs and will have a laboratory-specific catalog of parts that are incompatible with any proposed standard, such that transferring over those parts and knowledge base to a new standard will require a significant amount of effort, time and resources. Although the purpose of standardization is to streamline a process and ultimately make the integration of parts more reliable and efficient, any such payoff would be on a longer time scale. Therefore, younger laboratories with less of a historical backlog would find it easier to adopt any proposed standards. Second, the first proposed standard was not broadly accepted. As researchers worked with the initial physical assembly standard, they found that it was problematic for coupling certain types of parts together. This led to a feeling from some participants that standards were being imposed that were not applicable for many of the systems they would like to build. And third, the quality of the parts in the registry was not generally good, which presented a huge challenge to the major goal of the registry—the reuse of parts to support efficiency in design and construction. A glance through the registry will show that many parts have not been confirmed as working and do not have any, much less thorough, associated characterization data. Complaints and frustrations grew as teams attempted to use parts from previous years' projects and found that they did not work as designed or in some cases were not even the correct sequence. Although iGEM headquarters (currently at MIT) has more recently implemented a quality control check at the level of sequencing, the sheer number of parts received makes it impractical for iGEM staff to have a direct role in parts characterization and functional validation.

In response, the leadership began putting in place a value system within iGEM that would enable the community to address these challenges over time. In particular, mechanisms were put in place that rewarded team participation in the areas of contributing and documenting biological parts that were compatible with approved standards, contributing characterization data for these parts, and even for developing and documenting new and improved technical standards (see the BBF request for comment process below). Although prizes recognize specific achievements of a select number of teams (best in class), medals are also awarded to those who meet specified requirements; any iGEM team can earn a gold medal. In addition, the teams are provided with the medal requirements up-front, so that they know what the judging will be based on, and are asked to evaluate their own projects in terms of meeting these requirements. In addition to addressing the issues outlined above, the medal system rewards teams for helping another iGEM team, characterizing or improving existing registry parts, and developing advances in human practice issues as they relate to synthetic biology. This reward structure has worked extremely well in building the value system within the community. The iGEM competition has also used the rewards structure to explicitly celebrate

foundational and applied advances by setting up tracks (food/energy, environment, health/medicine, manufacturing, new application, foundational, information processing and software), where the best project in each track is awarded a prize. In addition, special prizes are awarded for specific contributions such as best part, standard, human practice advance, wiki and experimental measurement. The most valued prize for the teams is the Grand Prize, for which the team is awarded the BioBrick Trophy (a gigantic Lego-like machined aluminum brick with the names of winning teams etched on it, similar to trophies used in professional sports) that the winning team holds for a year and then passes off to the Grand Prize Winner at the following year's competition.

Community building

By engaging student researchers directly, an interesting thing has happened over time. The iGEM participants are forming a community and are invested in building out the necessary technologies supporting the engineering of biology. They are actively engaged in tackling the challenges and proposing solutions, as opposed to just complaining about the problems. Through the iGEM experience, they learn the importance of having high-quality, well-characterized parts and standards that support the sharing of these parts. And you can see it working in the community; the parts that work, the parts that are easy for others to take and build into their systems—these get picked up by other teams and used in new projects and new applications (in technology-driven work, this represents success). When teams waste precious time trying to work with poor-quality parts, they can share and document their experiences through iGEM, thus giving back important information to the community. This type of reuse, validation and feedback is often not available through traditional scientific reporting mechanisms, which generally celebrate novelty versus distilling processes to practice. When teams identify problems with existing standards, they can go through the process of identifying new standards that might address existing issues and then put them out to the community for use and comment. These collective experiences have over time helped build a sense of responsibility in many of the teams and have led to improvements in the quality and documentation of parts in the registry.

iGEM's initial goal of inventing and improving the underlying technologies of synthetic biology has expanded and evolved to become much more about education of students and teachers, community building and growing a standard biological parts collection. Many schools have developed courses in synthetic biology based on their experiences with iGEM and some have started or are starting entire research centers focused on the topic. Even the number of successful projects is increasing over time. Although several projects have resulted in peer-reviewed publications with significant follow-up work from the researchers after iGEM[2–5], many more are succeeding in pushing the limits of biological systems that can be engineered under time, financial and expertise constraints.

And this gets to the real test of iGEM: can the process of engineering biological systems be made so efficient and reliable that a team of undergraduates (or high school students) with little experience can successfully build an interesting and exciting system in several months? And, will these systems ever approach the complexity and scale of projects conducted through traditional genetic engineering tools that currently take on the order of 150 skilled researcher years to complete[6]?

Ongoing challenges

By most measures iGEM is a fantastic success; however, it is facing new challenges as a result of this success. iGEM headquarters and individual teams face challenges in continued financial support. Teams are responsible for their own fundraising, which includes fees associated with participating in iGEM and running the team and its research project. The international nature of iGEM, and the differences in fundraising models between countries make this particularly challenging. Funds supporting research through traditional federal agencies or foundations in many countries are scarce, and many of the more successful teams have significant buy-in from their universities or are turning to industrial sponsorship when they can. This is highlighted in the cramming of sponsor names and logos onto the backs of team T-shirts, giving iGEM a feel of NASCAR or professional sports. The differences in funding levels between teams and the intense competitive spirit associated with iGEM bring up questions as to whether something should be done to level the resource playing field (e.g., setting upper limits to budgets), such that huge disparities in resources do not lead many teams to feeling like hopeless participants in the competition. iGEM as an organization has run on lean resources, and at the organizational level additional resources could make a significant difference in the ability to improve the student experience through improvements to the registry, parts collection and educational materials.

There is also an ongoing question about the competition aspect of iGEM. Specifically, many of the students take the competition very seriously. Although this results in high-quality and impressive research projects—and has importantly not hindered the open and supportive culture—it also may have undesired personal consequences. Many students are so disappointed when their team does not make it to the list of finalists that they can be seen crying after the finalist announcements. There are also stories that the amount of time some teams dedicate to their projects is so intense it can be detrimental to other parts of their lives, often leading to the break up of personal relationships. Is this something that iGEM can or should try to change? Or, is this part of the human experience around competitions, especially of this scale? Without the competition part of iGEM, would the community invest as heavily? Finally, there are questions around the post-iGEM experience. After the students finish iGEM and return to their schools and plan for their future career goals, what community do the students find, if any, supporting synthetic biology beyond iGEM?

The BioBricks Foundation

The BBF is a not-for-profit organization that was started in 2004 by many of the people involved in iGEM to represent the public interest in the foundational technologies that help define the field of synthetic biology. The original goal of the BBF was to invent and bring to life a legal framework that accelerates and enables the accrual of an open collection of functional genetic elements encoding standard biological parts. However, as highlighted through experiences with iGEM, the successful development of an open technology platform requires several components to be in place, the first being, in particular, a community of people that supports the platform's development and benefits from its existence. In addition, an open technology platform based on standard biological parts requires that the technical standards that define the parts exist and are open. Therefore, the BBF has also directed its efforts to standards development for the field (legal and technical) and community engagement and development. The subsequent text explores each of these activities in more detail.

To encourage the development and use of technical standards in synthetic biology, the BBF has run several workshops on the topic. These workshops were organized to discuss the importance of technical standards in bio-

tech and prioritize areas most critically in need of standards. From these discussions, the BBF developed and launched a process by which people can define and propose technical standards for biotech through the BBF request for comment (RFC) process, which was inspired by the Internet Engineering Task Force (Fremont, CA, USA) RFC process. An RFC can propose a standard, describe best practices/protocols, provide information, or comment, extend, or replace an earlier RFC. The RFC document is made available online through the BBF website (http://biobricks.org/) and feedback and comments are collected for each RFC.

In its first year, 51 RFCs have already been published. Many of these have been submitted by iGEM teams, as one of the optional tasks for teams to earn a gold medal is to develop and define a new technical standard through the RFC process. Glancing through the list of RFCs, they cover concepts as broad as standard definitions, assembly strategies, part characterization and reporting methods, visual description languages, modeling languages and design tools. Over time, as more knowledge is gained regarding best technical standards, the BBF will likely need to play a role in filtering through the RFCs and determining the smaller set of standards to be used by the field.

A legal framework

The BBF and a team comprising Lee Crews and Mark Fischer of the law firm Fish & Richardson (Boston), Drew Endy of Stanford University (Stanford, CA, USA), David Grewal of Harvard University (Cambridge, MA, USA) and Jennifer Lynch and Jason Schultz of the University of California, Berkeley (Berkeley, CA, USA) have also been developing a legal framework that supports an open collection of biological parts. The final draft of this framework—the BioBricks Public Agreement (BPA)—is now available online through the BBF website for comments (http://hdl.handle.net/1721.1/49434). The BBF felt that an ownership, sharing and innovation framework based on patents (the property rights mechanisms most commonly used in biotech) had substantial limitations in the context of an engineering process based on the reuse of thousands of different components across many different systems. Specifically, the cost and time to define and obtain patent-based protection is too great to support the engineering of many-component, integrated, genetic operating systems. In addition, the costs associated with freedom-to-operate searches become prohibitive for systems comprising dozens of genetic components, never mind anticipated genome-scale engineering projects. Although the BBF has advocated for the need to consider new property rights law in support of the future of biotech, they developed the BPA to support the immediate maturation of an open technology platform supporting genetic engineering.

Mark Fischer was involved in helping to draft the legal frameworks for free software in the 1980s. However, the differences in property rights between biotech (patents) and software (copyright) presented several challenges to a direct translation of the licenses used to support open and free software. As a result, the BPA represents a bilateral agreement (or contract) between the contributor and user. The language within the BPA allows the contributor to acknowledge invention over the uses of a part, disclose information on whether there is a patent on it or not, and promise not to assert any property rights against others under certain conditions of use, so that the part can be freely used. The BPA also allows the user to state acceptance of use of the part and promise to use it according to the conditions put in place by the agreement. However, the agreement does not put any encumbrance on downstream uses, such as a give-back or share-alike clause. In doing so, the BBF hopes that the BPA will support the development of a shared open platform that both academics and industry can use, while still allowing proprietary systems to be built upon this open platform. The aim of the BPA is to reduce the legal ambiguity around the use and reuse of standard biological parts, and the BBF hopes that the BPA will encourage both industry and academia to support and play a role in the development of a next-generation open technology platform in biotech.

Community engagement

The BBF has also worked to support the broader synthetic biology community. In particular, the BBF has recently taken on the role of lead organizer of the synthetic biology conference series (most recently SB4.0 in Hong Kong; http://sb4.biobricks.org/). BBF's leadership of this conference series has allowed many diverse communities to learn about and engage with issues of safety, security, equity and ethics relating to the field of synthetic biology. In addition, the activities of the BBF in iGEM, technical standards and technical standards workshops play an important role in building and engaging the community. The BBF also directs efforts to developing educational materials regarding policy issues related to synthetic biology, in particular ownership, sharing and innovation frameworks underlying biotech.

Because of the diversity of backgrounds represented in the synthetic biology community, the BBF plays a key role in providing leadership and a focal point for this growing field. The BBF has focused its early efforts on addressing very challenging concepts in the field. Property rights law in biotech through patents is well established and entrenched, such that work to change this system to one that might be more appropriate for a future biotech meets significant resistance. The BPA is a step toward building a community that supports an open technology platform in biotech. However, for this vision to truly succeed, high-quality open parts are needed. In addition, most people (including foundations and companies) look to biotech as a set of applications. The BBF is working below the level of applications. Although their work in community building and legal and technical standards supports all biotechnological applications, raising funds for such foundational work is typically much more challenging.

Conclusions

The systematic application of engineered biological systems to the problems posed by hunger, disease, environmental quality and finite resources remains both extremely compelling, yet challenging, given the current state of tools supporting biotech. Both iGEM and the BBF are leading different, but synergistic, efforts focused on developing community, sharing and open technology platforms supporting biotech. Importantly, although advanced technologies can be used for good or harmful purposes, the activities of iGEM and the BBF, including education, outreach and community building, are directed toward biasing systems heavily in favor of constructive outcomes. The ideology, values, tools and culture realized by iGEM and the BBF seem likely to continue to make important contributions to the foundations of synthetic biology going forward.

ACKNOWLEDGMENTS

I would like to thank D. Endy, M. Fischer, D. Grewal, R. Rettberg and P. Silver for discussions on iGEM and the BBF.

1. Knight, T. *Draft Standard for Biobrick Biological Parts* (OpenWetWare, MIT, Cambridge, MA, USA, 2007). http://hdl.handle.net/1721.1/45138
2. Levskaya, A. *et al. Nature* **438**, 441–442 (2005).
3. Haynes, K.A. *et al. J. Biol. Eng.* **2**, 8 (2008).
4. Baumgardner, J. *et al. J. Biol. Eng.* **3**, 11 (2009).
5. Ciglič, M. *et al. IET Syn. Biol.* **1**, 13–16 (2007).
6. Ro, D.K. *et al. Nature* **440**, 940–943 (2006).

Our synthetic future

Nigel M de S Cameron & Arthur Caplan

Two prominent ethicists provide their views on the ethical debates surrounding synthetic biology.

What do public attitudes to new technologies tell us about synthetic biology and its potential impact on society? To what extent will new capabilities in biological engineering empower the research community to realize applications that the public find most troubling? And how should engagement with the public on the implications of genome synthesis and engineering be managed going forward? Two ethicists provide their perspectives.

Beyond the silos

Nigel M de S Cameron

When a hot technology prospect like synthetic biology gets the *New Yorker* treatment[1], it has plainly arrived—at least in the conversation of the cognoscenti. This is something of a surprise because engagement with the implications of science and technology (aside from gadget worship) seems to be curiously absent from polite American conversation. Yet it is the United States that dominates global emerging technology R&D on a vast scale, and on which both the US economy and US security depend more than most people imagine.

This conversational failure is not limited to cocktail parties, or indeed to the United States, though Europeans are generally more predisposed to discuss such subjects. Moreover, the absence of serious dialog evinces a worrying cycle of disinterest that threads right through the high (and low) culture of the twenty-first century—including the media, culture's lens; and, of course, the political classes, to which generally falls the task of shaping our various national conversations.

The stakes could hardly be higher. By bringing engineering and biology to a common focus, synthetic biology offers the prospect of the design and manufacture of biological organisms with properties that are selected and may supersede, and indeed entirely eclipse, those present within the natural order of things. As pioneer Drew Endy of Stanford University sums it up in the *New Yorker* in a masterly epitome of both enthusiasm and soul-searching: "It's scary as hell. It's the coolest platform science has ever produced, but the questions it raises are the hardest to answer."

The past generation has witnessed the slow emergence of a fragmented science and technology policy agenda that bears assorted labels, most notably nanotechnology, neuroscience, artificial intelligence and now synthetic biology—often captured together in the tag 'converging technologies'. Convergence entails, among other things, an emerging commonality in the policy and ethical agenda that, increasingly, is seen to mirror technological development and that has come to haunt the more reflective scientists involved. That is to say, if the question is the reengineering of human life to give members of *Homo sapiens* new capacities, it may be secondary whether the process comes about through nanoscale engineering of neuroprosthetics or mechanisms of biological design.

Our culture is presented with what the lawyers would call a case of first impression. Although we have flirted since the advent of the nuclear age with the possibility of destroying ourselves, never before has the species faced the option of reinvention to order. And pressing attendant questions emerge: where is it that such a conversation should be located? How should a species make such a choice (assuming, of course, that there is a choice to be made)? While politics is focused on so many (other) issues of immediate moment, how are leaders to frame issues that seem so far ahead, fraught with uncertainty, yet gargantuan in their import? These are matters that should keep us awake at night—both leaders in science and technology and those in government and wider culture. And certainly not just in 'ethics'.

A big downside of the coinage and institutionalization of 'bioethics' has been the addition of yet another silo to public culture and policymaking, in which disaggregated units of conversation make all fundamental problems harder to tackle. If ethics bodies are to have a role in framing our conversation about the human future, they will need a new level of integration with a newly focused policy apparatus. 'Bioethics' as a public policy phenomenon has tended to offer a way of shunting issues off, not onto, the policy agenda. Yet in the democracies, policy represents an ineluctably ethical enterprise. Like it or not, practical ethics is the daily domain of the policy community.

Like nanotechnology, which has drawn a good deal more recent attention, synthetic biology offers a door to possibilities beyond our imagination that could flow from present, useful and relatively modest achievements (such as the development of new drugs). Like the genetics on which it builds, itself still stained disturbingly by the eugenics that shaped its past in the early twentieth century, synthetic biology offers the prospect (distant, but acknowledged) of designer choices by some humans in respect of others. The *New Yorker* illustrates its article with a full-page

Nigel M. de S. Cameron is at the Center for Policy on Emerging Technologies, Washington, DC, USA.
e-mail: ***nigel.cameron@c-pet.org***

picture of a couple building their child from blocks of Lego.

Of course, it is not as if no one at all has been noticing. The lead synthetic biology critic as yet has been the memorably named Canadian 'Action Group on Erosion, Technology and Concentration' (ETC Group) that is active in international nongovernmental organization circles and made its name pressing for a nanotechnology moratorium. Its 2007 report is entitled "Extreme Genetic Engineering"[2]—a coinage that could misfire in a nation given to enthusiastically embracing 'extreme' sports and makeovers. I carried around a copy of the report during a visit to a leading synthetic biology lab, and, as I had hoped, it handily sparked some conversations. As research supports the commonsense view that how issues are framed has a lot more to do with how people assess new developments than we might wish, the question "who brands the conversation early?" is a vital one.

There were plainly no branding consultants present at the naming of synthetic biology "synbio", or the homonym would never have been allowed. In religious America, 'SinBio' might just catch on as the label 'Frankenfood' has in gourmet Europe—in an informal branding exercise that, for better or worse, has severely hobbled the spread of genetically modified (GM) crops. One of the lessons Europeans learned from the GM furor was to encourage 'upstream' discussion of emerging technologies, and get the critics to make their points early—so they can be either heeded or disputed, and help create a more mature public grasp of what is at stake. In the United States, despite wide agreement that the ELSI program (funding ethical, legal and social issues arising from the human genome project) was either a success or, at worst, a harmless investment in risk management, the two more recent big centers of controversial gravity—nanotechnology and now synthetic biology—have had far less generous (far-sighted?) attention.

There are three basic dimensions to the policy and ethics questions raised by synthetic biology. First, there is risk—essentially the issues raised by any technology: specifically, if something goes wrong, or, alternatively, if *someone* goes wrong. In the 9/11 century, haunted as we are by the prospect of retail weapons of mass destruction, what new capacities might synthetic biology put into the hands of smart dissidents—not least, in the context of 'open wetware'? The recent news that a scientist at a leading European research facility has been charged with Al

This cartoon and an article appearing in the pages of the *New Yorker* magazine this September signaled the arrival of synthetic biology in the intellectual mainstream, if not yet in the wider public's consciousness.

Qaeda involvement underlines the nontrivial nature of this anxiety. But what of innocent miscalculation, in a synthetic biology version of the risk scenarios discussed by Bill Joy in his provocative *Wired* essay "Why the future doesn't need us"[2] or, on the other side of the Atlantic, Sir Martin (now Lord) Rees's 2003 book *Our Final Century*[3] (set breathlessly before the US public in the less credible guise of *Our Final Hour*)?[4]

At the heart of the risk discussion lies the problem of uncertainty as to what future developments will arise, and the lack of any consensus as to the level of risk that the public is prepared to tolerate. People tolerate high levels of road fatalities as the price for the motor car's contribution to freedom of movement, but they expect essentially risk-free air travel and public transportation. Where on the spectrum will transformative technologies lie?

Second, there are concerns as to legal and non-accidental uses, especially by governments. This represents a subset of a vast and neglected question, as year-by-year technological advances place greater powers in the hands of governments—both over their citizens and to deploy in pursuit of security and other ends in the wider world.

And third, what are the implications of these new manipulative possibilities for the human future? How do they apply both in the design of individuals and to the shift in the nature and scope of the human experience and lifespan?

So what to do? As with nanotechnology, synthetic biology offers an entire arena of possibilities that complicate the policy discussion—and any approach to establishing norms, whether through regulation or otherwise. This is important as suggestions emerge for a similar approach to that of the storied 1975 Asilomar conference that faced head-on the implications of recombinant DNA and was key in building awareness of ethical and risk issues into the development of the technology—while mitigating public concerns as to its misuse.

So it is not possible simply to suggest a new Asilomar, although something like it—on an international scale—would take us a useful step down the road. What is required in parallel is continual capacity-building in the key agencies handling both technology policy and its ethical and social dimensions—within individual jurisdictions, and also within the relevant multilateral agencies (intergovernmental organizations). Such capacity building is especially important with respect to the public communications functions of these organizations and their engagement across both scientific disciplines and individual departments of government. The ELSI parallel also is useful, although the synthetic biology conversation needs to generate the kind of social and ethical discussion that ultimately shapes all policy and is too consequent to be shuffled off into an 'ethics' silo or simply contracted down through grant mechanisms into individual research efforts. On the global scale, the United Nations Educational, Scientific and Cultural Organization (UNESCO) Universal Declaration on Bioethics and Human Rights[5] offers a useful, if modest, point of departure; for the key to human engagement with the technological wonders of the twenty-first century is likely to lie in our classic concerns for human rights and dignity—always with an eye cast over our shoulders at the shadow of eugenics that so besmirched genetics a century ago—as we ponder our embrace of the new powers that we are being offered.

1. Specter, M. A life of its own. Where will synthetic biology lead us? *The New Yorker* 28 September 2009, 56.
2. <http://www.wired.com/wired/archive/8.04/joy.html>
3. Rees, M. *Our Final Century: Will the Human Race Survive the Twenty-first Century?* (Heinemann, London, 2003).
4. Rees, M. *Our Final Hour: A Scientist's Warning: How Terror, Error, and Environmental Disaster Threaten Humankind's Future In This Century—On Earth and Beyond* (Basic, New York, 2003).
5. <http://unesdoc.unesco.org/images/0014/001461/146180E.pdf>

Moving ahead but with greater controls

Arthur Caplan

A robust societal commitment to synthetic biology promises to yield all manner of benefits—the creation of adequate sources of cleaner fuels, the reduction of carbon emissions, the production of more and cheaper food, the identification of more efficient ways to create medicines, more fresh water and the building of bugs that will attack pests and pestilences that do so much harm to plants, animals and us. Indeed, according to some practitioners of synthetic biology, it is only our hubris about our own genome and ignorance of the microbial world around us that keeps the field from occupying center stage in the debates over where the biggest breakthroughs are most likely to occur in the coming decades.

So how could anyone play the role of ethical spoilsport when we have the means to solve our most pressing problems almost in our grasp? Still, some say no to the apparent Eden that lies before us if we will only permit microbial tweaking to energetically commence. Apparently immune to the huge promise invoked for synthetic biology, they counsel against moving forward with the creation of novel, designer life forms. Synthetic biology has engendered a bit of a moral backlash built mainly around the idea that it is not our place to make new life forms.

Some worry that engineering life is an activity that ought not be pursued because it is not appropriate for any power other than the divinity to engage in creation. Such concerns, however, are not likely to curtail synthetic biology. Nor should they. The issue of novel creation and humankind's role in it was settled long ago. There has simply been too large an impact on the constitution of the earth's living beings resulting from human intervention—tangerines, passenger pigeons, roses, collies and Louise Brown (the world's first test tube baby), among others. No major religion is opposed in principle to humanity trying to alter the natural environment. It is mainly secular critics of synthetic biology who invoke the divine in expressing ethical anxiety about synthetic biology.

Given its promise, synthetic biology should not be derailed by talk of the danger of 'playing God'. Scientists stuck writing grants year after year to continue their synthetic biology research do not see themselves as divine beings. And they are, as scientists, deadly serious rather than playful about extracting benefit from synthetic biology. The degree to which synthetic biologists are 'playing' when it comes to creating new life is tiny.

So, if metaphysical cautions are not going to derail things, is there nothing to worry about from the point of view of ethics and public policy as scientists begin aggressively manipulating viruses, bacteria, algae and other microbes to suit human purposes?

Once God is sent to the ethical bench, some serious sources of worry emerge—not serious enough to stop synthetic biology from moving ahead, but sufficient to warrant answers before the field goes much further.

Two worries in particular stand out. First, can we be sure that whatever is made will stay where its creators want it to? And second, can we be sure that those whose aims are malevolent will not gain access to techniques for designing life that could do enormous harm?

There is very little about the history of human activities involving living organisms that provides confidence that we can keep new life forms in their place. We do not have the national or international oversight and regulation requisite to minimize the risk of the creations of synthetic biology causing harm by showing up uninvited owing to accident, inadvertence or negligence. People have been introducing new life forms for hundreds of years into places where they create huge problems. Rabbits, kudzu, starlings, Japanese beetles, snakehead fish, smallpox, rabies and fruit flies are but a short sample of living things that have caused havoc for humanity simply by winding up in places we do not want them to be. Sometimes, those involved in creating new life forms have accidently lost track of the animals, insects or plants they were working with, as happened with the introduction of 'killer bees' into South, Central and North America. And in other cases inadequate attention to oversight allowed life forms to escape and wind up in places they were most certainly not wanted, such as the appearance in the food chain of genetically modified 'Starlink' corn containing the insecticidal Cry9C protein unapproved for human consumption.

A huge problem that has not been adequately addressed is what standards of control should govern the creation, introduction and release of novel life forms. Should there be specific restrictions on the kind of life forms that can be engineered so as to minimize threats to human, animal and plant health? Should synthetic life forms be engineered when possible to use a different amino acid code from 'natural' organisms or to expire after a finite period of time (an idea pioneered by Monsanto (St. Louis) with genetically modified seed containing terminator genes, which proved controversial as a way to protect intellectual property)? And if these rules are articulated, which agencies will have clear responsibility and authority for enforcing them? And can enforcement be made uniform, coordinated and transparent?

Not only is there a lack of agreed-upon regulations and regulators in place to help manage the products of synthetic biology, few provisions have been made to ensure that the techniques involved or the knowledge generated do not fall into the wrong hands. In an age of terrorism and bioweaponry, that may not be ethically sound public policy.

With the appearance of the nuclear bomb at the end of the Second World War, great efforts were made by the United States and other nations to keep secret the knowledge of how to create these deadly weapons. International organizations sought treaties that would control the proliferation of these weapons and even attempt to place the creation of some forms of weapons off limits. National restrictions were placed on who could work on nuclear weapons and what could be published about them. None of this has been done for synthetic biology, despite the potential danger posed by the creation of weaponized microbes, germs and viruses that might be engineered to decimate our food supply, poison our water or cause pandemic horror in human populations.

Both environmental control and protections against misuse merit more attention than they have received. International coordination is essential if the public is to feel comfortable that both matters are being managed. Neither poses an insurmountable obstacle to the advancement of synthetic biology. But a failure to vigorously attend to both could set the field back just as the promise of synthetic biology, if somewhat over-hyped, is ready to deliver much good.

Arthur Caplan is at the Center for Bioethics, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
e-mail: caplan@mail.med.upenn.edu

Should moral objections to synthetic biology affect public policy?

Gregory E Kaebnick

Moral concerns as to the relationship of synthetic biology with nature do not provide a convincing basis for more stringent regulatory oversight of the field.

There is a growing view that synthetic biology not only promises to engineer organisms that serve purposes ranging from medicine and agriculture to industry and environmental remediation, but also threatens—perhaps more than any other technology—to change the human relationship to the 'natural' world in morally undesirable ways[1–3]. Others dismiss this concern out of hand. Drew Endy, one of the leaders in the field, has asserted that "the questions of playing God or not are so superficial and embarrassingly simple that they're not going to be useful"[4].

Certainly, the concern about the human relationship to nature needs to be articulated more clearly. It can, in fact, be spelled out in three different ways, which are based on very different philosophic claims, make different assumptions about what 'nature' means, and have different implications for the public regulation of synthetic biology. Some are grounded in large claims about the nature of reality, some only on moral values and some depend crucially on possible consequences. None is superficial or simple. At the same time, once spelled out, none easily generates any special regulatory constraints on synthetic biology.

Metaphysical mistakes

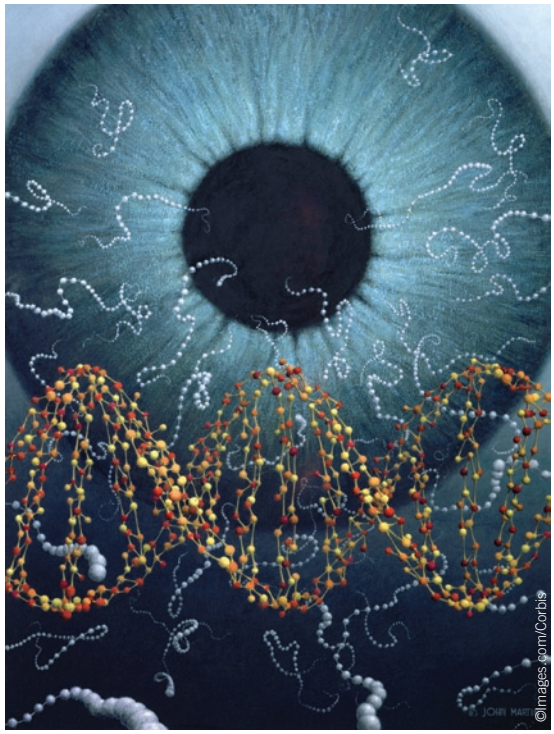
The first possible form of a concern about how synthetic biology might change the human relationship to nature is a metaphysical claim—a claim, that is, about the nature of reality. The claim has two parts, one about the categories of things that exist and another about the moral significance of those categories.

Gregory E. Kaebnick is at The Hastings Center, Garrison, New York, USA.
e-mail: kaebnickg@thehastingscenter.org

Much of the language casually thrown around in debates about synthetic biology suggests some version of a metaphysical claim. The term 'playing God', for example, insinuates that humans are inappropriately stepping outside their proper role in the cosmos—that is, they are making a mistake about the category to which humans belong in the order of things, and in so doing making a moral mistake. The medical ethicists Joachim Boldt and Oliver Müller[1] come very close to this position, if they do not actually hold it, when they argue that synthetic biology is ethically more problematic than other biotechnologies because it constitutes not merely the manipulation of life but the very creation of life. With the emergence of synthetic biology, they write, the human role in the cosmos changes "from '*manipulatio*' to '*creatio ex existendo*'," which is a "fundamental change in our way of approaching nature"[1].

One kind of metaphysical mistake that might be imputed to synthetic biology, then, is the inappropriate elevation of humans. Another is the inappropriate degradation of life. By 'creating life', according to this version of the claim, humans are making a mistake about the category to which living things belong. Often, this claim is merely suggested. Prince Charles touched on it when he lamented that biotech seemed to be leading to "the industrialisation of Life"[5]. In critiques concerned primarily with the social consequences of synthetic biology, Canadian nongovernmental organization the Action Group on Erosion, Technology and Concentration (ETC Group) conveys it obliquely with such titles and subheads as "Who Owns Nature? Corporate Power and the Final Frontier in the Commodification of Life"[6] and "Original Syn?"[7].

To object to synthetic biology along either of these lines, as a failure to recognize the



Watchful oversight of emerging approaches in biological engineering to a large extent should not be based on the various moral concerns that relate to the effects of the technology on the natural world.

appropriate metaphysical category either of humans or of life, is to believe that a very serious moral error is imminent because synthetic biology violates the very structure of reality, perhaps one dictated by a deity (for example, God). But defending either of these forms of the concern also requires defending a larger metaphysical position that makes sense of it. That defense, which will require explication of an overall worldview and perhaps of God's role in creating the world and the proper relationship between God and humans, is not likely, in a modern liberal democratic society, to serve as the basis for public policy that limits or bans the field.

A further problem with both lines of thought is that whether synthetic biology represents a shift from manipulating to creating is at best debatable. Several existing biotechnologies converge in synthetic biology, and the heart of the field is arguably just the ongoing refinement and extension of research on genetic engineering[8]. Furthermore, even the work that is most dramatically 'synthetic'—the creation of a protocell and a minimal genome to put into it—still starts with existing materials. By contrast, the kind of creating with which God is credited is creation *ex nihilo*.

Intrinsic wrongs

A concern about how synthetic biology changes the human relationship to nature can also be understood as a merely moral claim. In this form of the concern, the moral standards at stake are a product of human culture or reason rather than of the structure of reality.

One way to articulate a purely moral concern about synthetic biology would be to show that synthetic biology undermines morally significant concepts. For example, Mildred Cho and coauthors[3] wrote that synthetic biology might threaten the perceived specialness of life. Alternatively, Boldt and Müller[1] argue that synthetic biology might, by making humans "creators" of nature, unjustifiably inflate humans' understanding of themselves.

These contrasting claims are obviously speculative and may prove unfounded. One reason for skepticism is that synthetic biology in its current form is concerned almost exclusively with the engineering of single-celled organisms, which is likely less troubling than the engineering of more complex organisms[9]. If the work is also restricted to the laboratory and the factory, and the release of organisms into the wild forbidden or restricted, then it might not be seen as broadly changing humans' views of other living things. Finally, one might wonder why the work's possible conceptual implications generate a moral objection. It would not be the first time that science has challenged humans' views about life and their place in the cosmos. A very cogent argument must be given to explain why the conceptual implications are so problematic that they generate special regulatory constraints.

Another way to articulate a purely moral concern about synthetic biology would be to argue that 'nature' refers, not to metaphysical categories, but just to the natural environment, more or less independent of human intervention, and that some human interventions into nature are morally undesirable in themselves—intrinsically undesirable, that is. To understand the concern this way is to see synthetic biology as analogous, for example, to the extirpation of naturally occurring species or the destruction of wildernesses—other environmental interventions that many consider intrinsically undesirable.

Defense of the view that synthetic biology is intrinsically morally undesirable rests on two key stipulations, having to do with the meaning of nature and the bounds of morality.

The environmental philosopher Christopher Preston[2] objects to synthetic biology along these lines. He argues that synthetic biology intervenes in nature in a way that "traditional molecular biotechnology" does not. "The relevant difference," he explains, "is that traditional biotechnology has always started with the genome of an existing organism and modified it by deleting or adding genes"[2]. By contrast, because synthetic biology would "create an entirely new organism," it crosses a cherished line: it "departs from the fundamental principle of Darwinian evolution, namely, descent through modification."

Defense of the view that synthetic biology is intrinsically morally undesirable rests on two key stipulations, having to do with the meaning of nature and the bounds of morality. First, the term 'nature' must be understood as distinguishing what does not result from human intervention from what does. A bright line is surely not possible, given the extent of human influence in the modern world, but a bright-line definition may not be necessary, just as it is not for the morally significant concepts of 'personhood' and 'lying.' Perhaps the distinction can be rendered usable by drawing on an assortment of widely accepted examples—wolves and the Alaskan backcountry on one side, Chihuahuas and Midwestern American farms on the other—while admitting that many examples fall into a gray area in between. Furthermore, we might be able to regard the distinction as a matter of convention rather than a timeless fact; we might be able to hold, that is, that something is natural when the degree of human intervention in it does not cross socially established bounds. 'Natural' is used this way in the labeling of produce: anything available in the supermarket is to some degree a product of human intervention, but most people allow that organic orange juice can be labeled natural but not the fizzy beverage Tango. Similarly, although a restored creek or prairie is achieved only through human interference, most people would consider it 'natural'.

Second, moral opinions about human interventions into nature must be possible. It must be possible, for example, that driving a species into extinction, logging a forest or perhaps even just altering a geological feature can be considered intrinsically morally undesirable. The theories of morality dominant in Western philosophy—utilitarianism and Kantian theories, in particular—do not easily accommodate this kind of valuation. Even so, contemporary attitudes concerning the environment and public policy (for example, the US Endangered Species and Wilderness acts) suggest that many people have a wider view of moral value.

Although far from unassailable, these stipulations are now reasonably widely accepted. Accepting them, however, does not mean agreeing that synthetic biology is intrinsically morally undesirable. They are only necessary conditions for that view, and they are consistent with thinking that synthetic biology is acceptable, or that at least it should not be publicly restrained.

Indeed, there are several reasons to think that synthetic biology should be tolerated, at least at the level of public policy. First, even among those who have intrinsic objections to synthetic biology, many would still be willing to weigh them against the possible benefits. The value at stake in any human intervention into nature might be overridden by other moral considerations; we might still log an old-growth forest.

Second, the strength of the objection must be assessed; even if we agree that synthetic biology is undesirable, it might not be deeply undesirable. The human–nature issues that

have most alarmed the public, and that have led to public policy, have concerned damage to the natural world, and perhaps permanent and universal damage: when the passenger pigeon was killed off, for example, its place in nature was probably gone forever, and its absence could be observed and felt by anybody. In contrast, the creation of a new kind of organism is a creative act. Arguably, the natural world would remain unchanged, at least if the organism remained confined to the laboratory or manufacturing setting.

Environmental concerns are therefore the wrong analogy for showing that synthetic biology is intrinsically undesirable. Bearing in mind that synthetic biology is arguably only a refinement and extension of gene transfer, one should look instead to the debates over genetically modified (GM) crops and livestock. That debate, however, also does not support a restrictive view of synthetic biology, as there is no consensus that GM foods and animals should be banned. Even some commentators who take seriously concerns about how GM foods affect the human relationship to nature recommend merely that the relevant food products be labeled[10,11].

Preston's[2] objection to synthetic biology echoes the explanation offered by the science writer Michael Pollan[12] of why he found GM potatoes troubling: when new varieties are created through conventional breeding, he argued, they can be seen both as products of human creativity and as an example of adaptation to fill a special ecological niche; with genetic modification, only the story of human creativity makes sense. The evolutionary story is suppressed[12]. But Pollan did not draw the conclusion that the potatoes should be banned, or even that they were clearly wrong. He just didn't eat the potatoes.

Environmental harms

One final way of understanding the concern synthetic biology raises about the human relationship to nature collapses it, in effect, into concerns about the field's possible consequences. One frequently mentioned fear is that synthesized organisms might leak out of the laboratory or factory, perhaps mutate and become established in the wild[13].

Unlike other forms of the concern about how synthetic biology might change our relationship to nature, understanding it as a straightforward concern about how synthetic biology might damage actual living things in the world around us seems to be a very plausible candidate for grounding public policy.

Another fear is bioterrorism: particularly if terrorists targeted agriculture, a synthesized pathogen might be suitable, or evolve to become suitable, for other hosts in the environment. These possibilities are worth taking seriously for the same reasons we take seriously the environmental threats posed by other kinds of industry or agriculture; among these reasons is the intrinsic value widely given to nature.

Unlike other forms of the concern about how synthetic biology might change our relationship to nature, understanding it as a straightforward concern about how synthetic biology might damage actual living things in the world around us seems to be a very plausible candidate for grounding public policy. It requires no special defense beyond that already offered for policies to protect rare species and undeveloped lands. For the same reason, however, no special policy implications seem to follow from it beyond already familiar concerns about environmental impact.

Conclusions

The upshot, then, is that the different forms of the moral concern about synthetic biology's effect on nature have very different implications. Of the three forms considered here, all may be worth taking seriously as personal moral positions, but the two that have radical implications for public policy are also implausible bases for policy, whereas the one that is a plausible basis for policy would support a policy position that is identical to our present approach. This third position merely gives reason to ensure that the cost-benefit assessment of synthetic biology includes the possible consequences for environmental destruction or amelioration in addition to those for human well-being.

ACKNOWLEDGMENTS

This work was funded by grants from the Alfred P. Sloan Foundation and from the US National Endowment for the Humanities.

1. Boldt, J. & Müller, O. *Nat. Biotechnol.* **26**, 387–389 (2008).
2. Preston, C.J. *Environ. Values* **17**, 23–39 (2008).
3. Cho, M.K., Magnus, D., Caplan, A.L., McGee, D. & the Ethics of Genomics Group. *Science* **286**, 2087–2090 (1999).
4. <http://www.edge.org/documents/archive/edge237.html>.
5. Charles, Prince of Wales. Questions about genetically modified organisms. *Daily Mail* 1 June 1999.
6. <http://etcgroup.org/upload/publication/707/01/etc_won_report_final_color.pdf>.
7. <http://etcgroup.org/upload/publication/602/01/synbioreportweb.pdf>.
8. Brent, R. *Nat. Biotechnol.* **22**, 1211–1214 (2004).
9. Royal Academy of Engineering. *Synthetic Biology: Public Dialogue on Synthetic Biology* (Royal Academy of Engineering, London, 2009).
10. Kaebnick, G.E. *Perspect. Biol. Med.* **50**, 572–584 (2007).
11. Streiffer, R. & Rubel, A. *Public Aff. Q.* **18**, 223–248 (2004).
12. Pollan, M. *The Botany of Desire: A Plant's Eye View of the World* (Random House, New York, 2001).
13. <http://www.scienceprogress.org/2008/11/synthetic-biology/>.

From synthetic biology to biohacking: are we prepared?

Gaymon Bennett, Nils Gilman, Anthony Stavrianakis & Paul Rabinow

The emergence of synthetic biology, and off-shoots such as DIYbio, make the need for a rigorous, sustained and mature approach for assessing, and preparing for, the broad range of associated dangers and risks all the more pressing.

In the past year, a spate of articles has reported on the growth and formalization of 'DIYbio'[1–7]. Alternately portrayed as techno-progressive, rogue and, above all, hip, this global cadre of DIYbio practitioners or biohackers is stylized as being capable of doing at home what just a few years ago was only possible in the most advanced university, government or industry laboratories[8]. The degree to which such capabilities have been, or can be, actualized remains an open and empirical question. What is clear is that the emergence of DIYbio and synthetic biology add urgency to the creation of a framework for systematically evaluating the risks and dangers of biological engineering. To proceed in that direction, more sustained reflections on the problems and objects at issue is a mandatory prerequisite.

Given the increasing ease and availability of biological engineering, the community needs to spend more time and effort in assessing and anticipating the dangers and risks associated with the technology.

DIYbio versus synthetic biology

The media attention surrounding DIYbio has served to brand the endeavor just as synthetic biology was branded. Both embrace the goals of making biology 'easy to engineer' and ensuring materials and know-how circulate in an 'open source' mode—"biology for the people" as the platitude has it. The association is not surprising or accidental. DIYbio and synthetic biology, after all, share institutional and personal connections. Leading research institutions, such as the National Science Foundation–funded Synthetic Biology Engineering Research Center, of which three of us (P.R., G.B. and A.S.) are a part, have made these two goals central to their strategic plans. Additionally, leading figures in synthetic biology have informally served as impresarios to some in the biohacker movement, notably through their sponsorship and promotion of the International Genetically Engineered Machines (iGEM) competition, which in 2009 has drawn over 100 teams of undergraduate bioengineers from five continents. In the light of the growth of DIYbio and the publicity that it has generated and received, however, the directors of the iGEM competition have banned DIYbio teams from participating in the competition.

The connections and convergences can no doubt be overstated. Self-definitions vary, and not all synthetic biologists would define their field as fostering "mechanisms for amateurs to increase their knowledge and skills," as a prominent DIYbio website (http://diybio.org/) puts it. Conversely, not all DIY biologists design "new biological parts, devices, and systems," as synthetic biology has sometimes been defined (http://syntheticbiology.org/). Nevertheless, it's certainly fair to say that accessible, easy-to-engineer biology is becoming the proverbial name of the game. Those synthetic biologists and DIYbio practitioners who object to being grouped together need to speak up in their own name.

The good news is that open access biology, to the extent that it works, may help actualize the long-promised biotechnical future: growth of green industry, production of cheaper drugs, development of new biofuels and the like. The bad news, however, is that making biological engineering easier and available to many more players also makes it less predictable, raising the specter of unknown dangers.

Biosecurity issues

A range of researchers and research institutions have raised the issue of biosecurity.

Gaymon Bennett, Anthony Stavrianakis & Paul Rabinow are at the Synthetic Biology Engineering Research Center and the Department of Anthropology, University of California at Berkeley, Berkeley, California, USA. Nils Gilman is at Monitor 360, San Francisco, California, USA.
e-mail: gaymon.bennett@berkeley.edu

Two different consortia of companies, for example, have proposed competing screening frameworks to deal with new capacities in biosynthesis technology[8]. The trouble with many of these responses, however, is that they take the increase in technical capacities, *per se*, to basically be the heart of the matter[9]. Ergo: technical solutions are proffered as adequate to technical problems. This technical approach is framed as 'dual use': there are good uses and bad uses, good users and bad users. Given this frame, a double challenge is posed: how to either design things biological, such that the 'bad guys' can't reengineer what the 'good-guys' have made, or set-up screening procedures so that the good guys can effectively keep the bad guys out. The goal: prevention through technical and organizational blockage.

> What is clear is that the conditions of the life sciences have changed dramatically since the 1975 conference on recombinant DNA at Asilomar.

Such responses no doubt have their place. The trouble is these responses don't actually address the problem at hand (leaving aside the intractable difficulty of discerning who is good and who isn't). The real concern in all of this is the fact that dangerous events, whether intentional or accidental, are facilitated through an increase in ease and access. Much to their credit, the authors of the widely circulated Sloan report[10] on synthetic biology and biosecurity have made this same point.

To the extent that more and more people in less and less formal and visible settings are able to engineer biological systems, the possibility of predicting the form and timing of such dangerous events, and thereby preventing them, becomes intractable. In certain respects, DIYbio is a 'black swan' waiting to happen: it portends events whose probability might seem low, but whose negative impact is likely to be quite high[11].

The insufficiency of current responses is reflected in, and reinforced by, a trend toward polemics. On one side, there are activists, such as those that form the Canadian nongovernmental organization, the Action Group on Erosion, Technology and Concentration, who evoke the 'precautionary principle'. These activists want to shut down all research programs for which the range of negative outcomes cannot be reliably determined in advance. Given that science, by definition as well as practice, is experimental in character, such a position amounts to shutting down bioengineering altogether, whether synthetic biology, DIYbio or otherwise[12].

On another side, there are enthusiasts, practitioners of synthetic biology, and biohackers. Enthusiasts subscribe to what has been called a 'proactionary principle', which invokes a 'right to innovate'[13]. Enthusiasts are often unwilling to frankly address dangers posed by easy-to-engineer and open-source biology. To the extent that such possible dangers are acknowledged, an attitude of 'trust us' pervades. Enthusiasts and entrepreneurs are willing to concede the need for some forms of indirect self-regulation. Policymakers, however, should leave it to the biologists to develop norms and protocols.

Polemics may not be strictly representative. They are, however, consequential. These polemics introduce a shell game in which the facts of the matter hide under one analogy after another, each coming in quick succession. Synthetic biology, activists say, is just like giant agribusiness. It's really all about ownership of nature, destruction of biodiversity and devastation of marginalized farming communities. Or, maybe it's Frankenstein that should worry us. Garage biologists will create designer organisms, fashioned to the maker's will. The implications are familiar: violated nature will reap its own revenge.

On the other side, enthusiasts use a Lego analogy: bioengineering will be made child's play; order your kit and get to work. Or, when addressing a more skeptical audience, the analogy becomes the computer industry: yesterday, we were building PCs in our garages; today, we have iPhones. Message: if you want your iPhone, put up with the potential for the equivalent of a few computer viruses. The only trouble is that the analogy between computer viruses and bioengineered viruses is not at all apt: computer viruses can't kill people, at least not directly.

What gets covered over by activists and enthusiasts alike is that the contemporary admixture of bioengineering and biosecurity forms a combination with distinctive and distinctively troublesome characteristics. It's safe to bet that synthetic biology and DIYbio will only intensify these characteristics. The point is that today, we've got a distinctive problem on our hands; attending to its particularities is a demand of the first order.

The taken-for-granted credibility of Asilomar-like self-governance, as some senior researchers have recognized, is highly contestable and does not stand up to critical evaluation. The organizers of the annual 'SynBio' conferences, for example, have invited scrutiny and criticism from nonengineers and nonbiologists. And the charter for the National Science Advisory Board allows the participation of nontechnical experts. Whether this is a mere gesture or will contribute to better governance remains to be seen.

What is clear is that the conditions of the life sciences have changed dramatically since the 1975 conference on recombinant DNA at Asilomar. The exclusion of the public is no longer even imaginable in the age of the internet. Gentleman's agreements of a kind that were common in 1975 are no longer imaginable, given the rise of patenting in the biotech industry. Assurances by patriarchs that safety issues can be handled through expertise and containment are no longer plausible given the global conditions of security. And so on.

It follows that, among other things, safety by design and screening technologies alone won't cut it. Technical capacities are increasing, to be sure. And these technical capacities need to be responded to at a technical level. But such a task, difficult and worthwhile as it may be, is only one vexing aspect of the current situation. The increase in technical capacities is just a first vector that makes the current problem distinctive and troublesome. Here are some others that warrant careful reflection:

- Moral arrogance. Many elite researchers and self-styled hackers tacitly concur that all technical advance is worthwhile and that only malicious people will do bad things. Arrogating moral goodness to the bioscientific side of the ledger overly simplifies a moral landscape that needs to be analyzed in all its complexity and contradiction.

- New actors and actions. The post-9/11 security environment is characterized by new actors and actions. For the past eight years, US citizens have had to face what most of the rest of the world has confronted daily for some time: that the difference between who and what is dangerous, and who and what is not, is a blurred and ever-shifting matter.

- Existing global access. Global capital and the internet have taken cutting-edge biology into laboratories around the globe. Even without DIYbio, bioengineers in countries all over the world have access

to materials and know-how. If Iran can fund developments in nuclear technology, they can certainly foot the bill for a few synthetic biology laboratories.

- Shifts in governance. For a decade now, national and multinational regulators and planners have increasingly been turning their attention to 'low-probability/high-impact' events rather than civil defense. Preparedness for such events, whether 9/11, Hurricane Katrina or H1N1 flu, seems to be the order of the day everywhere but in the laboratories and fraternities of advanced bioengineering.

All told, facilitating DIY capabilities for designing and constructing biological systems makes all of these factors even more difficult to deal with, to say the least.

Another approach

We argue that developments in synthetic biology and DIYbio call for another approach. Beyond the denunciation of the activists and the hype of enthusiasts, we need the vigilant pragmatism of what we have called 'human practices' (http://www.synberc.org/humanpractices). Such an approach consists of rigorous, sustained and mature analysis of, and preparation for, the range of dangers and risks catalyzed by synthetic biology and DIYbio. Preparedness activities might include on-the-ground tracking of the ramifications of synthetic biology research, or training in emergency response to biological events. Less familiar activities might include scenario development and stakeholder war-gaming (e.g., see http://www.gbn.com/; http://360.monitor.com/).

In the coming years, the intertwined growth of synthetic biology and DIYbio will further limit the scope of the current 'dual-use' framing of biological threat assessment and mitigation based on guarding key facilities, establishing export controls and monitoring technical experts. In its stead, policy makers will need to develop new analytic and policy frameworks, frameworks calibrated as much to preparation for unlikely but damaging events as to the design of technical safeguards[14].

We simply do not know the full extent of dangers on the near-future horizon, or of opportunities for that matter. We cannot be certain how biotechnological capacities will expand and ramify. We cannot be certain of the extent to which synthetic biologists and biohackers will successfully make biology easy to engineer or open source. We can be certain, however, that the stakes are high for everyone involved—above all for the enthusiasts. Those unwilling to prepare for dangerous events are exposing themselves, professionally and personally: if and when an untoward bio-event takes place, the so-called experts who failed to prepare will take the lion's share of collective blame. Studies, laboratories and careers are likely to be policed or even terminated.

The central challenge today is to neither shut things down, nor simply trust the experts. Rather, the challenge is to foster sustained and engaged inquiry that takes the pragmatic conditions of this techno-science and its practitioners as its primary object of concern, rather than hype or fear about an imagined future. Following this human practices mode, we might be able to anticipate and specify how to prepare regulations, normative frameworks and ethical responses adequate to the demands of the day.

AUTHOR CONTRIBUTIONS

All authors contributed equally to the work presented in this paper. P.R. directed the work of the Berkeley research team. N.G. conducted primary research on DIYbio. G.B. was the lead writer. A.S. was a secondary writer. All authors discussed and revised the manuscript at all stages.

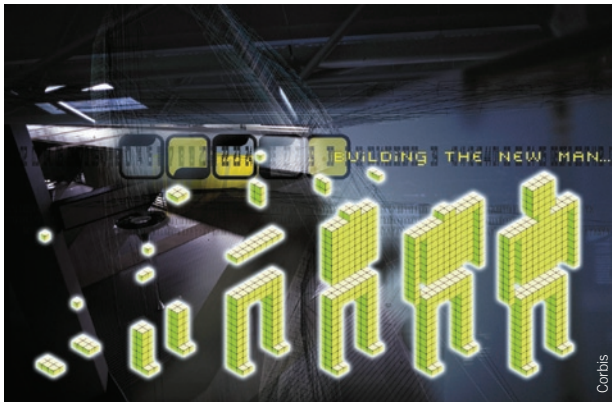
1. Johnson, C.Y. Hackers aim to make biology household practice. *Boston Globe* September 15, 2008.
2. Degasne, S. ADN: les biohackers créent des monstres dans leur garage. *Rue 89* November 3, 2008.
3. Anonymous. Biohackers attempt to unstitch the fabric of life. *Associated Press* December 27, 2008.
4. McKenna, P. The rise of the garage genome hackers. *New Scientist* January 7, 2009.
5. Bloom, J. The geneticist in the garage. *The Guardian* March 19, 2009.
6. O'Brien, D. Hacking into the building blocks of life itself. *Irish Times* April 10, 2009.
7. Whelan, J. In attics and closets, 'biohackers' discover their inner Frankenstein. *Wall Street Journal* May 12, 2009.
8. Hayden, E.C. *Nature* **461**, 22 (2009).
9. Nair, N. *Nat. Med.* **15**, 230–231 (2009).
10. Garfinkel, M.S., Endy, D., Epstein, G.L. & Friedman, R.M. *Synthetic Genomics: Options for Governance* (J. Craig Venter Institute, The Center for Strategic and International Studies, Massachusetts Institute of Technology, October 2007). http://www.jcvi.org/cms/fileadmin/site/research/projects/synthetic-genomics-report/synthetic-genomics-report.pdf
11. Taleb, N.N. *The Black Swan: The Impact of the Highly Improbable* (Random House, New York, 2007).
12. http://etcgroup.org/upload/publication/602/01/synbioreportweb.pdf
13. Schmidt, M. *et al. Syst. Synth. Biol.* **2**, 7–17 (2008).
14. Ostfield, M.L. Biosecurity Bioterrorism: Biodefense Strategy, *Practice Sci.* **6**, 261–267 (2008).

Engineering a new business

Mike May

As the market for DNA on demand continues to grow, increases in the scale and efficiency of new genome engineering approaches promise to accelerate product discovery and even open up new commercial opportunities.

A congruence of innovation in the fields of microfluidics, miniaturization, automation and DNA synthesis, assembly and sequencing promises to provide new capabilities to companies focused on engineering innovative new products for pharmaceuticals, bioenergy, agriculture and beyond. At the same time, the nascent approaches underlying this technology still pose significant challenges in terms of reduction to practice, regulatory concerns and public perception.



Building blocks. Synthetic biology means different things to different people, but designing new biological parts and systems for useful purposes captures the essence.

Three broad classes of companies are emerging. First, many companies are making DNA parts for sale as reagents to academia and industry. The majority of these companies manufacture synthetic oligonucleotides (or oligos), but some are specializing in larger assemblies, even complete synthetic genes. In recent years, the synthetic oligo market has continued to grow. For example, one of the biggest synthetic oligo companies (**Table 1**), GenScript (Piscataway, NJ, USA), sold nearly twice the number of base pairs this year compared to last, according to Sally Wang, executive vice president. Demand is expected to increase for longer stretches of oligos with lower numbers of base errors. At the same time, the cost per DNA base is likely to keep dropping; as a result, synthetic oligos seem on their way to commoditization, and some companies are already selling oligos to fund other types of work.

A second group of companies is exploiting synthetic biology to advance processes that were previously performed with genetic engineering or metabolic engineering (**Table 2**). For example, an enzyme maker can now use computational approaches plus gene synthesis to design more effective compounds. So instead of arduously searching through thousands of enzymes to perform a task, the company can zero in on the best ones faster. Even with today's synthetic biology abilities, however, those designer enzymes still require fine tuning through traditional wet-lab techniques.

A last group, including some multinational biotechs and pharmas are now exploring advanced biological engineering approaches in their own R&D work or to sell products that can be used by others in the field (**Table 3**). To gain the needed expertise, these large companies often develop collaborations with smaller, innovative biotechs that specialize in cutting-edge approaches.

Thus, the products of synthetic biology seem poised for broader application. But for commercialization to succeed, business models must be found that are sustainable (**Box 1**) and industry and academia must address tough sociological, dual-use—peaceful and military—and safety issues that surround dissemination of the technology (**Box 2**).

What's new?

Synthetic biology is not so much a new field, as an evolving one. Previous capabilities in genetic and metabolic engineering paved the way for synthetic biology. As John Mulligan, CSO at Blue Heron (Bothell, WA, USA), puts it, "Synthetic biology is used to cover a wide range of modern manipulative molecular biology experiments, making the definition a bit problematic." In his view, the goal of synthetic biology "is to develop molecular and computational tools that will allow biologists to design systems, implement them using standard parts, and achieve predictable results."

Some researchers expect synthetic biology to deliver a greater level of control over genomes as well as provide tools for carrying out genetic manipulation at a scale and efficiency that is unprecedented. According to George Church, professor of genetics at Harvard Medical School (Boston), within a few years, synthetic biology is likely to provide the "ability to make essentially any genome and have it behave in a manner consistent with computer aided–design tools." This is not just about synthesizing a stretch of DNA, but about making it fully functional in a living cell. "We can already make about anything we want," Church says. The process, though, is not always efficient. So researchers need better algorithms to design sequences and better ways to make what they want.

As the applications of synthetic biology expand, so too will the overall market. A June 2009 BCC Research report[1] defines the field as "enabling technologies that are critical for synthetic biology (e.g., DNA synthesis or DNA sequencing); synthetic biological components (e.g., synthetic genes, synthetic functional DNA constructs and synthetic parts); integrated systems (e.g., synthetic chromosomes, genomes, cells and organisms); and products enabled by synthetic biology tools (e.g., pharmaceuticals, biofuels and chemicals)." Within that frame-

Mike May is a freelance writer based in Houston.

work, they find that this field as a whole created a $233.8 million market in 2008 (**Fig. 1**). But that's just a start; they extrapolate that the market for synthetic biology components and enabled products will reach $2.4 billion by 2013, which requires an annual increase of 59.8%. For now, chemicals and energy make up the leading market segment, accounting for $80.6 million in 2008. Biotech and pharmaceuticals came in a close second at $80.3 million, but this segment is expected to grow to $594 million by 2013 (ref. 1).

If the synthetic biology market is to reach such levels by 2013, John Bergin, author of the BCC Research report, points out that several things are needed, including a continued decrease in the cost of synthesizing DNA. Bergin says that the increasing availability of gene sequencing creates more and larger electronic gene databases. This drives demand for protein-expression systems, directed evolution and metabolic engineering, which creates demand for synthetic biology technologies and tools. In short, Bergin expects synthetic DNA to form a foundation for the future of this field. As Bergin notes, "Demand for synthetic genes is already robust and growing due to their utility in genomics."

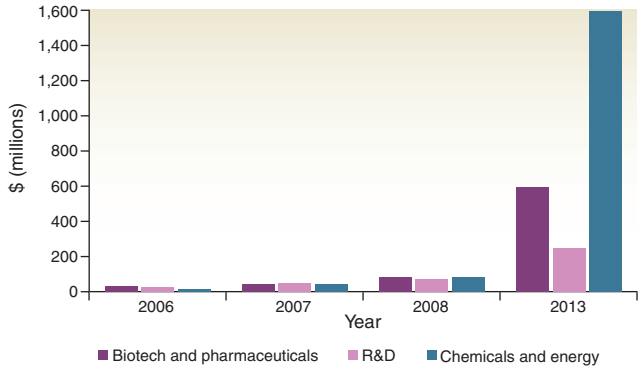


Figure 1 Actual and projected global market in synthetic biology. The projected figures are based on interviews with academic and industry leaders and government, industry and trade publications. Source: BCC Research.

Oligos to order

Roughly a couple dozen companies around the world make synthetic oligos. Those companies, however, range from ones that focus just on churning out custom oligos to others that make complete genes, as well as some companies that do a bit of both. Peer Staehler, president of febit synbio (Heidelberg, Germany), says that this market is worth $75–150 million a year, and he adds that it is growing rapidly. The field is also

Table 1 Selected oligo and synthetic gene suppliers

Company (year founded) Location Website	Employees	Oligo or gene size	Company description
Alpha DNA (1997) Montreal http://www.alphadna.com/	5–10	<180 bases	Custom oligos, and mutagenesis services, plus catalog of reagents and kits, primers, and recombinant proteins
Ana-Gen (1994) Duluth, GA, USA http://www.ana-gen.com/	5–9	7–25 bp	Designs, synthesizes and purifies oligos needed for a complete gene and mutagenesis services, plus catalog of reagents and kits, primers, and recombinant proteins
ATG Biosynthetics (2001) Freiburg, Germany http://www.atg-biosynthetics.com/	6	<100 bases (*n* x10) kb pairs	Contract gene synthesis. Consultancy. 'EvoMag' program for codon optimization reverse genetic engineering and synthetic biology applications.
Biolegio (1996) Nijmegen, The Netherlands http://www/biolegio.com/	20	Up to 210 bases	Specializes in long oligos and offers a broad range of modifications and dyes
BioNexus (1999) Oakland, CA, USA http://www.bionexus.net/	25	Up to 140 bases	Gene synthesis and other genomics-related products and services, various modifications to synthesized DNA, fluorescent dyes for labeling DNA oligos, phosphorothioate and mixed base oligos
Biosearch Technologies (1993) Novato, CA, USA, http://www.biosearchtech.com/	90	Up to 120 bases	Oligos for real-time qPCR and molecular diagnostics. Supplies the reagents and modifications needed to synthesize oligos. Engineers its own DNA synthesis instruments
Bioserve Biotechnologies (1990) Beltsville, MD, USA http://www.bioserve.com/	68	Up to 125 bases	Genotyping, DNA and RNA extractions from tissues, maintains biorepository of normal and diseased tissues from 120,000 individuals from several countries
Biosynthesis (1984) North Dallas, TX, USA	70	Up to 2–3 kb pairs	Custom and PCR oligos, custom peptides, RNA, polyclonal antibodies, organic synthesis
Blue Heron (1999) (products distributed by Invitrogen, Carlsbad, CA, USA) http://www.blueheronbio.com/	<50	Up to 52 kb pairs	Gene synthesis using proprietary GeneMaker technology generates synthetic structures that range in size from 60 to tens of thousands of base pairs
CyberGene (1995) Stockholm http://www.cybergene.se	4	Up to 80 bases	Manufactures quantitative PCR kits for prenatal diagnostics which are registered in the EU with CE mark such as ChromoQuant, for prenatal diagnosis
DNA2.0 (2003) https://www.dna20.com/	25–50	Up to 35 kb pairs	Gene synthesis, gene design assistance, expression optimization and protein engineering
Epoch Biolabs (2001) Sugar Land, TX, USA http://www.epochbiolabs.com/	35	5–100 bp/up to 50 kb pairs for genes	Gene synthesis and molecular services, variant library construction, protein expression and purification, DNA sequencing, SNP analysis

(continued)

Table 1 Selected oligo and synthetic gene suppliers (continued)

Company (year founded) Location Website	Employees	Oligo or gene size	Company description
Eurofins MWG Operon (1990) Ebersberg, Germany http://www.eurofinsdna.com/	250	5–20 bases up to 3 kb pairs	DNA sequencing, oligos, siRNA and gene synthesis
febit synbio (2005) http://www.febit.com/	90	Up to 3.5 kb bases	Synthetic genes produced from oligos based on microarrays in a 60-mer format; developing a new production platform called "MegaCloner," which will be used to offer building blocks that will be 40–400 bp in size
Geneart (2006) Regensburg, Germany http://www.geneart.com/	190	Up to ~20 kb pairs	DNA engineering and processing; produces optimized synthetic genes, generates gene variants and gene libraries, and produces DNA-based active agents
Gene Link (1993) http://www.genelink.com/	16	Up to 260 bases	Synthetic DNA, RNA, siRNA and antisense oligos; ultra-modified oligos with modifications in backbone, bases and fluorescent dyes
GeneScript (2002) Piscataway, NJ, USA http://www.genscript.com/	600	N.D.	Custom gene and oligo synthesis, bio-assays, Optimum gene proprietary codon optimization software
GeneWorks (1996) Adelaide, Australia http://www.geneworks.com.au	N.D.	5–100 bases	Custom oligos
Integrated DNA Technologies (1987) Coralville, IA, USA http://www.idtdna.com/	500	20 bases	Custom oligos
The Midlands Certified Reagent Company (1974) Midland, TX, USA http://www.oligos.com/	13	DNA 3-180, RNA 3-65 locked nucleic acids (LNA) 3-165, genes in 40-mers, any number	DNA, RNA, peptide nucleic acid synthesis, 75 polymers. More than 300 modifications that are commercially available, homegrown or out-licensed modifications, all LNA oligos
Trilink Biotechnologies (1996) San Diego http://www.trilinkbiotech.com/	85	Up to 180 bases	Modified nucleic acid, highly modified and mid-scale oligos, modified dNTPs

N.D., not disclosed. Source: websites and company press releases.

becoming more competitive. In general, all of the companies rely on the same basic chemistry for synthesizing oligos. In fact, Ali A. Javed, director of R&D at Gene Link (Hawthorne, NY, USA), says, "This market has matured so much that the innovation is reduced. Products are at a commodity level, a disposable-product level." Consequently, companies in this market must find ways to distinguish themselves.

Although no commercial maker of synthetic oligos will say just how they do it, they all follow the same general process. A customer sends in a desired sequence, maybe just a number from GenBank or a computer-designed, completely novel stretch of nucleotides. The oligo maker then screens the DNA sequence against databases to identify sequences that might code for toxins or other problematic agents. In the United States, for example, the Centers for Disease Control and Prevention (Atlanta) and Animal and Plant Health Inspection Service (Riverdale, MD, USA) maintain the National Select Agents Registry, which lists dangerous toxins and biological agents that can be obtained only by registered users. So if an ordered sequence of DNA encodes a biological agent or toxin on this list, that could be made only for someone or a facility registered for that agent. In addition, there are some agreements among countries that attempt to prevent the misuse of this technology. For example, the Australia Group (http://www.australiagroup.net/en/index.html)—an informal trade group that seeks to limit the proliferation of chemical and biological weapons—now includes guidelines about exporting oligos that code for toxins. The ultimate control for now, however, lies with the oligo makers, who try to determine an oligo's legality or potential for danger.

Once the oligo maker decides to move forward with an order, the company turns to its own design process, which includes various elements—determining how to break a large sequence into pieces for manufacturing, and picking the methods to make and assemble the fragments—all aimed at optimizing the process in production and cost. As design turns to manufacturing, other processes must be added, including error removal. Chemical synthesis of oligos might produce sequences with error rates of 1 in 300 base pairs, but for some applications, such as for which the product must be nearly perfect with ≥1 error in 10 million base pairs, this would be unacceptably high. Most companies rely on software and purification techniques—typically all proprietary—to reduce the error rate of completed sequences.

Some companies are going beyond the usual methodology and developing newer, faster platforms for synthesizing oligos. In Germany, febit synbio is developing a microarray-based process that will synthesize large numbers of oligos in parallel. Staehler says, "Many teams have failed to extract good DNA from microarrays, but we have teamed up with labs around the world and have shown—at least in proof of principle—that you can produce DNA at an incredible speed and complexity." He adds that incorporating miniaturization and parallelism makes the difference.

For others, oligos are capital generators. DNA2.0 (Menlo Park, CA, USA) started out in 2003 using its computational power to engineer proteins, but it was unable to raise any venture capital. Instead, it started selling the custom genes made with the same technology that it was using to improve proteins. "We've watched our synthetic-gene market go up by tenfold in the past six years," says Jeremy Minshull, DNA 2.0's president. So even without any startup funding, this company turned a profit in its first 18 months of operation.

As more companies enter this field, each looks for ways to get an edge. For example, Mulligan of Blue Heron says, "We focus on a fully automated process. So we use protocols that are easier to do on robots." The Blue Heron robots include off-the-shelf ones and a few that the company designed and built. This company's technology also allows a wide range of oligo lengths. "We've

Table 2 Selected companies with R&D that incorporates advanced engineering approaches

Company (year founded) Location Website	Employees	Company description	Products	Funding source
Agrivida (2003) Cambridge, MA, USA http://www.agrivida.com/	32	Agbiotech company developing crops to produce chemicals, fuels and bioproducts from non-food cellulosic biomass. Enables the delivery of low-cost sugars for the production of a wide variety of industrial biotech products	None	Series B funding in 2009, led by DAG Ventures
Amyris (2003) Emeryville, CA, USA http://www.amyris.com/	200	Renewable products company focused on the production and use of renewable chemicals and transportation fuels. Combines technology, production and distribution to commercialize and scale products across the supply chain through its Brazilian subsidiary, Amyris do Brasil Pesquisa e Desenvolvimento Biocombustiveis. Building distribution capabilities, through its US subsidiary Amyris Fuels	None	Private funding including venture capital
Biodesic (2005) Seattle http://www.biodesic.com/	2	Provides technologies and knowledge to transform business and society through the development and distribution of biological technologies. Developing novel technologies, such as ProDNA, a system for protein measurement that is expected to be as sensitive and accurate as the existing methods for RNA and DNA	None	Bootstrapped and now internally funded through consulting
Biotica (1996) Cambridge, UK http://www.biotica.com/	23	Drug discovery and developer, using its polyketide engineering platform. Has a library of naturally occurring polyketides, which are optimizable using its technology platform	None	Venture capital plus collaboration license deals
Codexis (2002) http://www.codexis.com/	300	Clean technology company that develops industrial biocatalysts, including enzymes and microbes, for use in the energy industry to enable next generation, non-food biofuels and for cost-effective manufacturing of human therapeutics. Develops biocatalytic processes that can reduce manufacturing costs across a broad range of industries	Markets enzyme products and technology to pharmaceutical companies including Merck, Pfizer and Teva	Privately held with funding from corporate and venture investors
Ginkgo BioWorks (2008) Boston http://www.ginkgobioworks.com/	6	Instrument and consulting company, focused on making biology easier to engineer. Commercializing a suite of proprietary DNA assembly technologies intended to simplify the rapid construction of metabolic pathways and gene networks	BioBrick Assembly Kit (co-developed with New England Biolabs), which includes the reagents needed to assemble BioBrick standard biological parts	Started with seed funding, including an SBIR grant, grant from the city of Boston, and now working off revenue and consulting fees
Genomatica (2000) http://www.genomatica.com/	35–40	Chemical company that commercializes novel biomanufacturing processes to produce a variety of industrial chemicals for all major industries. Had a proprietary integrated bioprocess engineering platform and SimPheny, a metabolic modeling and simulation system	None	Privately held and backed by Mohr Davidow Ventures, Alloy Ventures and Draper Fisher Jurvetson
Global Bioenergies (2008) Evry, France http://www.global-bioenergies.com/	13	Renewable products company that transforms renewable resources into hydrocarbons, targeting fuels, plastics and rubbers, using classical or proprietary synthetic biology technologies	None	Venture capital
Metabolix (1992) http://www.metabolix.com/	107	Bioscience company focused on providing sustainable solutions for manufacturing plastics, chemicals and energy, using a systems approach, from gene to end product, integrating biotech with advanced industrial practice. Has a proprietary platform technology for biobased, biodegradable plastics from corn for many market applications	Mirel Bioplastics	Publicly traded on the NASDAQ under MBLX
Synthetic Genomics (2005) La Jolla, CA, USA http://www.syntheticgenomics.com/	~100	Synthetic biology company that develops and commercializes genomic-driven advances related to energy, chemicals and high-value agricultural products. Designing next generation fuels and biochemicals from carbon dioxide, plant biomass and coal, developing a biological solution to increase production or recovery of subsurface hydrocarbons, high yielding and disease resistant feedstocks	None	Privately held company that, in 2007, closed its Series B round of financing with BP and the Asiatic Centre for Genome Technology
Verdezyne (2005) http://www.verdezyne.com/	25	Industrial biotech company that uses a combinatorial approach to designing and engineering enzymes, metabolic pathways and microorganisms that produce target chemicals. Has a patented process for the design and synthesis of self-assembling genes directly from commercial oligos	None	Venture capital

Source: Company websites and press releases. SBIR, small business innovation research.

Table 3 Selected large corporations exploring advanced engineering R&D approaches

Name Location Website	General description	Selected synthetic biology projects
Bayer CropScience Monheim am Rhein, Germany http://www.bayercropscience.com/	Crop science company focusing on crop protection, nonagricultural pest control, seeds and plant biotech. It has a global workforce of more than 18,000, and it is represented in more than 120 countries	Entered a technology alliance with Chromatin to apply mini-chromosome technology for crop improvement
ExxonMobil Irving, TX, USA http://www.exxonmobil.com/	Largest publicly traded international oil and gas company	Entered a multi-year research and development agreement with Synthetic Genomics to develop next-generation biofuels using photosynthetic algae
Merck http://www.merck.com/	This global research-driven pharmaceutical company was established in 1891, and it employs more than 55,000 people. Merck discovers, develops, manufactures and markets a wide range of vaccines and medicines	Formed an ongoing collaboration with Codexis to incorporate synthetic approaches to biocatalysis, which can be used in pharmaceutical basic research and manufacturing
Monsanto St. Louis http://www.monsanto.com/	An agricultural company that focuses on the application of modern biology to seeds, especially ones with incorporated technology, such as pest resistance. This company also makes herbicides, including Roundup	Works with Protabit (Pasadena, CA, USA), which developed a computational–protein design platform. Through this collaboration, Monsanto hopes to develop new traits for crops
Pioneer Hi-Bred Johnston, IA, USA http://www.pioneer.com/	This DuPont business develops advanced plant genetics to increase productivity, profitability and develop sustainable agricultural systems. Pioneer provides services to customers in nearly 70 countries	Collaborating with Arzeda, which can develop new enzymes *de novo*. Pioneer Hi-Bred plans to use these enzymes as starting points for its own technologies, including directed evolution
Syngenta Basel http://www.syngenta.com/	In 2000, Novartis and AstraZeneca merged their agribusinesses to form Syngenta, which focuses on two main types of products: seeds and crop protection	Licensed mini-chromosome technology from Chromatin to improve the traits of corn, and is now working on sugarcane

Source: Company websites and press releases.

had orders as small as 60 base pairs," Mulligan says, "and our largest product was 52 kb, which took several levels of assembly." But Blue Heron could go even higher, at least to a couple hundred kilobases, the company claims.

Like any business, synthetic oligos must be economical to survive, let alone grow. Some biotechs and pharmas—even academic labs—already outsource oligo synthesis as it gets more cost effective. However, many universities still maintain core facilities to serve their faculty's needs for oligos, although there are fewer such cores than there were a few years ago. According to Anthony Yeung, an officer with the Association of Biomolecular Resource Facilities (Bethesda, MD, USA), the number of core facilities offering oligo synthesis as a service today has dropped to roughly half the number that existed in 2005, whereas the number expressing an interest in DNA synthesis actually increased by 20%, "asserting the continued importance of the technology to core facilities," he says. The core facilities also report an increase in volume, which in some cases leads to outsourcing where volume and pricing are favorable. But in-house synthesis is still in demand when confidentiality and local expertise are needed.

In the commercial sector, BCC's Bergin sees the companies with more advanced technologies having the best prospects long term, although for simple oligos, price and delivery remain key. "Companies offering downstream products like synthetic genes or other biological parts and who have their own in-house quality oligos supply capability, or a strong oligos supply partner, will be in the best position moving forward," he says. Some customers come to commercial makers to get more-complex jobs done quickly. For example, a customer might want to try a dozen substitutions at 50 positions in an antibody. In such cases, says Mulligan, "A commercial maker can be two to three times faster and at a fraction of the cost of doing it in your lab." He adds, "The business is growing because the prices are coming down." Those companies that can handle a variety of orders and reduce the oligo failure rate (through in-house production of high-quality oligos) will come to the forefront of this business, Bergin predicts.

As the customer base grows, so do the capabilities of the industry. Blue Heron plans to expand its capacity by tenfold in the next 12–18 months. "And our staffing will stay about the same," says Mulligan, "as we add capacity with increased automation."

Despite the growth in commercial oligos, some in the field envision even more improvements in the future, especially in terms of length. "We're still dependent on relatively expensive synthesis," says genomics innovator J. Craig Venter. Less-expensive synthesis along with

Box 1 Flash in the business plan?

In March, startup Codon Devices (Cambridge, MA, USA)—the company that blazed the commercial trail for synthetic biology and whose scientific advisory board read like a *Who's Who* for the field—announced it was closing its doors, just five years after its founding. In fact, the diversity of interests and approaches embodied in the founders—George Church of Harvard Medical School, bioengineer Drew Endy now at Stanford University, physicist Joseph Jacobson of the Media Lab at the Massachusetts Institute of Technology and chemical engineer Jay Keasling at the University of California, Berkeley—may have been part of the problem. At the time of the company's closing, experts and analysts pointed to the difficulty of trying to do too many things at once as the likely culprit. Moreover, the leaders at Codon Devices seemingly reached a similar conclusion. Less than a year before going out of business, Codon Devices abandoned its synthetic-oligo side to concentrate on developing applications. At that time, the change in business strategy and a $31 million infusion of funds from its investors, which included Khosla Ventures (Menlo Park, CA, USA) and Alloy Ventures (Palo Alto, CA, USA), looked sufficient to keep Codon Devices afloat. But just one year later, the board closed it down. After the closing, Church told *Nature*[15] that the company should have stuck with applications and forgone synthetic oligos. So like any other fledging field of research, even a stellar conjunction of capital and science is no guarantee of commercial success.

Box 2 Can there be safety in synthesis?

As companies succeed in making synthetic oligos in the 50-kb range, they reach the size of many viruses listed on the US National Select Agents Registry, which regulates the use of toxins and biological agents. Once companies can readily make synthetic oligos in the 200-kb range, that will cover every virus on that list. "As we venture into assembling whole bacterial genomes," says Blue Heron's CSO John Mulligan, "the concerns grow over the possibility that this technology will allow access to pathogens that wouldn't otherwise be available to people with malicious intent."

Synthetic-gene companies are working together to standardize a process for screening potentially dangerous agents. For example, febit synbio and several other companies joined forces as the International Association of Synthetic Biology. "We wanted to start working on a framework for governments and regulatory groups—something that shows what to do and what not to do," says febit synbio's Staehler. This group alerts companies about potential risks and distributes information about best practices for screening synthetic oligos. Staehler says, "We are starting to interact with the FBI in the US and several government authorities in Germany." Some believe that self-regulation is sufficient. For instance, Paula Olsiewski, program director at the Alfred P. Sloan Foundation (New York), says, "I applaud the industry for the good work they are doing."

But the difficulty comes in identifying every potentially dangerous sequence. That would require an inclusive, constantly updated list. Another problem is that one can create a dangerous agent starting with a set of short oligos, ordered from different companies, according to John Dileo, lead scientist at the MITRE Corporation (Bedford, MA, USA), a not-for-profit technology company that supports the US government. To make it harder to accomplish such a task, Dileo and James Diggans, group leader for computational biology at MITRE, developed the DNA order tracking system (DOTS). This software would gather oligo orders from companies to see if any sequences could be combined to make something illegal or dangerous. "Long genes can be screened relatively easily," says Diggans. "The harder part comes with short oligos."

So far, DOTS works in simulated runs at MITRE. To work in the real world, though, all synthetic-oligo companies would have to submit each order they receive to a general database. But Diggans says, "There is a lot of concern about the centralization of orders, because of confidentiality with customers." As a next step, MITRE will try out their software in field tests.

Safety concerns are universal. Bärbel Friedrich, a microbiologist at Humboldt University (Berlin), and her colleagues from several other German organizations developed a position paper about the opportunities and risks behind synthetic biology (http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/2009/download/stellungnahme_synthetische_biologie.pdf). In this work, Friedrich distinguishes biosafety from biosecurity issues. She believes that existing regulations handle much of the biosafety concerns, but due to the rapid advancement in the field, there needs to be a monitoring system. "We also need research on the impact of artificial cells, novel biomaterials and so on," she says. For biosecurity, she advocates that the synthesis of DNA sequences be kept safe by using a general database for identifying dangerous sequences and following a standardized commercial procedure. Enforcing such regulations, however, may not be so easy. "How to do this worldwide is a problem," she says.

other technological advances, however, will spawn the use of even longer sequences. That, too, brings new challenges. For example, as DNA gets longer, it gets more brittle. So scientists must develop ways to handle these longer stretches of bases.

In addition, future technology could do a better job even with shorter oligos. As Javed of Gene Link says, "We can endlessly design an oligo to perform better." In addition to adjusting the codons for a particular amino acid, he'd like to see more nucleotides to consider. He says, "There should be an arsenal of modifications—like 16 bases instead of just 4—for customers to chose from, and it should not be inhibitory because it is so expensive."

Chromosomes on demand?

Although synthesizing oligos and genes is familiar territory for biotech, a radical new goal, pioneered by J. Craig Venter, Hamilton Smith and Clyde Hutchison and their colleagues at the J. Craig Venter Institute (Rockville, MD, USA), is to synthesize an entire chromosome from scratch and then reboot it in a recipient cell chassis. The idea is that in the context of an artificial, controlled environment, a 'chassis' organism with a minimal genome would be capable of devoting many more resources to a synthetic product pathway of interest, enabling higher yields.

Several steps in this project have already been attained. In 2003, the Venter team assembled their first complete genome—that of the bacteriophage ϕX174—by stitching together short oligos using an adaptation of PCR[2]. Four years later, they provided the first demonstration of genome transplantation using native donor DNA from *Mycoplasma mycoides* to reprogram a related species *Mycoplasma capricolum*[3]. The researchers have since successfully cloned a complete synthetic bacterial (*Mycoplasma genitalium*) chromosome in a yeast cell (*Saccharomyces cerevisiae*)[4]. In their latest work, after cloning a native *M. mycoides* genome in yeast, through the addition of a yeast centromere to the bacterial genome, the researchers showed that treatment of donor DNA with specific methylase from the donor bacterium allows successful transplantation back into a different bacterium (*Mycoplasma capricolum*), whose genome had been removed[5]. This work thus moved a genome from a prokaryote to a eukaryote and back. When asked how significant this feat is, Venter says, "It depends in part on how extendable it is to other types of bacteria." Then he adds, "We have no reason to believe that it won't be." Nonetheless, Venter thinks that the yeast could create a roadblock. As he says, "There may be a limit of what can go in yeast, but we don't know the limit."

Steps toward using larger collections of DNA, however, are already underway. In 2007, James Birchler and his colleagues at the University of Missouri (Columbia, MO, USA) described a method for making maize mini-chromosomes—a centromere with telomeres—to which they added genes (**Fig. 2**) (ref. 6). As Birchler explains, "We start with an endogenous centromere, and then we can add onto it whatever we want." What can be added, however, is limited by the amount of DNA—about 35–40 kb—that can be injected into a maize cell in one transformation. Birchler hopes to soon be able to repeatedly add pieces of DNA into a cell, one mini-chromosome at a time, thereby allowing the incorporation of more genes. Birchler adds, "Depending on the nature of what is added and the purpose, one could use endogenous promoters or engineer the genes to be under the control of promoters that would be coordinately expressed. This of course is still in the future."

Such a process could improve corn by adding

the genes for drought resistance or for nitrogen utilization, complex traits that require multiple genes. With corn, it is possible to add one gene to a maternal lineage and one to a paternal lineage, and then cross them to make a hybrid that includes both genes. But mini-chromosome technology bypasses tedious and time-consuming crosses, in adding multiple genes at once.

Chromatin (Chicago), is already producing plant mini-chromosomes of up to 200 kb (ref. 7), but even larger mini-chromosomes are also feasible, according to Daphne Preuss, founder and CEO. Chromatin has licensed its plant mini-chromosomes to several companies, including Syngenta (Basel) for transforming sugarcane. Sugarcane is grown commercially as a vegetative crop, which means that it gets propagated through cuttings. So, as Preuss explains, "It's not practical to add one gene to one sugarcane plant and another gene to a different plant and then cross them to get both genes in one plant like you can with corn," Preuss explains. "To get multiple genes in sugarcane you want to do it all at once."

Artificial chromosomes have also been produced in animal systems. At Hematech (Sioux Falls, SD, USA), researchers combined fragments from human chromosomes 2, 14 and 22 to make an artificial chromosome, which is essentially a vector that includes the full repertoire of human antibody genes, according to company president Eddie Sullivan. Hematech scientists use this human artificial chromosome to create transchromic cattle, which serve as human antibody production systems[8]. The size of cows alone makes them a good factory. "You can collect up to 60 liters of plasma per month from an adult animal," Sullivan says. With a human antibody–producing cow, Hematech can expose the animal to, say, a human infectious disease, or maybe even cancer cells, and those antigens could produce specific antibodies in the cow. The company is in early product development and has already done some preclinical testing in the biodefense area.

Souped-up engineering?

Traditional approaches to metabolic engineering still dominate work under way in industry. For example, Archer Daniels Midland (Decatur, IL, USA) and Metabolix (Cambridge, MA, USA) will use metabolic engineering in the technology behind a plant being built in Clinton, Iowa, where starch from corn will fuel engineered bacteria to generate natural versions of polyhydroxyalkanoate (PHA), which are traditionally petroleum-based plastics. This plant should begin operating by the end of this year.

Among the most ambitious metabolic engineering attempts, artemisinin—a component of an antimalarial remedy—remains the poster child. Traditionally, pharmaceutical scientists extract this drug from the Asian plant sweet wormwood, a process that is affected by the vagaries of weather and drought, and which costs too much to serve many populations most affected by malaria. In 2006, chemical engineer Jay Keasling and organic chemist Richmond Sarpong, (both of the University of California, Berkeley) reported the engineering of a complete biosynthetic pathway for making artemisinin in yeast (**Fig. 3**)(ref. 69). To turn this technology into a product, Keasling founded Amyris Biotechnologies (Emeryville, CA, USA). In 2004, the Bill & Melinda Gates Foundation provided a $42.6 million grant to the nonprofit pharmaceutical company Institute for OneWorld Health (San Francisco), which helped scale-up the manufacturing process for biosynthetic artemisinin. In 2008, through a license agreement with Sanofi-Aventis (Paris), the company built a plant to make this drug. According to Keasling, this should lead to commercially available biosynthetic artemisinin in the next couple years.

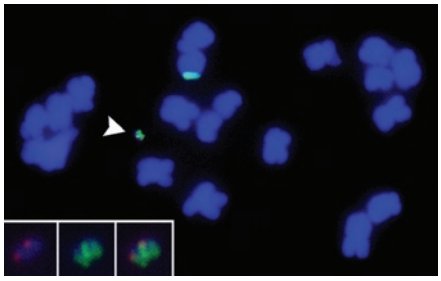


Figure 2 Artificial chromosomes. The arrow points to a mini-chromosome (green is centromere-specific probe; red, the transgene; and blue, DAPI-stained DNA. (Image provided by Jim Birchler, University of Missouri).

Other companies are applying existing approaches, such as directed evolution, to drug manufacturing. In 2006, Codexis (Redwood City, CA, USA) used directed evolution of three biocatalysts to improve the production of atorvastatin, the active ingredient in Pfizer's cholesterol-lowering drug Lipitor. According to Codexis, this technology generated a 4,000-fold improvement in the productivity of one reaction in this drug-making process.

Pfizer is not the only adopter of Codexis's platform. In 2007, Merck (Whitehouse Station, NJ, USA) started a collaboration with the company to produce biocatalysts. "In the pharmaceutical business," says Greg Hughes, an associate director at Merck, "biocatalysis can help minimize the environmental impact of manufacturing processes." Hughes would not divulge specifics about any ongoing projects, but said, "We look to apply biocatalysis from basic research to manufacturing."

However, many of the early adopters of directed evolution techniques have had disappointing results. According to Eric Schmidt, a biotech and healthcare analyst at New York-based Cowen & Co., "I would say that directed evolution has not met with much, if any, success. Companies like Maxygen (Redwood City, CA, USA) and Advanced Molecular Evolution (AME) have not panned out as hoped." (In October, Maxygen restructured into a joint venture with Astellas Pharma (Tokyo) after experiencing a capital crunch; AME was bought by Eli Lilly (Indianapolis) in 2004).

Not all the experience has been negative, however. For example, since purchasing AME, Lilly claims to use AME's directed evolution approaches to design and engineer new biologics in a variety of programs, for autoimmune diseases, diabetes and cancer. Currently, 8 of the ~60 molecules in Lilly's clinical pipeline and 4 in preclinical development involved work from AME, according to company spokesperson Judy Kay Moore. What's more, one of the reasons Merck turned to Codexis was because of its capacity to use a variety of genetic tools, including directed evolution through DNA shuffling, to increase enzyme efficiency. In fact, Codexis looks for ways to improve the efficiency of entire pathways. "Overall, this technology works so well," explains David Anton, senior vice president of R&D at Codexis, "because we can get improved enzymes in a few weeks rather than months. This triggers fast progress."

The next generation?

It is arguable whether any of the approaches used by Codexis, Metabolix and AME in the above applications represent the type of technological leap in engineering that might be possible if gene circuit design *in silico*, DNA synthesis, assembly and sequencing at the genome scale all become routine parts of product development. A key aspect of making this leap will be our ability to create effective synthetic regulatory mechanisms for increasingly complex, multigenic systems. A pilot project undertaken by Kristala Jones-Prather, a chemical engineer at the Massachusetts Institute of Technology (Cambridge, MA, USA), focuses on developing a strain of bacteria that can produce glucaric acid. Three genes—one each from bacteria, mouse and yeast—are needed to create the pathway in *Escherichia coli*. But initially, when the three enzymes were expressed in the bacterium, glucaric acid yields were limited by differences in the catalytic efficiencies of the different enzymes. Rather than trying to enhance the activities of the less efficient enzymes in the pathway, Jones-Prather decided instead to colocalize the three enzymes and optimize their relative abundances. This was accomplished by

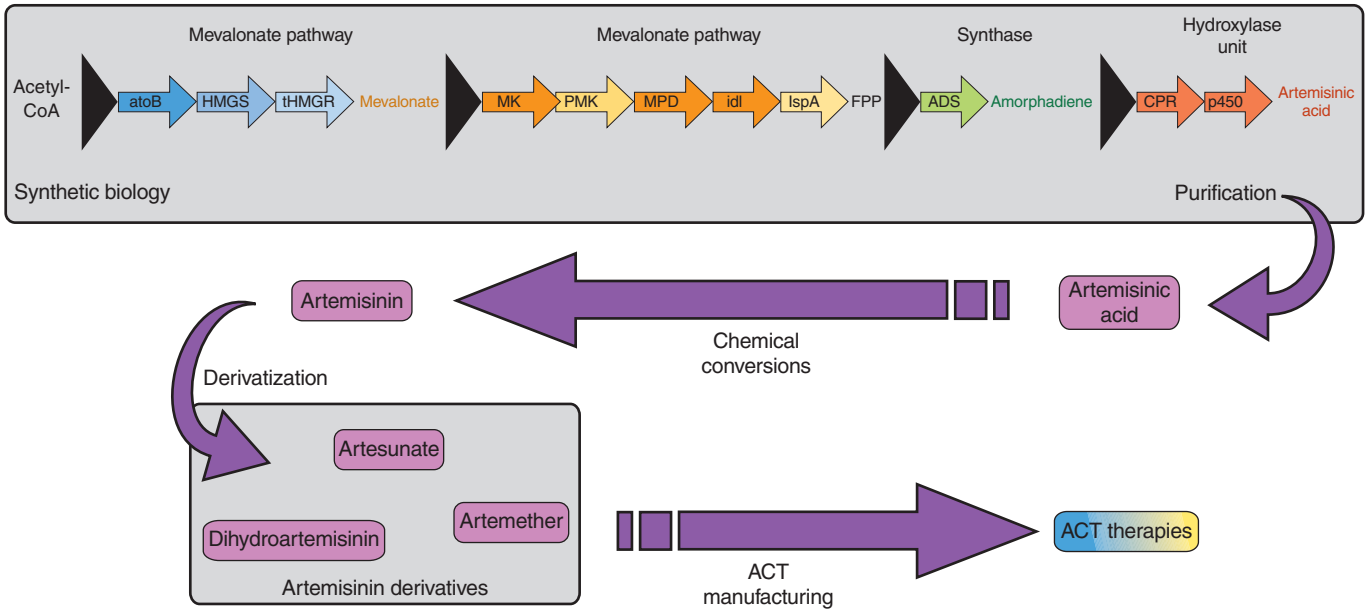


Figure 3 Making artemisinin. The process for the microbial production of artemisinin. Using synthetic biology, the metabolism of the microbe is engineered to produce artemisinic acid, a precursor to artemisinin. The artemisinic acid is released from the microbes, purified from the culture media then chemically converted to artemisinin. Once the artemisinin is produced, it must be further chemically converted into a derivative such as artesunate or artemether, which are integrated into artemisinin-based combination therapies (ACT) for the treatment of malaria. Copyright © 2007 The American Society of Tropical Medicine and Hygiene.

tagging each enzyme with a protein ligand and targeting these to a scaffold designed to recruit the enzymes in an optimal ratio. "So without figuring out specifically why our system wasn't doing what we wanted, we thought that bringing together the enzymes would make it work better, and it did," Jones-Prather says. It increased the product output by threefold[10].

Some other technologies being developed in academic laboratories also give a glimpse of the scale and efficiency of genome engineering that might have important industrial applications. Church and his colleagues, for example, have developed a way of combining directed evolution with synthetic oligos designed to target specific sites within the genomes—a technique they call multiplex automated genome engineering (MAGE)[11]. Starting with a set of genes in a particular pathway, they modify the strength of regulatory elements that control expression levels of the genes by using recombination to substitute short stretches of the host cell's DNA in the genes' ribosome binding sites, done robotically and iteratively (**Fig. 4**). This introduces changes in the targeted sites throughout the genome all at once, and with their microfluidic machine, they can continuously monitor the phenotype. When Church and colleagues applied this strategy to 20 genes required for lycopene accumulation in *E. coli*—which conveniently turns the cells red, making its synthesis easy to detect—they needed only three days to generate a bacterium that produced five times more pigment than an unoptimized strain.

Church describes MAGE as a "demonstration of accelerated evolution targeted by metabolic engineering knowledge for industrial-scale production." He adds, "MAGE is an attempt to expedite two kinds of research. One is building a genome that has a particular sequence. The other application is providing a number of possible solutions to a genome."

Other companies are working on ways to reduce industry's need for petroleum-based products. At Genomatica (San Diego), for example, CEO Christophe Schilling and his colleagues are building a computer model that simulates a metabolic system. "If we want to make a chemical, we use a computer model to see how it can be done and to see which path would give the highest yield and which organism would be the best to use," he says. They take that blueprint to the lab, where they fine-tune the process. With this approach, Genomatica engineered microbes to turn sugar into 1,4-butanediol, which is used in the plastics and fiber industries, where it is made from petroleum feedstock. According to Schilling, the process is "nearing the levels that are being targeted to provide a cost advantage when commercially produced."

Other companies are also developing computational tools to engineer efficiency. For some projects at Verdezyne (Carlsbad, CA, USA), scientists use computer-designed oligos that self-assemble into full-length genes, which are then expressed at high levels in microorganisms. CSO Stephen Picataggio explains, "We're developing a yeast-production platform, optimizing the conversion of sugars from various feedstocks for biofuels and chemicals." Picataggio knows that the green side of sugar feedstocks encourages chemical makers to adopt it, but he adds, "it has to be cost-advantaged." Apparently, investors believe that Verdezyne can turn its technology into such a cost cutter, because the company runs on venture capital, along with some internal investment.

What lies ahead?

The prospect of engineering new pathways and even new organisms may open up exciting possibilities for new products with new activities, but the commercial promise will have to be bal-

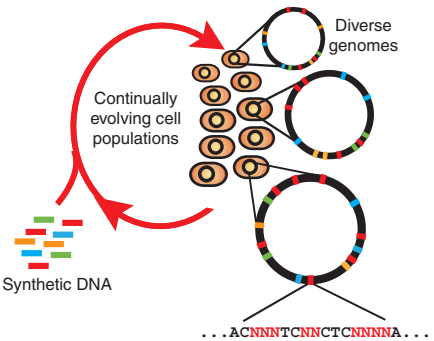


Figure 4 Multiplex automated genome engineering. The process enables the rapid and continuous generation of sequence diversity at many targeted chromosomal locations across a large population of cells through the repeated introduction of synthetic DNA. (Reprinted from ref. 11.)

anced with the dangers inherent in unfettered dissemination of genome engineering technology (**Box 2**). So far and for years, companies have been attempting to address the problem through self-regulation. The International Association of Synthetic Biology (Heidelberg) in November finally finished drafting a code of conduct (not yet available on their website), and so far, the four companies that were involved in its development are signing on. But getting everyone on board with a single set of standards may be problematic[12].

Beyond safety issues, synthetic biology also faces legal and regulatory challenges. "The patents involved in synthetic biology intellectual property have not been tested," explains J. Mark Waxman, a partner with Foley & Lardner (Boston), "and a number of them make some very broad claims." In their splashy 2007 report "Extreme Genetic Engineering," the nonprofit Action Group on Erosion, Technology and Concentration (ETC) Group identified a range of already patented products and processes related to synthetic biology, including methods for building synthetic oligos and genes, engineering biosynthetic pathways and making novel nucleotides. The report concludes, "Some of these patents cast an extremely wide net[13]." As an example of that, they point to US patent 6,521,427 issued to Glen Evans of Egea Biosciences[14] (San Diego), which covers chemical synthesis and assembly of genes and genomes. The ETC Group call this "potentially a description of the entire synthetic biology endeavor."

That early stage of IP mirrors similarly unanswered questions in the regulatory environment. Waxman points out that many existing regulations—such as state and Federal statutes on pesticides—affect synthetic biology. Nonetheless, more regulatory discussions lie ahead. "We need to reach a consensus on what ought to be regulated and how," Waxman says. That will probably be much more difficult to do than it is to say. With gene synthesis, says Waxman, "the problem may be difficult to solve," especially as this technology becomes less expensive and more widely available.

Finally, many of these technologies remain in their infancy, so commercialization is likely to be fraught with challenges. Because so many of the details remain to be resolved, BCC's Bergin thinks the market for developing products derived from synthetic or partially synthetic organisms may take several years to emerge. For those that are first to market with products and a solid and defensible intellectual property position, the commercial rewards are likely to be great.

1. Bergin, J. *Synthetic Biology: Emerging Global Markets* (BCC Research, Wellesley, MA, USA, June 2009).
2. Smith, H.O., Hutchison, C.A. III, Pfannkoch, C. & Venter, J.C. *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445 (2003).
3. Lartigue, C. *et al. Science* **317**, 632–638 (2007).
4. Gibson, D.G. *et al. Science* **319**, 1215–1220 (2008).
5. Lartigue, C. *et al. Science* **325**, 1693–1696 (2009).
6. Weichang, Y. *et al. Proc. Natl. Acad. Sci. USA* **103**, 17331–17336 (2007).
7. Carlson, S.R. *et al. PLoS Genet.* **3**, 1965–1974 (2007).
8. Kuroiwa, Y. *et al. Nat. Biotechnol.* **20**, 889–894 (2002).
9. Ro, D.-K. *et al. Nature* **440**, 940–943, 2006.
10. Dueber, J.E. *et al. Nat. Biotechnol.* **27**, 753–759 (2009).
11. Wang, H.H. *et al. Nature* **460**, 894–898 (2009).
12. Lok, C. Gene-makers put forward security standards. *Nat. News*, published online, 10.1038/news.2009.1065 (Nov. 4, 2009).
13. Anonymous. *Extreme Genetic Engineering: An Introduction to Synthetic Biology* (ETC Group, Ottawa, ON, Canada, January 2007). http://etcgroup.org/upload/publication/602/01/synbioreportweb.pdf
14. Evans, G.A. Method for the complete chemical synthesis and assembly of genes and genomes. US patent 6,521,427 (2003).
15. Check Hayden, E. & Ledford, H. *Nature* **458**, 818 (2009).

The sorcerer of synthetic genomes

Andrew Marshall

J. Craig Venter reflects on an effort spanning decades to create a living cell from chemically synthesized building blocks.

Dovetailed into J. Craig Venter's other scientific accomplishments—pioneering the use of novel sequencing approaches to decode tissue transcripts, microbial genomes and ultimately the human genome, not to mention his more recent exploits to catalog and sample the microbial diversity of the world's oceans—has been more than 15 years of work aimed at synthesizing a living organism from simple, chemical building blocks. This culminated with his most recent paper in *Science*[1], published together with his collaborators Hamilton Smith and Clyde Hutchison at the J. Craig Venter Institute (Rockville, MD, USA), that finally demonstrates the feasibility of transferring a genome from a prokaryote to yeast and then back into a different prokaryote. Venter and his team are now poised to take the last tantalizing step—constructing a genome synthetically and then rebooting that to life. *Nature Biotechnology* talked to him about the work and its implications for the future of biological engineering.

J. Craig Venter and his group at the JCVI are forging new ground in the field of synthetic genomics.

How did the synthetic genomics effort first come together?

J. Craig Venter: It started back in 1995 when we sequenced the first two genomes in history. The first genome was *Haemophilus influenzae* that had about 1,800 genes. After it was clear that our new method worked, we looked for a second genome to sequence that year. So the question came up: What would be the most interesting organism to sequence for the first genome comparison? Ham [Smith] and I got talking and we heard about Clyde Hutchison's work, where he'd been characterizing *Mycoplasma genitalium*, which he claimed had the smallest genome of any independently self-replicating organism, something that is still true today. And so after sequencing *Haemophilus*, we quickly sequenced the *Mycoplasma genitalium* genome. On doing the first-ever genome comparisons, we immediately started asking questions like: How small could a genome be and was there a minimal operating system?

What other kinds of questions centered around minimal genomes?

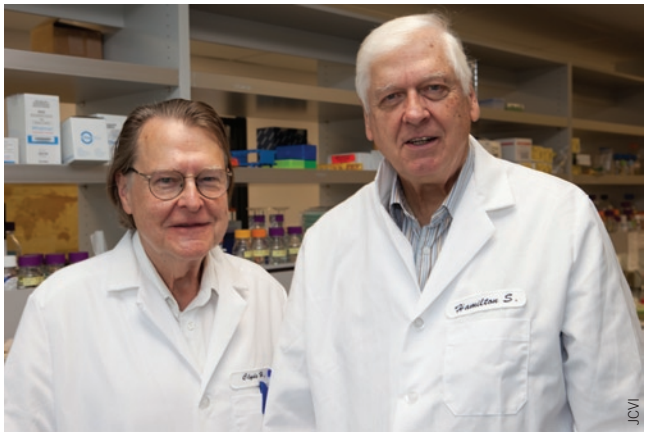
J.C.V.: To put our thinking into context: right after sequencing the *M. genitalium* genome, we started sequencing the third genome, *Methanococcus jannaschii*, the first Archaea genome that was published in 1996. At the same time, some NASA [National Aeronautics and Space Administration; Washington, DC, USA] scientists claimed they'd discovered these fossils of 'nanobacteria' in Martian meteorites. It turned out to be a complete artifact. But we sat down (like a lot of other scientists around the world) and did some calculations about the volume of those hypothetical nanobacteria and asked could they have any volume at all and could they even support small DNA molecules? The answer was no. As a result of all this, we had a great deal of thinking and discussion about minimal genomes.

How did the minimal genome work get underway?

J.C.V.: After we'd sequenced *M. genitalium*, we decided to start knocking out genes in the mycoplasma to see how many genes it could dispense with. That's one of those ideas that's very easy to say, but it's been extremely hard and frustrating to carry out, in part due to the lack of a genetics system in *M. genitalium*. So we (primarily Clyde Hutchison) developed this new approach that we called 'whole transposon mutagenesis' where we randomly inserted transposons into the

Andrew Marshall is Editor, Nature Biotechnology.

genome and then selected for cells that could survive the insertions in their genome. We then sequenced using a primer off the transposon to find out where it was inserted into the genome[2]. This was a whole new approach that you could only do if you had a sequenced genome. But a major limitation of this method was we could only knock out genes one at a time. While we collected more and more knockouts, we found that it did not tell us whether the genes could all be knocked out together, due to a lack of selectable markers. As a result, we decided the only way one could make a minimal genome would be to chemically synthesize the chromosome and then physically vary the gene content—and so that was the start of the field of synthetic genomics.



Clyde Hutchison (left) and Ham Smith (right), who have spearheaded the work by the teams at JCVI aiming to create a living organism from chemical building blocks.

What led you to ΦX174 as the first genome to synthesize?

J.C.V.: It actually had a very slow, difficult start. Clyde Hutchison was in Fred Sanger's lab when they sequenced ΦX174—the first DNA virus ever sequenced—and because of its historic import, we decided to synthesize that genome primarily as a test to see whether we could accurately synthesize genomes. That simple idea ended up being about a ten-year project—in part because we stopped for two years to sequence the human genome. But we just thought it would be simple and that we'd just make PCR primers that had sufficient overlap, anneal them together and then PCR copy the whole thing. When we did this, we obtained DNA molecules the right size (5 kb), but nothing worked. Even with selection by infectivity—where one molecule out of a million would have seen virus particles made—we got nothing. And it turns out that there are just too many errors in DNA synthesis.

After sequencing the human genome, Ham and I started back in on the project and then recruited Clyde up from North Carolina [Chapel Hill] to join us. As a result, we spent many years, particularly Ham and Clyde, working out error correction in the synthesis. That culminated with our report when it took two weeks going from the DNA sequence in the computer to synthesizing the ΦX174 genome, which was activated by injecting it into *E. coli.* The *E. coli* cellular machinery read the synthetic genetic DNA and produced the viral proteins, which self assembled to form the active virus[3].

Could you talk a little more about error correction?

J.C.V.: What we described in the ΦX174 paper[3] were some nice elegant methods for doing repair in real time off of a correct strand, but we still had to select clones and sequence them to ensure the correct order of bases. What ΦX174 gave us was the confidence that we could build accurate DNA units of 5 kb; our assumption was that we could assemble the smaller units using homologous recombination.

So we had a team of several scientists working on this problem. One of the early genomes we sequenced was *Deinococcus radiodurans,* which has a phenomenal DNA repair system that can take these huge doses of ionizing radiation (up to 3 mrads), blowing its chromosome apart with several hundred double-stranded DNA breaks, and then over 12 to 24 hours reassemble the genome as it was before. We spent years trying to isolate the DNA repair genes out of *Deinococcus*, and cloning them, to attempt to create an *in vitro* expression system to assemble our DNA fragments. But we never got it working outside of the intact cells.

It was about that same time that the team led by Dan Gibson discovered that we could assemble our synthetic DNA in yeast using its recombination system. This multi-year work culminated almost two years ago now with the complete synthesis of the *Mycoplasma genitalium* genome[4].

At every stage, we've had to develop new methodology and tools. Over 100 kb, the synthetic DNA segments were too big to clone in *Escherichia coli.* We were looking for another system and discovered if we just used an artificial yeast centromere we could convert bacterial genomes into yeast chromosomes. Work led by Gwyn Benders allows us to clone complete bacterial genomes in yeast artificial chromosomes just by adding a yeast centromere to them[1].

Tell us more about the most recent step of the work.

J.C.V.: The way we had originally envisioned it, we were just going to have the synthetic genome that we assembled on the lab bench and then we were going to transplant it into a recipient cell. But because we ended up doing the final genome assembly in yeast using homologous recombination, we now had to develop whole new methods for isolating our synthetic bacterial chromosome from yeast and then transplanting it.

In our original genome transplantation study, we isolated *Mycoplasma mycoides* genome and transplanted it into a closely related species[5]. But after cloning the *M. mycoides* chromosome in yeast and then isolating it, it would not transplant. It took 20 people two years to solve that little riddle of why it would not transplant from yeast but it would from *M. mycoides* cells[1]. We knew something was happening to the DNA in the *M. mycoides* cell that wasn't happening in yeast. It turns out the secret was DNA methylation.

What kind of approaches did you use to solve the riddle?

J.C.V.: While we were attempting to transfer and boot-up DNA derived from yeast, different members of the team worked out the methods for cloning bacterial chromosomes in yeast—something no one had ever done before. So those are pretty cool methods on their own. The methylation work involved the development of a number of new methods, including trying cellular extracts and using them to methylate the chromosome out of yeast. We purified and cloned all the specific methylases and used them to methylate the bacterial genomes cloned in and extracted from yeast. None of this is trivial as you cannot just pipette entire chromosomes and keep them intact as supercoiled DNA. We have to move and modify the genomes in gel blocks. All the enzymology has to take place in gel blocks. And it takes careful handling not to destroy the chromosomes. The team has done absolutely phenomenal work. As with all things in science, it's the little tiny breakthroughs on a daily basis that make for the big breakthrough.

Did you look at anything else other than methylation?

J.C.V.: We did all these studies to make sure that there were no proteins needed for transplantation. Because you could envision DNA-binding proteins—the equivalent of histones or some similar mechanism in our genomes—required to stabilize the genome. And so we used proteinases to digest all the proteins associated with the extracted DNA and we still found we could get the *M. mycoides* chromosome to transplant. We also worked out that it had to be supercoiled DNA, and if it wasn't supercoiled, clean DNA, it would not transplant.

What have we learned concerning the compatibility of a donor genome with a recipient cell?

J.C.V.: We learned very early on from our first genome sequencing—that of *Haemophilus*—that there are gaps in microbial sequences with the initial assemblies. Genes or sequences that might be toxic to *E. coli,* such as the *Haemophilus* ribosome genes, would delete or truncate when cloned. So it took a huge effort to totally close those genomes because we had to find ways to get clones so that we could sequence walk across the deleted sequence. One of several advantages of the new sequencing techniques is that we don't need cloning in *E. coli*.

When we had the synthetic quarter molecules of the synthetic *Mycoplasma genitalium* genome (when we had pieces of 175,000 base pairs), we got two of the four to clone initially in *E. coli*. All four have now been grown in *E. coli*, but for some reason, passing them through yeast first made them clonable, which we don't understand. We have to solve each of these riddles one at a time. That's why it's so slow. The good news is instead of just gee whiz quickly getting a synthetic cell, we're really learning the processes of life and now being able to move what we call the software of life across the branches of life. So, in a way, it's good that it's taken us 15 years to do because we've just learned so much that's really critical for the next stages by struggling to get through it. We just lucked out with one system and it worked. It just as easily could have led us down a blind alley for a long time.

What about evolutionary divergence and chromosome compatibility?

J.C.V.: The difference between *M. mycoides* and *M. capricolum* is roughly the same as the difference between mice and humans. We don't know how different the DNA can be. Basically, we're assuming we could eventually design a system to be a universal recipient. One where you have the right tRNAs and have the ability to start reading the genetic code. It's a fundamental aspect of life that you see when you throw the ΦX174 genome into *E. coli* and it just starts reading the DNA and making the viral proteins and they self-assemble. I think the key thing is probably just having the protection against the restriction enzymes in the cell first and foremost, and then having the right machinery to be able to read that genetic code and express those genes. But we don't know how far afield we can go. I'm betting quite far as long as we stick to those fundamental rules. That's the kind of thing we're just starting to test right now.

What role does codon usage play in stability of the clones?

J.C.V.: Did the different codon usage in *M. mycoides* and lack of transcription in *E. coli* facilitate the stable cloning of assembled mycoplasma fragments in *E. coli*? Was it just the UGA codon that yeast didn't recognize so a lot of the proteins weren't translated? With *E. coli,* you always assume that when you have proteins that are expressed that could interfere with *E. coli* biology, that the bacteria essentially goes back and deletes them. But we're doing some work now that suggests it may not be an issue at all. It might just be a matter of selecting for stable clones. We have a different mycoplasma genome that we've cloned in now with basically the same codon usage as yeast, and it is totally stable as well.

> "Every time we've tried to reboot a synthetic genome, we've come up with a new set of challenges."

But presumably, recoding starting genomes is going be very important?

J.C.V.: What works with one cell is probably going to take some engineering for each new one we take on.

What about the final step of rebooting a synthetic genome? How close are you to that?

J.C.V.: Every time we've tried to reboot a synthetic genome, we've come up with a new set of challenges. Our view now is that we've solved them all, but we'll only know that when we actually have the cell totally controlled by a completely chemically made genome. As yet we do not have that. There's still a chance it will happen this year—which I somewhat optimistically remind people I've said for the past two years now.

Will everything in the future be made from DNA synthesizers?

J.C.V.: No. We'll start with our repertoire of 20 to 40 million genes, and some of them will need to be synthesized and the rest we'll make PCR copies. We'll have 40 million bottles of genes and we'll pull those down for assembling genomes in the future. But I think for proof of concept, it's important for us to start with 4 bottles of chemicals and to watermark the genome—to make it absolutely foolproof that it's really the synthetic chromosome that is controlling the cells. It is essentially important as a theoretical concept, but we're not quite there yet.

Do you anticipate ethical controversy once rebooting of a synthetic genome is demonstrated?

J.C.V.: We have asked and driven the ethical discussion from the beginning. We've been trying to bring the community along with us every step of the way. We think once we do activate a genome that yes, it probably will impact people's thinking about life. But I think it already has, as we've progressed in a logical fashion with each step of these studies. Perhaps one of the good things about it taking so much time for us to do all this work is that we've had time to have the in-depth ethical discussions before we get to that key experiment. For example, the Sloan Foundation report [http://www.synbioproject.org/library/publications/archive/synbio3/], the NSABB [National Science Advisory Board for Biosecurity; http://oba.od.nih.gov/biosecurity/pdf/Final_NSABB_Report_on_Synthetic_Genomics.pdf], the Fink Report out of the US National Academy of Sciences and the report out of the Royal Academy in the UK [http://www.raeng.org.uk/news/publications/list/reports/Synthetic_biology.pdf]. We've watched the mistakes of others, for example with the issues that GMOs [genetically modified organisms] have had. We've worked really hard to bring the world along with us as we develop each step of the technology.

How far are we from understanding how to regulate complex genetic circuits in a synthetic system?

J.C.V.: Using the principle of in-the-lab evolution, if you have a minimal chassis you should be able to replicate billions of years of evolution by adding back components. When you look at what the various

groups—Jay Keasling's group and their work to engineer yeast to produce artemisinin, and what they did at DuPont (Wilmington, DE, USA) for PDO [1,3 propanediol] synthesis in *E. coli*—much of that effort was in shutting off pathways that interfere with the flow of carbon. It would be ideal to build life based on first principle. We don't want to start with complex systems and try to unwind them; we're trying to build things we truly understand and know how to control them. You want to start with a minimal system and add what you need to it. Or have some universal chassis that really does work. This is what our team is now working on, including using new amino acids. Some whole groups are developing biological circuits. This will be like using the transistors and the capacitors of the past. And the future will be taking biological components off the shelf and designing the circuitry. We're just trying to establish the first basic principles to do that. I've seen over again in science, once that happens there's a very, very rapid transition to the next stages.

What would a commercial operation using synthetic organisms look like?

J.C.V.: Currently, companies are looking at large facilities to convert sunlight to fuels or oil or production of foodstuffs or production of clean water—it will start with natural organisms or some pretty simple modifications of existing organisms. If you invest billions into infrastructure using version 1.0 of a cell—that is, a cell that's been modified with classic genetic engineering or metabolic engineering or even putting in some synthetically made pathways that have been proven to work, then the future will be totally based on synthetic genomics.

Because once you have that multibillion dollar infrastructure built for a facility, then the design of the biology will become the single most important thing for the future economic value.

The changes will be rapid. It's just like any industry. Look how simplistic what we were doing 15 years ago in DNA sequencing was compared to today. Our early work was considered such great milestones. At TIGR [The Institute for Genome Research; now the JCVI in Rockville, MD], we had the biggest facility in the world capable of running 100,000 sequences in one year. At Celera [Rockville, MD], we did 100,000 a day with 300 machines, which is now possible with single next-generation sequencing machines. The same changes are going to happen with synthetic biology. The tools everybody's using now are very primitive. I think the concepts behind what our team is doing today are important because they prove that what we're saying is possible.

> "There's not a single aspect of human life that doesn't have the potential to be totally transformed by these technologies in the future."

How does your work fit with other research in synthetic biology?

J.C.V.: There's a lot of basic molecular biology that many groups now call synthetic biology. It's sort of a bit like systems biology; systems biology used to be called physiology. There are some people like Lee Hood and others who are pushing the limits and defining what systems biology should be. But too many people just now think it's a sexy term, they stopped doing physiology and they're all doing systems biology. These are new catchy sexy terms so everybody wants to pretend they're doing systems biology and synthetic genomics.

What do you see as the ultimate significance of this work?

J.C.V.: The initial goal of our work was to understand how basic cellular life works and if it could be replicated. I think our findings, that what you need for life is the DNA information molecule and the ability to read the information to produce proteins that self assemble into living cells, are very important. We are starting with living cells and reprogramming them with new DNA software of life but we are not creating life from basic elements. I think it is very surprising to many that we can reprogram cells into new species simply by changing out the software.

I've defined synthetic genomics in a very precise way compared with synthetic biology—which can be anything from molecular biology to genetic engineering to gene circuits. But with synthetic genomics, the goal is to start in the computer in the digital world from digitized biology and make new DNA constructs for very specific purposes. That's why the proof of concept of being able to do that is so critical. We're not there yet, but we are close. It can mean that as we learn the rules of life we will be able to develop robotics and computational systems that are self learning systems. By doing combinatorial genomics using the 20 million genes in our databases, single robotic systems can learn more biology than in the previous decade. It's the beginning of the new era of very rapid learning. If science moves forward in a linear fashion, we've all failed. There's not a single aspect of human life that doesn't have the potential to be totally transformed by these technologies in the future.

1. Lartigue, C. *et al.* Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693–1696 (2009).
2. Hutchison, C.A. III *et al.* Global transposon mutagenesis and a minimal mycoplasma genome. *Science* **286**, 2165–2169 (1999).
3. Smith, H.O., Hutchison, C.A. III, Pfannkoch, C. & Venter, J.C. Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445 (2003).
4. Gibson, D.G. *et al.* Complete chemical synthesis, assembly and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
5. Lartigue, C. *et al.* Genome transplantation in bacteria: changing one species into another. *Science* **317**, 632–638 (2007).

Recent patent applications in synthetic biology

Patent number	Description	Assignee	Inventor	Priority application date	Publication date
JP 2009142273	An RNA-protein complex containing a substrate RNA derived from an RNA-protein complex interaction motif and a fusion protein including an amino acid sequence that couples specifically with RNA; useful for controlling cell function, as a raw material in synthetic biology for reconstructing a biomolecule, and as sensor, or molecular switch, in electronics, imaging, nanotechnology and medical treatment applications.	Japan Science & Technology Agency (Saitama, Japan)	Inoue M, Kikuta M, Kuramitsu S, Saito H	11/22/2007	7/2/2009
WO 2009048971	A method of expressing a protein containing nonstandard amino acids comprising providing a host organism and an orthogonal tRNA system comprising a nonstandard tRNA, a nonstandard aminoacyl-tRNA synthetase and a nonstandard amino acid. This invention combined with the capacity to synthesize whole genomes has important implications in synthetic biology, as it allows the rewriting of the genetic code of existing or newly designed organisms.	Synthetic Genomics (La Jolla, CA, USA)	Glass JI, Krishnakumar R, Merryman CE	10/8/2007	4/16/2009
US 20090061520	A method of creating a synthetic genetic circuit, comprising providing a host cell, a first vector and a second vector, and applying the second vector to the host cell such that recombination occurs between the selectable marker sequence in the first vector and the first and second homologous sequences.	University of Michigan (Ann Arbor, MI, USA)	Mayo AE, Ninfa A, Selinsky S, Song QX, Woolf P	11/3/2006	3/5/2009
WO 2008144060, US 20090047718	A first recombinant solventogenic organism comprising an altered expression of a gene involved in a solvent production pathway relative to the expression in the first organism strain prior to its transformation.	Advanced Biofuels (Chicago), Blaschek HP, Shi Z, Stoddard SF, TetraVitae Bioscience (Chicago)	Blaschek HP, Shi Z, Stoddard SF	5/17/2007	11/27/2008, 2/19/2009
US 20070264688, WO 2008024129	A method of constructing a synthetic genome for making synthetic cells for use in generating synthetic fuels, e.g., hydrogen or ethanol, by assembling nucleic acid cassettes that comprise portions of the synthetic genome.	J. Craig Venter Institute (Rockville, MD, USA), Hutchison CA, Smith HO, Venter JC	Hutchison CA, Smith HO, Venter JC	12/6/2005	11/15/2007, 2/28/2008
US 20070031942 CN 101133166	Producing polymers of nucleic acids by hybridizing the oligonucleotide mixture comprising oligonucleotides to capture probes and joining the nicking and gapping sites contained in the hybridizing duplex using ligation; enables parallel multiplex ligation and amplification on surfaces for making assemblies of nucleic acids of various biological applications and for analysis of biological samples such as DNA, RNA and proteins.	Gao X, Sheng N, Zhang X, Zhou X, Zhu Q	Gao X, Sheng N, Zhang X, Zhou X, Zhu Q	3/1/2005	2/8/2007, 2/27/2008
US 20070269862, WO 2008016380	A method of making a synthetic cell comprising introducing a genome or partial genome into a cell or cell-like system.	Assad-Garcia N, Glass JI, Hutchison CA, Lartigue C, Smith HO, Venter JC, Young L, J. Craig Venter Institute (Rockville, MD, USA)	Assad-Garcia N, Glass JI, Hutchison CA, Lartigue C, Smith HO, Venter JC, Young L	12/23/2005	11/22/2007, 2/7/2008
US 20070087366, WO 2007085906	A new composition comprising a loxP recombination element having a left inverted repeat region, a right inverted repeat region and a spacer region comprising spacer regions; useful for carrying out multiple non-cross-reacting recombination reactions in synthetic biology and metabolic engineering.	Holt RA, Missirlis PI, BC Cancer Agency (Vancouver, BC, Canada)	Holt RA, Missirlis PI	10/13/2005	4/19/2007, 8/2/2007

Source: Thomson Scientific Search Service. The status of each application is slightly different from country to country. For further details, contact Thomson Scientific, 1800 Diagonal Road, Suite 250, Alexandria, Virginia 22314, USA. Tel: 1 (800) 337-9368 (http://www.thomson.com/scientific).

nature biotechnology

PERSPECTIVE

Next-generation synthetic gene networks

Timothy K Lu[1–3], Ahmad S Khalil[3] & James J Collins[3,4]

Synthetic biology is focused on the rational construction of biological systems based on engineering principles. During the field's first decade of development, significant progress has been made in designing biological parts and assembling them into genetic circuits to achieve basic functionalities. These circuits have been used to construct proof-of-principle systems with promising results in industrial and medical applications. However, advances in synthetic biology have been limited by a lack of interoperable parts, techniques for dynamically probing biological systems and frameworks for the reliable construction and operation of complex, higher-order networks. As these challenges are addressed, synthetic biologists will be able to construct useful next-generation synthetic gene networks with real-world applications in medicine, biotechnology, bioremediation and bioenergy.

Ten years since the introduction of the field's inaugural devices—the genetic toggle switch (J.J.C. and colleagues)[1] and repressilator[2]—synthetic biologists have successfully engineered a wide range of functionality into artificial gene circuits, creating switches[1,3–9], oscillators[2,10–12], digital logic evaluators[13,14], filters[15–17], sensors[18–20] and cell-cell communicators[15,19]. Some of these engineered gene networks have been applied to perform useful tasks such as population control[21], decision making for whole-cell biosensors[19], genetic timing for fermentation processes (J.J.C. and colleagues)[22] and image processing[23–25]. Synthetic biologists have even begun to address important medical and industrial problems with engineered organisms, such as bacteria that invade cancer cells[26], bacteriophages with enhanced abilities to treat infectious diseases (T.K.L. and J.J.C.)[27,28], and yeast with synthetic microbial pathways that enable the production of antimalarial drug precursors[29]. However, in most application-driven cases, engineered organisms contain only simple gene circuits that do not fully exploit the potential of synthetic biology. There remains a fundamental disconnect between low-level genetic circuitry and the promise of assembling these circuits into more complex gene networks that exhibit robust, predictable behaviors.

Thus, despite all of its successes, many more challenges remain in advancing synthetic biology to the realm of higher-order networks with programmable functionality and real-world applicability. Here, instead of reviewing the progress that has been made in synthetic biology, we present challenges and goals for next-generation synthetic gene networks, and describe some of the more compelling circuits to be developed and application areas to be considered.

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [2]Harvard-MIT Health Sciences and Technology, Cambridge, Massachusetts, USA. [3]Howard Hughes Medical Institute, Department of Biomedical Engineering, Center for BioDynamics, and Center for Advanced Biotechnology, Boston University, Boston, Massachusetts, USA. [4]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA. Correspondence should be addressed to T.K.L. (timlu@mit.edu).

Published online 9 December 2009; doi:10.1038/nbt1591

Synthetic gene networks: what have we learned and what do we need?

The engineering of mechanical, electrical and chemical systems is enabled by well-established frameworks for handling complexity, reliable means of probing and manipulating system states and the use of testing platforms—tools that are largely lacking in the engineering of biology. Developing properly functioning biological circuits can involve complicated protocols for DNA construction, rudimentary model-guided and rational design, and repeated rounds of trial and error followed by fine-tuning. Limitations in characterizing kinetic processes and interactions between synthetic components and other unknown constituents *in vivo* make troubleshooting and modeling frustrating and prohibitively time consuming. As a result, the design cycle for engineering synthetic gene networks remains slow and error prone.

Fortunately, advances are being made in streamlining the physical construction of artificial biological systems, in the form of resources and methods for building larger engineered DNA systems from smaller defined parts[22,30–32]. Additionally, large-scale DNA sequencing and synthesis technologies are gradually enabling researchers to directly program whole genes, genetic circuits and even genomes, as well as to re-encode DNA sequences with optimal codons and minimal restriction sites (see review[33]).

Despite these advances in molecular construction, the task of building synthetic gene networks that function as desired remains extremely challenging. Accelerated, large-scale diversification[34] and the use of characterized component libraries in conjunction with *in silico* models for a priori design[22] are proving useful in helping to fine-tune network performance toward desired outputs. Even so, in general, synthetic biologists are often fundamentally limited by a dearth of interoperable and modular biological parts, predictive computational modeling capabilities, reliable means of characterizing information flow through engineered gene networks and test platforms for rapidly designing and constructing synthetic circuits.

In the following subsections, we discuss four important research efforts that will improve and accelerate the design cycle for next-generation synthetic gene networks: first, advancing and expanding the toolkit of available parts and modules; second, modeling and fine-tuning

the behavior of synthetic circuits; third, developing probes for reliably quantifying state values for synthetic (and natural) biomolecular systems; and fourth, creating test platforms for characterizing component interactions within engineered gene networks, designing gene circuits with increasing complexity and developing complex circuits for use in higher organisms. These advances will allow synthetic biologists to realize higher-order networks with desired functionalities for satisfying real-world applications.

Interoperable parts and modules for synthetic gene networks. Although there has been no shortage of novel circuit topologies to construct, limitations in the number of interoperable and well-characterized parts have constrained the development of more complex biological systems[22,31,35,36]. The situation is complicated by the fact that many potential interactions between biological parts, which are derived from a variety of sources within different cellular backgrounds, are not well understood or characterized. As a result, the majority of synthetic circuits are still constructed *ad hoc* from a small number of commonly used components (e.g., LacI, TetR and lambda repressor proteins and regulated promoters) with a significant amount of trial and error. There is a pressing need to expand the synthetic biology toolkit of available parts and modules. Because physical interconnections cannot be made in biological systems to the same extent as electrical and mechanical systems, interoperability must be derived from chemical specificity between parts and their desired targets. This limits our ability to construct truly modular parts and highlights the need for rigorous characterization of component interactions so that detrimental interactions can be minimized and factored into computational models.

Engineered zinc fingers constitute a flexible system for targeting specific DNA sequences, one which could significantly expand the available synthetic biology toolkit for performing targeted recombination, controlling transcriptional activity and making circuit interconnections. Zinc-finger technology has primarily been used to design zinc-finger nucleases that generate targeted double-strand breaks for genomic modifications[37]. These engineered nucleases may be used to enhance recombination in large-scale genome engineering techniques[34]. A second and potentially very promising use of engineered zinc fingers is as a source of interoperable transcription factors, which would greatly expand the current and limited repertoire of useful activators and repressors. In fact, zinc fingers have already been harnessed to create artificial transcription factors by fusing zinc-finger proteins with activation or repression domains[38,39]. Libraries of externally controllable transcriptional activators or repressors could be created by engineering protein or RNA ligand-responsive regulators, which control the transcription or translation of zinc finger–based artificial transcription factors themselves[18]. These libraries would enable the construction of basic circuits, such as genetic switches[1], as well as more complex gene networks. In fact, several of the higher-order networks we describe below rely on having multiple reliable and interoperable transcriptional activators and repressors for proper functioning.

Even so, these engineered transcription factors have not yet been fully characterized, and if they are to be used as building blocks for complex gene networks, then knowledge of their *in vivo* kinetics and input-output transfer functions would be beneficial. In addition, much of the rich dynamics associated with small, synthetic gene networks is attributable to the cooperative binding or multimerization of transcription factors, and it is not yet clear what further engineering is required to endow zinc-finger transcription factors with such features.

Nucleic acid–based parts, such as RNAs, are also promising candidates for libraries of interoperable parts because they can be rationally programmed based on sequence specificity[7,40,41]. Novel circuit interconnections could be established using small interfering RNAs (siRNAs) to control the expression of specific components. Recombinases, which target specific DNA recombinase-recognition sites, also represent a fruitful, underutilized source of interoperable parts. Recombinases have been used in the context of synthetic biology to create memory elements and genetic counters[9]. However, more than 100 natural recombinases are known, and these can be engineered by mutagenesis and directed evolution for greater diversity and sequence specificity[42–45].

Libraries of well-characterized, interoperable parts, such as transcription factors and recombinases, would vastly enhance the ability of synthetic biologists to build more complex gene networks with greater reliability and real-world applicability. In addition to libraries of individual parts, it would be of great value to have well-characterized and interoperable modules (e.g., switches, oscillators and interfaces) that could be used in a plug-and-play fashion to create higher-order networks and programmable cells. As the number of parts and modules expands, high-throughput, combinatorial efforts for quantifying the levels of interference and cross-talk between multiple components within cells will be increasingly important as guides for choosing the most appropriate components for network assembly.

Modeling and fine-tuning synthetic gene networks. Integrated efforts for modeling and fine-tuning synthetic gene circuits are useful for ensuring that assembled networks operate as intended. Such approaches will be increasingly important as more complex circuits are constructed along with the expanded development of interoperable parts. Although studies have shown that in some cases, component properties alone are sufficient for predicting network behavior[22,31,46], others have demonstrated the need for modeling and fine-tuning networks after their basic topologies have been established[1,22]. A multi-step design cycle that involves creating diverse component libraries, constructing, characterizing and modeling representative network topologies, and assembling and fine-tuning desired circuits, followed by subsequent refinement cycles[22], will be crucial for the successful design and construction of next-generation synthetic gene networks.

The fine-tuning of biomolecular parts and networks can be achieved by developing diverse component libraries through mutagenesis followed by in-depth characterization and modeling[22,47–51]. Significant progress has been made in tuning gene expression by altering transcriptional, translational and degradation activities. For example, promoter libraries with a range of transcriptional activities can be created and characterized, plugged into *in silico* models and then used to develop synthetic gene networks with defined outputs, without significant post-hoc adjustments[22,47–51]. Alternatively, synthetic ribosome binding site (RBS) sequences can be used to optimize protein expression levels. Recently, Salis *et al.*[52] have developed a thermodynamic model for predicting the relative translational initiation rates for a protein with different upstream RBS sequences, a model that can also be used to rationally forward-engineer RBS sequences to give desired protein expression. In addition, protein degradation can be controlled by tagging proteins with degradation-targeting peptides that impart different degradation dynamics[53].

By automating the construction and characterization of biomolecular components, extensive libraries could be created for the rapid design and construction of complex gene networks. These efforts, coupled with *in silico* modeling, would serve to fast-track synthetic biology (more detailed discussions of modeling techniques for synthetic biology are found in refs. 22,31,54–57). However, to build reliable models of biomolecular parts and networks, new methods

for probing and acquiring detailed *in vitro* and *in vivo* measurements are needed, which we discuss below.

Probes for characterizing synthetic gene networks. Significant advances have been made in the development of new technologies for manipulating biological systems and probing their internal states. At the single-molecule level, for instance, optical tweezers and atomic force microscopes provide new, direct ways to probe the biophysical states of single DNA, RNA and protein molecules as they undergo conformational changes and other dynamical processes[58–62]. However, we lack similar tools for tracking the *in vivo* operation of synthetic gene circuits in a high-throughput fashion. Ideally, making dynamical measurements of biological networks would involve placing sensors at multiple internal nodes, akin to how current and voltage are measured in electrical systems. Furthermore, external manipulation of synthetic biomolecular systems is typically accomplished by the addition of chemical inducers, which can suffer from cross-talk[63], be difficult to remove and be consumed over time. As a result, inputs are often troublesome to control dynamically.

Microfluidic devices have been coupled to single-cell microscopy and image processing techniques to enable increasingly precise manipulation and measurement of cells, especially since inputs can be modulated over time[64,65]. These systems allow the rapid addition and removal of chemical inducers, enabling more sophisticated, time-dependent inputs than conventional step functions, while also enabling researchers to track and quantify single cells for long periods of time. These developments make possible the wider use of well-established engineering approaches for analyzing circuits and other systems in synthetic biology. For example, frequency-domain analysis, a technique used commonly in electrical engineering[66,67], can be employed with microfluidics to characterize the transfer functions and noise behaviors of synthetic biological circuits[66–68]. Additionally, small-signal linearization of nonlinear gene circuits can be achieved by applying oscillatory perturbations with microfluidics and measuring responses at the single-cell level[67,68].

Indeed, microfluidics provides a useful platform for perturbing synthetic gene circuits with well-controlled inputs and observing the outputs in high-resolution fashion. Without the proper 'sensors' (that is, for quantitatively and simultaneously probing all the internal nodes of a given gene circuit), however, this technology alone is not sufficient to bring full, engineering-like characterization to synthetic gene networks.

Thus far, probes enabling quantitative measurements of synthetic gene circuits have primarily focused on the use of fluorescent proteins for *in vivo* quantification of promoter activity or protein expression. With the advent of novel mass spectrometry–based methods that provide global, absolute protein concentrations in cells[69], quantitative transcriptome data can now be merged with proteome data, improving our ability to characterize and model the dynamics of synthetic gene networks. Global proteomic data may also assist synthetic biologists in understanding the metabolic burden that artificial circuits place on host cells. Further efforts to devise fluorescent-based and other types of reporters for the simultaneous monitoring of transcriptome and proteome dynamics *in vivo* are needed to close the loop on full-circuit accounting. Some promising tools under development include tracking protein function by incorporating unnatural amino acids that exhibit fluorescence[70,71], quantum dots[72] and radiofrequency-controlled nanoparticles[73].

As the field awaits entire-circuit probes, there are, in the meantime, several potentially accessible technologies for increasing the throughput and pace of piecewise gene-circuit characterization. Recent advances in engineering light-inducible biological parts and systems[23,24,74] have unlocked the potential for optical-based circuit characterization, expanding the number and type of tunable knobs available to synthetic biologists. For instance, by coupling a synthetic gene network of interest to a biological light/dark sensor as well as to fluorescent protein outputs, one could potentially measure the network's input/output transfer function in a high-throughput fashion using spectrophotometric microplate readers, without having to add varying concentrations of chemical inducers. In essence, both control and monitoring of biomolecular systems would be accomplished using reliable and high-speed optics that are typically associated with fluorescence readouts and microscopy. This is an exciting prospect, particularly in the context of microfluidic devices, which would facilitate the focusing of optical inputs and readouts to single cells.

Using electrical signals, in lieu of chemical or optical signals, for control and monitoring of biological systems would also present high-speed advantages. Recently, advances have been made in integrating silicon electronics with lipid bilayers containing transmembrane pores to perform electronic signal conduction[75]. This technology may eventually allow direct communication and control between engineered cells and electronic circuits by means of ionic flow. The incorporation of these and other technologies to perturb and monitor the *in vivo* performance of synthetic gene networks will enable us to achieve desired functionality faster and more reliably.

Test platforms for engineering complex gene circuits. Increasing complexity—whether assembling larger synthetic gene networks from smaller ones or engineering circuits into higher organisms—dramatically increases the number of potential failure modes. In the former case, combining multiple individually functioning genetic circuits into a single cellular background can lead to unintended interactions among the synthetic components or with host factors, and these various failure modes are often difficult to pinpoint and isolate from one another. In the latter case, engineering synthetic networks for mammalian systems poses additional challenges beyond engineering circuits for bacterial and yeast strains, which have comparatively well-characterized genomes, transcriptomes, proteomes and metabolomes. Mammalian systems are much more complex and possess substantially less well-characterized components for engineering[76], but for these and other reasons, constitute fertile ground for new applications and genetic parts.

The development of test platforms where engineered gene circuits can be designed and validated before being deployed in other or more complex cellular backgrounds would mitigate failure-prone jumps in complexity. These platforms could be used to verify or debug circuit topology and basic functionality in well-controlled environments. For example, cells optimized for testing may be engineered to have minimal genomes to decrease the risk of pleiotropic or uncharacterized interactions between the host and the synthetic networks[77–81]. The use of orthogonal parts that are decoupled from host cells may enable the dedication of defined cellular resources to engineered functions, which can simplify the construction and troubleshooting of gene circuits. For example, nucleic acid–based parts can be designed to function orthogonally to the wild-type cellular machinery[82–84]. Artificial codons and unnatural amino acids, which have enabled new methods for studying existing proteins and the realization of proteins with novel functions, could also be used to produce synthetic circuits that function orthogonally to host cells[85]. Simplifying backgrounds would additionally enable more accurate computational modeling of complex circuits before they are deployed into their ultimate environments. Furthermore, minimal cells could themselves contain synthetic circuits that provide useful testing functionalities, such as multiplexed transcriptional and translational controls and output probes.

Lower organisms can also be useful for the construction and characterization of synthetic gene networks before such systems are

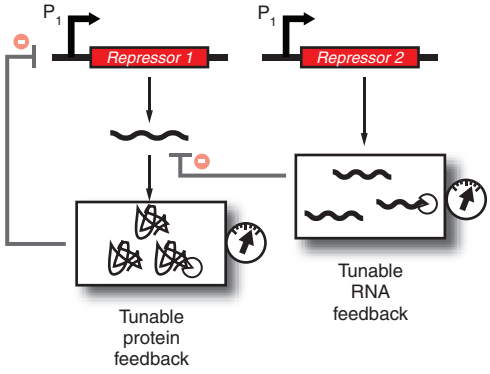


Figure 1 Tunable genetic filter. Filter characteristics can be adjusted by tuning the degradation of RNA and protein effectors in negative-feedback loops. Examples of RNA effectors include siRNAs, riboregulators and ribozymes. Examples of protein effectors include transcriptional activators and repressors. In this example, the P_1 promoter is suppressed by transcriptional repressor proteins expressed from the *Repressor 1* gene.

extended and deployed into higher organisms. In fact, several synthetic circuits, such as clocks and switches, were initially developed in bacteria and later translated into mammalian counterparts using analogous design principles[3,7,12]. Additionally, lower-organism test platforms could be endowed with certain features of interest from desired higher-organism hosts. For example, RNA interference–based circuits could be built first in *Saccharomyces cerevisiae* before being used in mammalian cells[86]. In one case, mitochondrial DNA was engineered into *Escherichia coli* before retransplantation into mammalian hosts[87]. Other biomolecular systems and components that are ripe for engineering in lower organisms include chromatin, ubiquitins and proteosomes.

The introduction of synthetic gene networks into higher organisms also runs the risk of compromising natural networks, which have evolved to maintain cellular robustness. Accordingly, methods for simplifying organisms for designing and testing synthetic circuits could be extended to engineer final deployment hosts, making them more conducive to synthetic gene circuits. Ultimately, *in vivo* directed evolutionary methods, based on repeated rounds of mutagenesis and selection within final cellular backgrounds, could be used to identify the optimal performance conditions of synthetic gene networks after their basic functionalities have been validated in earlier test platforms[34].

Next-generation gene networks

Advancing synthetic gene circuits into the realm of higher-order networks with programmable functionality is one of the ultimate goals of synthetic biology. Useful next-generation gene networks should attempt to satisfy at least one of the following criteria: first, yield insights into the principles that guide the operation of natural biological systems; second, highlight design principles and/or provide modules that can be applied to the construction of other useful synthetic circuits; third, advance the tools available for novel scientific experiments; and fourth, enable real-world applications in medicine, industry and/or agriculture. Below, we describe several next-generation gene circuits and discuss their potential utility in the context of the above criteria.

Tunable filters and noise generators. Fine-tuning the performance of a synthetic gene network typically means reengineering its components, be it by replacing or mutating its parts. Networks whose responses can be tuned without the reengineering of its parts, such as the biological version of a tunable electronic filter, would enable more sophisticated cellular-based signal processing. Synthetic transcriptional cascades can exhibit low-pass filter characteristics[16], and artificial gene circuits with negative autoregulation are capable of pushing the noise spectra of their outputs to higher frequencies, where it can be filtered by the low-pass characteristics of a downstream gene cascade[88]. Tunable genetic filters with respect to time could be implemented by tuning RNA and/or protein degradation in autoregulated negative-feedback circuits[66,89–91] (**Fig. 1**). Such circuits would be useful in studying and shaping noise spectra to optimize the performance of artificial gene networks.

Recently, an externally tunable, bacterial bandpass-filter has been described[17] that uses low-pass and high-pass filters in series to derive bandpass activity with respect to enzymes and inducer molecules. These types of filters, when coupled to quorum-sensing modules, can be used for spatial patterning applications[15,17]. They could also be readily extended to complex multicellular pattern formation by engineering a suite of different cells, each carrying filters that respond to different inputs. Synthetic gene circuits based on tunable filters may also make useful platforms for studying cellular differentiation and development, as artificial pattern generation is a model for how natural systems form complex structures[15,17].

Along similar lines, recent developments in stem cell biology have unlocked important potential roles for synthetic gene networks[92]. For example, it has been shown that stochastic fluctuations in protein expression in embryonic stem cells are important for determining differentiation fates[93]. Indeed, stochasticity might be harnessed in differentiation to force population-wide heterogeneity and provide system robustness, though it may also be detrimental if it causes uncontrollable differentiation.

The effects of stochasticity in stem cell differentiation could be studied with synthetic gene circuits that act as tunable noise generators. Lu *et al.*, for instance, considered two such designs for modulating the noise profile of an output protein[94]. This showed that the mean value and variance of the output can be effectively tuned with two external signals, one for regulating transcription and the other for regulating translation, and to a greater extent with three external signals, the third for regulating DNA copy number[94]. By varying noise levels while keeping mean expression levels constant, the thresholds at which gene expression noise yields beneficial versus detrimental effects on stem cell differentiation can be elucidated (J.J.C. and colleagues)[95].

Furthermore, the discovery of induced pluripotent stem cells (iPSCs), based on the controlled expression of four transcription factors (OCT4, SOX2, KLF2 and MYC) in adult fibroblasts, has created a source of patient-specific progenitor cells for engineering[92]. Genetic noise generators and basic control circuits could be used to dissect the mechanism for inducing pluripotency in differentiated adult cells by controlling the expression levels of the four iPSC-dependent transcription factors. Ultimately, these efforts could lead to the development of timing circuits[22] for higher-efficiency stem cell reprogramming.

Lineage commitment to trophectoderm, ectoderm, mesoderm and endoderm pathways are controlled by distinct sets of genes[93], and many interacting factors, including growth factors, extracellular matrices and mechanical forces, play important roles in cellular differentiation[96]. As differentiation pathways become better understood, synthetic gene cascades may be used to program cellular commitment with increased fidelity for applications in biotechnology and regenerative medicine.

Analog-to-digital and digital-to-analog converters. Electrical engineers have used digital processing to achieve reliability and flexibility, even though the world in which digital circuits operate is inherently analog. Although synthetic biological circuits are unlikely to match the computing power of digital electronics, simple circuits inspired by digital

and analog electronics may significantly increase the reliability and programmability of biological behaviors.

For example, biological analog-to-digital converters could translate external analog inputs, such as inducer concentrations or exposure times, into internal digital representations for biological processing. Consider, for instance, a bank of genetic switches with adjustable thresholds (**Fig. 2a**). These switches could be made out of libraries of artificial transcription factors, as described above. This design would perform discretization of analog inputs into levels of digital output. Depending on the level of analog inputs, different genetic pathways could be activated. Cells possessing analog-to-digital converters would be useful as biosensors in medical and environmental settings. For example, whole-cell biosensors[19], resident in the gut, may be engineered to generate different reporter molecules that could be measured in stool depending on the detected level of gastrointestinal bleeding. Expressing different reporter molecules rather than a continuous gradient of a single reporter molecule would yield more reliable and easily interpretable outputs.

Digital-to-analog converters, on the other hand, would translate digital representations back into analog outputs (**Fig. 2b**); such systems could be used to reliably set internal system states. For example, instead of fine-tuning transcriptional activity with varying amounts of chemical inducers, a digital-to-analog converter, composed of a bank of genetic switches, each of which is sensitive to a different inducer, might provide better control. If each activated switch enabled transcription from promoters of varying strengths ($P_{output,3} > P_{output,2} > P_{output,1}$), then digital combinations of inducers could be used to program defined levels of transcriptional activities (**Fig. 2b**). Such a circuit might be useful in biotechnology applications, where reliable expression of different pathways is needed for programming different modes of operation in engineered cells. In addition, digital-to-analog converters may be useful in providing a multiplexed method for probing synthetic circuits. For example, because each analog level is associated with a distinct digital state, a single analog output can allow one to infer the internal digital state of a synthetic gene network (**Fig. 2b**).

Adaptive learning networks. Synthetic gene networks that can learn or adapt to exogenous conditions could provide insight into natural networks and be useful for applications where adaptation to external stimuli may be advantageous, such as autonomous whole-cell biosensors[97,98]. Endogenous biomolecular networks in bacteria can exhibit anticipatory behavior for related perturbations in environmental stimuli[99,100]. This type of behavior and the associated underlying design principles could, in principle, be harnessed to endow transcriptional networks with the ability to learn[97], much like synaptic interconnections between neurons. A basic design that would enable this functionality involves two transcriptional activators (Activator A and Activator B), each of which is expressed in the presence of a different stimulus (**Fig. 3a**). Suppose that both transcriptional activators drive the expression of effector proteins (Effector A and Effector B), which control distinct genetic pathways. When both transcriptional factors are active, indicating the simultaneous presence of the two stimuli, a toggle switch is flipped ON. This creates

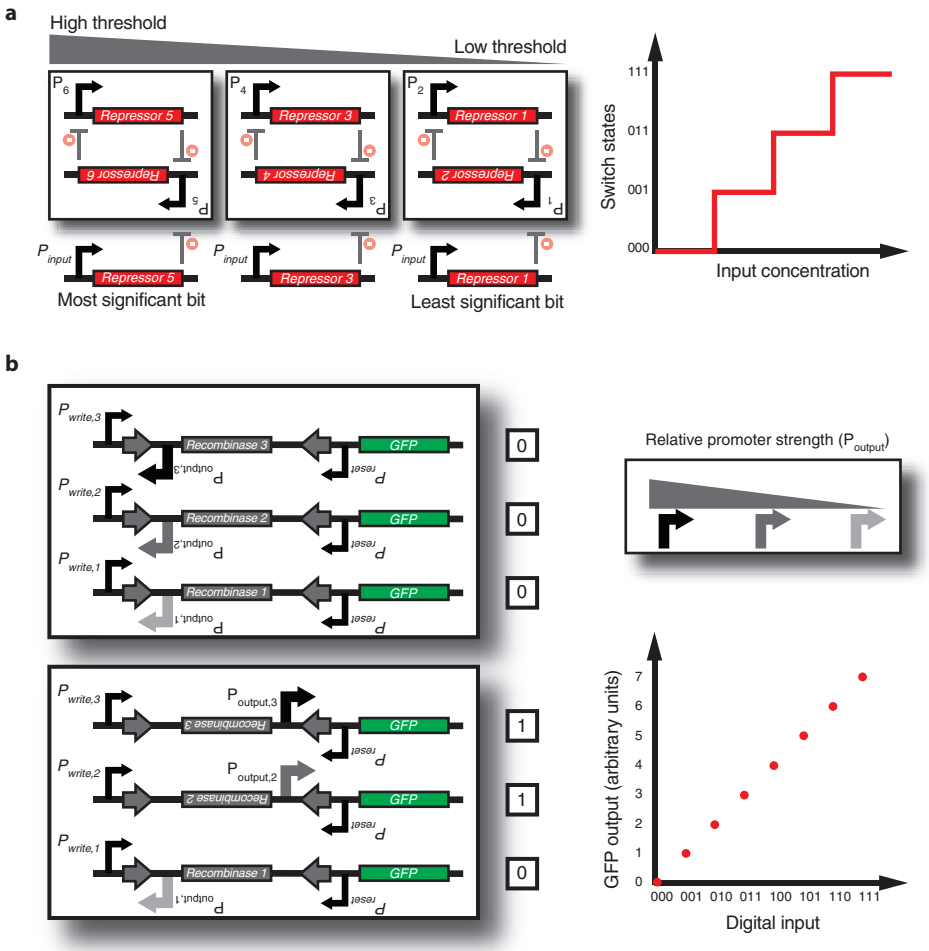


Figure 2 Genetic signal converters. (**a**) Analog-to-digital converter circuit that enables the discretization of analog inputs. The circuit is composed of a bank of toggle switches that have increasing response thresholds so that sequential toggling is achieved as input levels increase. The design could enable different natural or synthetic pathways to be activated depending on distinct input ranges, which may be useful in cell-based biosensing applications. Inputs into promoters and logic operations are shown explicitly except when the promoter (P) name is italicized, which represents an inducible promoter. (**b**) Digital-to-analog converter circuit that enables the programming of defined promoter activity based on combinatorial inputs. The circuit is composed of a bank of recombinase-based switches, known as single-invertase memory modules (SIMMs)[9]. Each SIMM is composed of an inverted promoter and a recombinase gene located between its cognate recognition sites, indicated by the arrows. Upon the combinatorial addition of inducers that activate specific P_{write} promoters, different SIMMs will be flipped, enabling promoters of varying strength to drive green fluorescent protein (GFP) expression. This allows combinatorial programming of different levels of promoter activity.

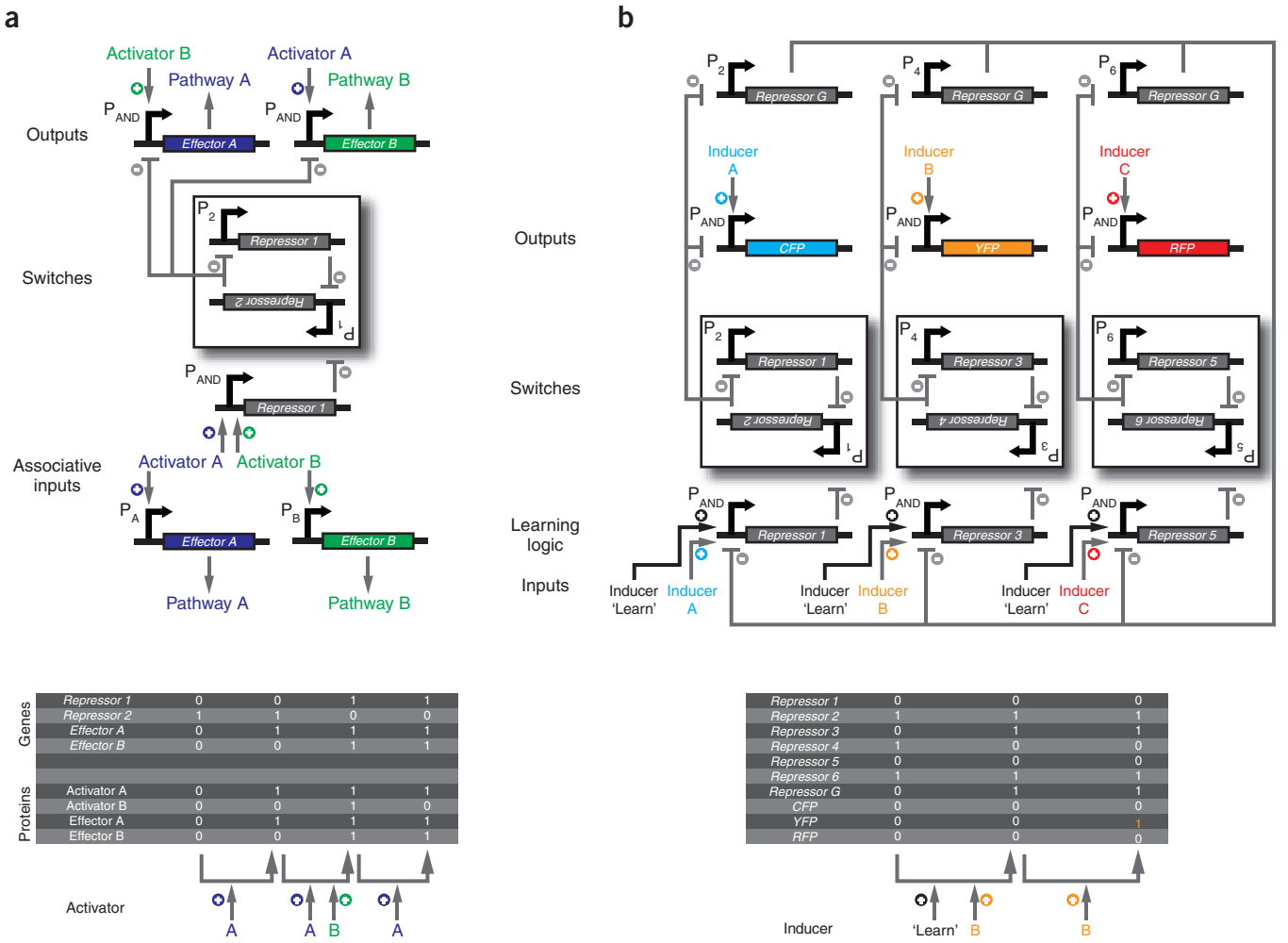


Figure 3 Adaptive learning networks. (**a**) Associative memory circuit enables association between two simultaneous inputs ('Activator A' and 'Activator B') so that the subsequent presence of only a single input can drive its own pathway and the pathway of the other input. Associations between inputs are recorded by a promoter 'P_{AND}' that is activated in the presence of Activator A and Activator B to toggle the memory switch. Inputs into promoters and logic operations are shown explicitly except when the promoter name is italicized, which represents an inducible promoter. (**b**) Winner-take-all circuit allows only one input out of many to be recorded. This effect is achieved by a global repressor protein that gates all inputs and prevents them from being recorded if there has already been an input recorded in memory.

an associative memory. Subsequently, if either of the transcription factors is activated, AND logic between the ON toggle switch and one transcriptional activator produces the effector protein that controls the pathways of the other activators. On the basis of this design, cells could be programmed to associate simultaneous inputs and exhibit anticipatory behavior by activating the pathways of associated stimuli, even in the presence of only one of the stimuli.

In another example of a learning network, one could design bacteria that could be taught 'winner-take-all' behavior in detecting stimuli, similar to cortical neural processing[101]. In this example, bacteria could be exposed to different types of chemical stimuli (Inducers A–C; **Fig. 3b**). An exogenously added inducer (Inducer 'Learn') acts as a trigger for learning and serves as one input into multiple, independent transcriptional AND gates, which possess secondary inputs for detecting the presence of each of the different chemical stimuli. Each gate drives an individual toggle switch that, when flipped, suppresses the flipping of the other switches. This creates a winner-take-all system in which the presence of the most abundant chemical stimuli is recorded. Furthermore, the toggle switch outputs could be fed as inputs into transcriptional AND gates, which once again possess secondary inputs for detecting the presence of the different stimuli. If these gates drive different fluorescent reporters when activated, then the overall system will associate only a single type of stimuli with the learning trigger and respond with an output only in the presence of the single type of stimuli in the future. This system could potentially be adapted to create chemotactic bacteria that 'remember' a particular location or landmark and only respond to the gradient of one chemoattractant.

In more complicated instances of learning networks, it is conceivable that synthetic gene circuits could be designed to adapt on their own, that is, without external mutagenesis or exogenous nucleic acids. For example, transcription-based interconnections could be dynamically reconfigured based on the expression of DNA recombinases[9]. Another design could involve error-prone RNA polymerases, which create mutant RNAs that could be reverse-transcribed and joined back into the genome based on double-stranded breaks created by zinc-finger nucleases. Specificity for where the mutations would occur could be achieved by using promoters that are uniquely read by the error-prone RNA polymerases, such as T7 promoters with a T7 error-prone RNA

polymerase, and zinc-finger nucleases that define where homologous recombination can occur[102]. In this design, enhanced mutagenesis frequencies could be targeted to specific regions of the genome.

Protein-based computational circuits. Beyond DNA- and RNA-based circuits, protein-based synthetic systems have the potential to enable flexible and fast computation through post-translational mechanisms[103–105]. Protein-based circuits are advantageous in that they can be designed to target synthetic activities to subcellular locations[24]. In this way, different sites within the same cell could have different protein circuit states rather than relying solely on shared cellular promoter states, thereby enabling researchers to explore the functional dynamics and consequences of cellular localization. Protein-based designs can also operate on much shorter time scales than genetic circuits because their operation is independent of the transcription and translation machinery[106]. Accordingly, it would be exciting to develop protein-based circuits that can act as rapidly responding logic gates, smart sensors or memory elements.

With regards to this last application, synthetic amyloids could serve as novel components for epigenetic memory circuits. By fusing a yeast prion determinant from Sup35 to the rat glucocorticoid receptor, a transcription factor regulated by steroid hormone, Li and Lindquist[107] demonstrated that the state of transcriptional activity from the fused protein could be affected and inherited stably in an epigenetic fashion. Given the increasing number of identified prionogenic proteins[108], there is an opportunity to create amyloid-based memory systems that transmit functionality from one generation to the next (**Fig. 4**). In these systems, aggregation could be induced by the transient expression of the prionogenic domain (PD), whereas disaggregation could be achieved by expressing protein remodeling factors, such as chaperones (heat shock protein 104). Though this system relies on the transcription and translation of prionogenic and disaggregating factors, it may enable the control of protein effectors that can operate on shorter time scales. For example, enzymes fused to a prionogenic domain may exhibit different activity levels depending on whether they are attached to an amyloid core.

Because genetic circuits and proteins function on different time scales, it would also be worthwhile to develop synthetic networks that couple both modalities. For example, the output of protein-based computation could be stored in recombinase-based memory elements[5,6,9]. It would also be conceivable to couple the two types of networks to harness their varied filtering capabilities. For example, the mitogen-activated protein kinase cascade contains both positive-feedback and negative-feedback loops that enable rapid activation followed by deactivation[109], thus acting like a high-pass filter. On the other hand, transcription- and translation-based gene networks operate on longer time scales rendering them effective low-pass filters. Thus, synthetic kinase/phosphatase circuits that in turn drive gene-based networks could be used to create bandstop filters, which could be coupled with other bandpass filters and used for complex patterning applications.

Intercell signaling circuits and pulse-based processing for genetic oscillators. Robust genetic oscillators with tunable periods have been developed through a combination of experimental and computational efforts[11,12,110]. In addition to shedding light on the design principles guiding the evolution of naturally occurring biological clocks and circadian rhythms, these synthetic oscillators may also have significant utility in biotechnology applications, such as in the synthesis and delivery of biologic drugs. Glucocorticoid secretion, for instance, has a circadian and ultradian pattern of release, resulting in transcriptional pulsing in cells that contain glucocorticoid receptors[111]. Therefore, pulsatile administration of hormones may have therapeutic benefit compared with synthetic hormones applied in a non-ultradian schedule.

An alternative to device-based periodic drug delivery systems could be engineered bacteria that reside in the human gut and synthesize an active drug at fixed time intervals. To realize such an application, one would need to develop and implement intercell signaling circuits for synchronizing and entraining synthetic genetic oscillators[112,113]. Such circuits could be based, for example, on modular components from bacterial quorum sensing systems. Along similar lines, one could engineer light-sensitive[23,24] entrainment circuits for synchronizing mammalian synthetic genetic oscillators. This may help in the construction of oscillators that can faithfully follow circadian rhythms.

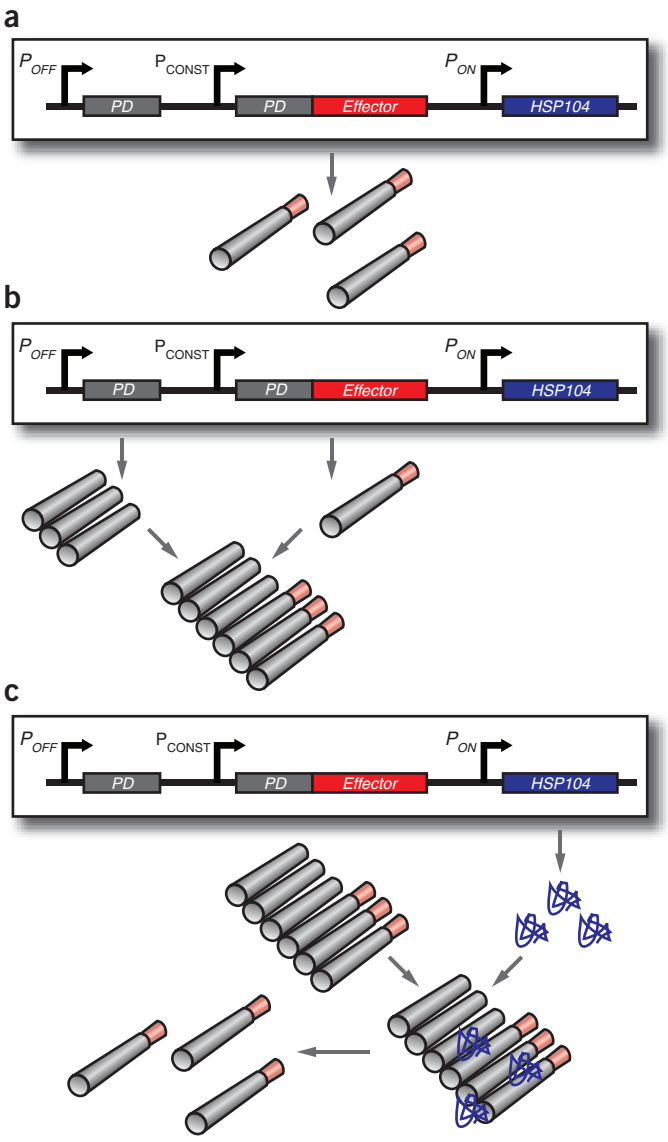


Figure 4 Amyloid-based memory. (**a**) Amyloid-based memory can be implemented by fusing a prionogenic domain (*PD*) to an effector gene, such as a transcriptional activator. (**b**) Overexpressing the prion-determining region via promoter 'P_{OFF}' causes aggregation of the fusion protein, rendering the effector inactive. (**c**) Subsequent overexpression of chaperone proteins (e.g., *HSP104*), which act to disaggregate amyloids, via promoter 'P_{ON}' releases the effector from the amyloid state and enables it to fulfill its function. Inputs into promoters and logic operations are shown explicitly except when the promoter name is italicized, which represents an inducible promoter.

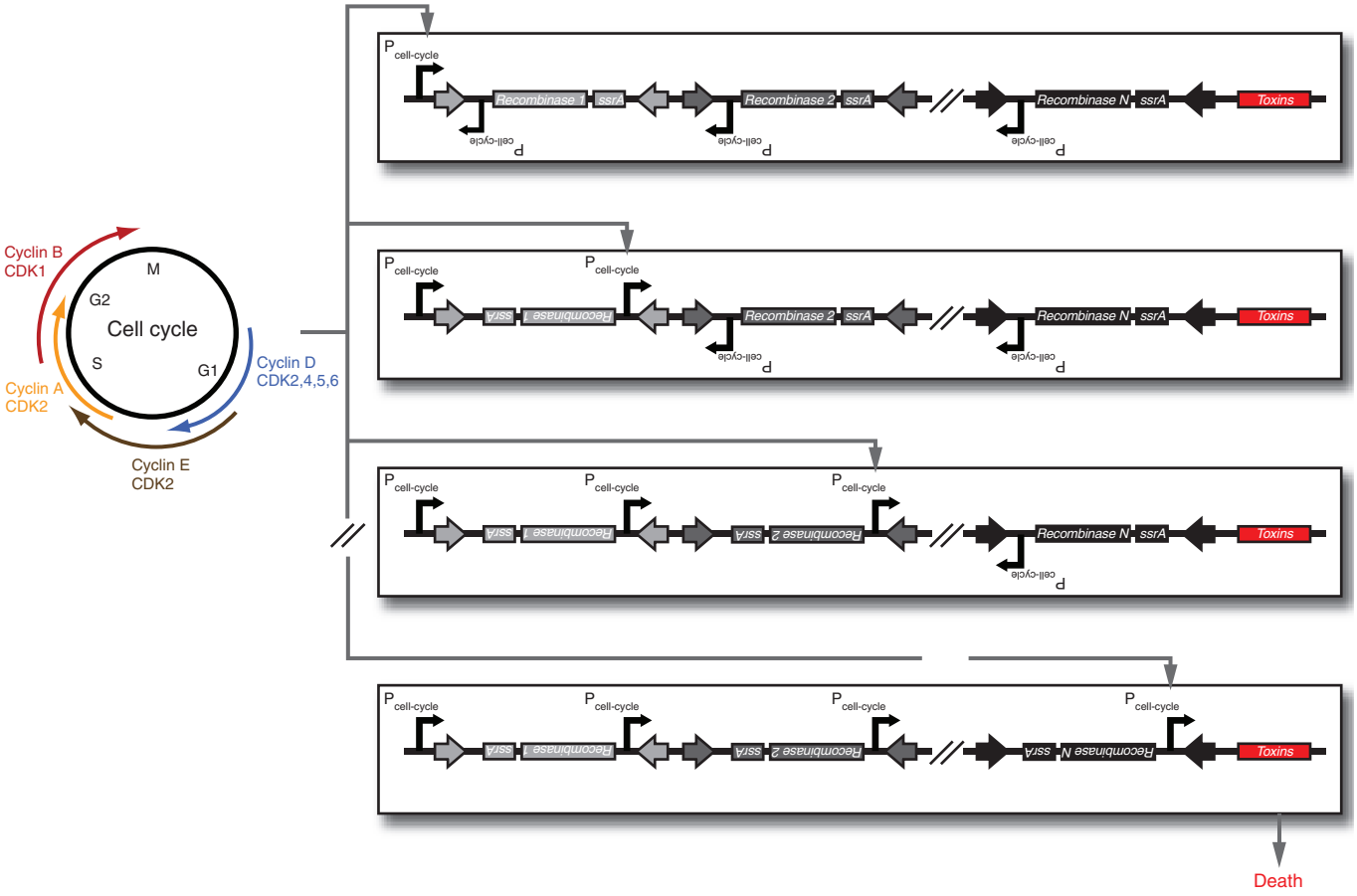


Figure 5 Cell-cycle counter for biological containment. Cell-cycle counting is accomplished with a cascade of single recombinase-based memory units (e.g., SIMMs[9]), each of which is driven by a cell cycle–dependent promoter. After *N* cell-cycle events are counted, the gene circuit unlocks the expression of a toxic protein triggering cell death. Protein degradation tags (*ssrA*) are fused to the recombinase genes to ensure stability of the circuit.

Spike- or pulse-based processing is present in neurons and has been adapted for use in hybrid computation in electrical systems, where interspike times are viewed as analog parameters and spike counts are viewed as digital parameters[114]. In synthetic gene circuits, pulse-based processing may open up exciting new methods for encoding information in engineered cells. For example, instead of transmitting information between cells by means of absolute levels of quorum-sensing molecules, the frequency of a robust genetic oscillator could be modulated. This might be useful in delivering information over longer distances, as frequency information may be less susceptible to decay over distance than absolute molecule levels. Representing signals in this fashion is analogous to frequency modulation encoding in electrical engineering.

Engineered circuits for biological containment. Biological containment, which refers to efforts for ensuring that genetically modified organisms do not spread throughout the natural environment, can be achieved by passive or active techniques. In passive containment, cells are engineered to be dependent on exogenous supplementation to compensate for gene defects, whereas in active containment, cells are engineered to directly express toxic compounds when located outside their target environments[115]. Synthetic genetic counters or timers for programmed cell death could be used as an active containment tool. Counting circuits could, for example, be designed to trigger cell suicide after a defined number of cell cycles or a sequence of events. Recently, we have developed two designs for synthetic counters—a recombinase-based cascade of memory units and a riboregulated transcriptional cascade—that could be adapted for this purpose[9]. In each case, one could incorporate into the counters promoters that are cell cycle–dependent and replace the output reporter proteins with toxic proteins (**Fig. 5**). Circuits of this sort would enable cells to be programmed to have limited, prescribed lifetimes.

Redundant circuits that implement digital logic allowing for the conditional survival of engineered cells only within their desired environments would also potentially reduce the failure rate of biological containment. If a broad set of interoperable parts were developed, multiple layers of control circuits could be built for increased reliability. As in electrical and mechanical engineering, quantitative analysis of failure rates in biological systems would enable improved systems-level design and robustness of synthetic gene networks. This could be accomplished, for example, by subjecting synthetic containment circuits to a variety of stressful conditions that would lead to increased mutation rates and thus improper functioning. Rational and directed evolutionary methods to engineer cells with decreased mutation rates or the application of redundant circuits could then be employed to minimize failure rates.

Whole-cell biosensors and response systems. Programmable cells that act as whole-cell biosensors have been created by interfacing engineered gene networks with the cell's natural regulatory circuitry[19] or with other biological components, such as light-responsive elements[23,24]. The development of novel or reengineered sensory modalities and

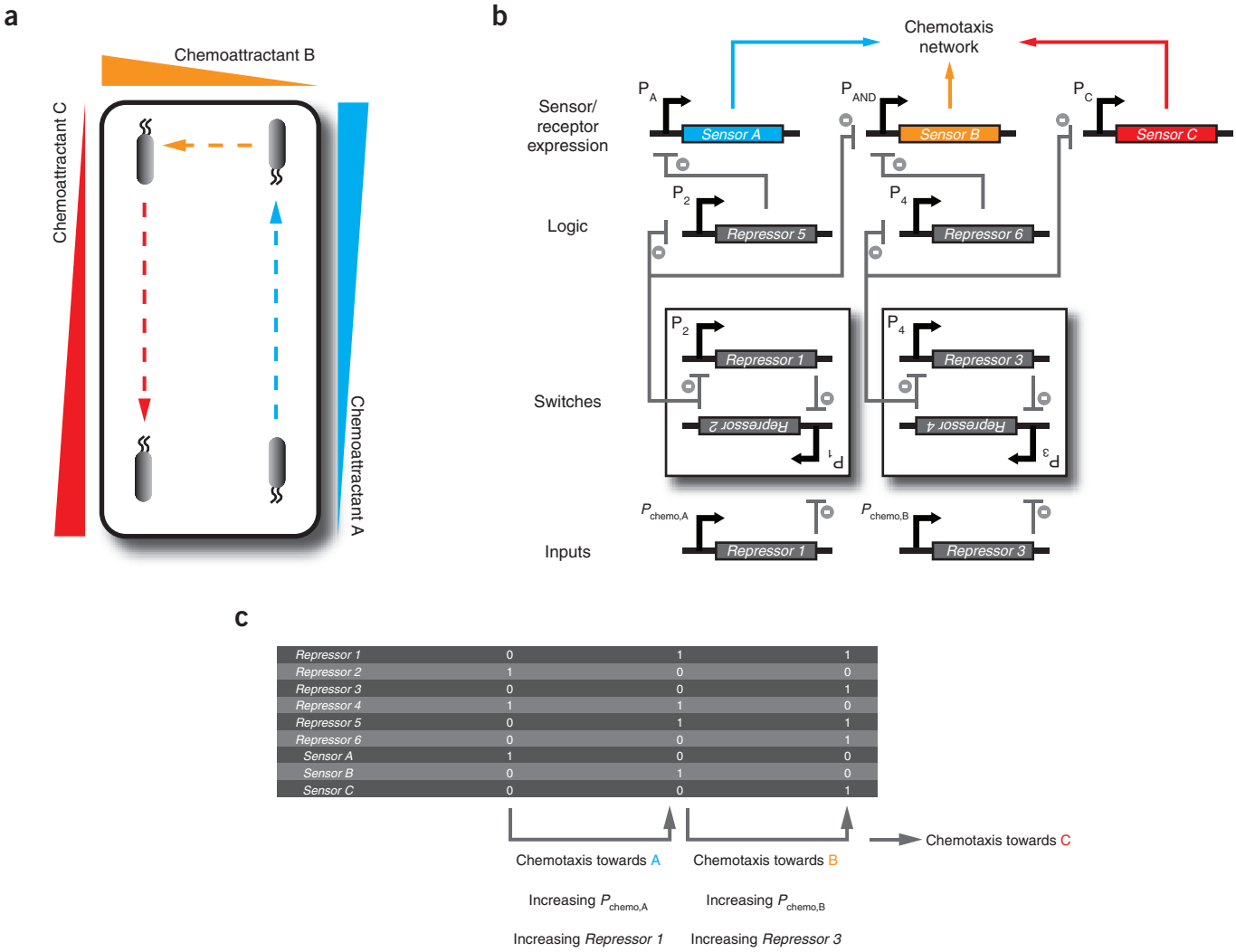


Figure 6 Autonomous chemotaxis. (**a**) Chemotactic environment made up of three chemoattractant gradients (A, B, C). (**b**) The synthetic gene network, whereby toggle switches control the sequential expression of three chemotaxis sensor receptors, for autonomously navigating bacteria down three chemoattractant gradients. Inputs into promoters and logic operations are shown explicitly, except when the promoter name is italicized, which represents an inducible promoter. (**c**) Boolean ON/OFF values for the network genes illustrate the sequential order of operations.

components would expand the range of applications that programmable cells could address. This could involve engineering proteins or RNAs to detect a range of small molecules[116,117], or designing protein-based synthetic signaling cascades by rationally rewiring the protein-protein interactions and output responses of prokaryotic two-component signal transduction systems[118].

The detection of electrical signals or production of biological energy (e.g., mimicking the operation of electrical electrocytes[119]) could also be enabled by incorporating natural or synthetic ion channels into engineered cells. In addition, magneto-responsive bacteria could play useful roles in environmental and medical applications[120]. Synthetic bacteria, designed to form magnetosomes and seek out cancer cells, could be used to enhance imaging, and magnetic bacteria could be engineered to interact with nanoparticles to enhance the targeting of cancer cells. Moreover, the introduction of mechanosensitive ion channels (e.g., MscL from *Mycobacterium tuberculosis* and MscS from *E. coli*) could endow designer cells with the ability to detect mechanical forces[121]. Such cells may be useful *in vivo* sensors for studying cellular differentiation signals or the effects of external stresses on the body.

Ultimately, programmable cells possessing novel sensory modules could be integrated with mechanical, electrical and chemical systems to detect, process and respond to external stimuli, and exploited for a variety of environmental and biomedical applications. For example, bacteria could be engineered to seek out hazardous chemicals or heavy metals in the environment, perform cleanup and return to their origin to report on the number of hazardous sites encountered via analysis by microfluidic devices. To eventually achieve such complex tasks, an intermediate goal might involve programming chemotactic bacteria to swim from waypoint to waypoint. A dish containing gradients of several chemoattractants would constitute the navigational course (**Fig. 6a**).

At the core of this design could be a synthetic gene network made up of a series of sequential toggle switches that control the expression of receptors needed for bacterial chemotaxis toward chemoattractants[122] (**Fig. 6b**). The programmable cells would initially express only a single chemoattractant receptor, and therefore would migrate up only one of the chemoattractant gradients[122]. To determine that a waypoint has been achieved, a threshold-based toggle switch would be turned ON upon reaching a sufficiently high concentration of the chemoattractant.

When the first toggle switch is ON, production of the first chemoattractant receptor would be suppressed and production of a second receptor allowed, resulting in cells swimming up the second chemical gradient. The ON switch would additionally prime the next toggle switch in the series to be switched ON when the second waypoint is reached. When that second toggle flips ON, the previous switch would be flipped OFF to ensure that only one chemoattractant is being followed at a time. The final chemoattractant would lead the bacteria back to its origin so that the engineered cells would complete a multi-stop round trip.

Designer circuits and systems for microbiome engineering. The human microbiome is fertile ground for the application of engineered organisms as scientific tools and therapeutic agents. There are unique bacterial populations residing in distinct locations in the human body that are perturbed in disease states[123,124]. Each represents an exciting opportunity for reengineering the human microbiome and designing targeted therapeutics for a range of conditions, including dermatologic, genitourinary, gastrointestinal, metabolic and immunologic diseases[125–127].

Recently, bacteria have been engineered to infiltrate cellular communities for the purposes of delivering probes, gene circuits or chemicals[128,129]. In a similar fashion, bacteriophages carrying synthetic gene circuits could transform existing microbiome bacteria with new functionalities. For instance, given that anaerobic bacteria are known to migrate to hypoxic and necrotic regions of solid tumors[130], bacteriophages could be designed to infect cancer-targeting bacteria. These bacteriophages could encode conditional expression of chemotherapeutic agents using synthetic logic gates or switches that are coupled to environmental sensors.

Bhatia and colleagues[131] recently have developed nanoparticles that perform Boolean logic based on proteolytic activity. Viruses that infect tumor cells or bacteria could carry synthetic gene circuits that regulate in a programmable fashion the expression of enzymes that trigger nanoparticle activity. In these ways, one could develop targeted therapies against cancer or infectious diseases that exploit the human microbiome and synthetic gene networks.

Switchboard for dynamically controlling the expression of multiple genes. Engineered cells have long been used to produce recombinant proteins and chemicals for the biotechnology industry, and one of the major applications of synthetic biology to date has been in enhancing microbial production of biofuels[132] and biomaterials[133–136]. Improving production from cells involves numerous engineering decisions related to the entire organism, including codon optimization, choosing whether or not to export recombinant proteins[137], rational or evolutionary methods for improving metabolic yields[138,139], and optimization of growth conditions. Often some or all of the genes required for production are non-optimal for bacterial expression and contain repetitive sequences that are unstable in bacterial hosts. Whole-gene synthesis techniques are increasingly being used to optimize coding sequences for recombinant production[136].

These innovative approaches, as well as more traditional knockout techniques, introduce hard-wired changes into the genomes of interest. However, for many industrial and bioprocess applications, there is a need to dynamically modulate and control the expression of multiple genes, depending upon the state of the bioreactor. These situations would benefit from the development of a synthetic switchboard, one that could tune the expression of many different genes simultaneously and independently. Such a switchboard could be made up of a series of adjustable threshold genetic switches, riboregulators or riboswitches, and designed to respond to different environmental and intracellular variables, such as pH, light intensity and the metabolic state of the cell. The switchboard design, which would integrate novel sensory modalities with tunable, interoperable genetic circuits, would have broad functionality. It could be programmed, for example, to shift carbon flux between different pathways depending upon cellular conditions, thereby optimizing the production of biofuels, specialty chemicals and other materials.

Conclusions

The past decade has witnessed the power of intelligently applying engineering principles to biology in the development of many exciting, artificial gene circuits and biomolecular systems. We are convinced that next-generation synthetic gene networks will advance understanding of natural systems, provide new biological modules and create new tools that will enable the construction of even more complex systems. Most importantly, if the current pace of progress in synthetic biology continues, real-world applications in fields such as medicine, biotechnology, bioremediation and bioenergy will be realized.

ACKNOWLEDGMENTS

We would like to thank the Howard Hughes Medical Institute and the National Institutes of Health Director's Pioneer Award Program for their financial support. We also thank the reviewers for their insights and suggestions.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturebiotechnology/.

Published online at http://www.nature.com/naturebiotechnology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
2. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
3. Kramer, B.P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22**, 867–870 (2004).
4. Isaacs, F.J., Hasty, J., Cantor, C.R. & Collins, J.J. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA* **100**, 7714–7719 (2003).
5. 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).
6. 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).
7. 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).
8. Ajo-Franklin, C.M. *et al.* Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276 (2007).
9. Friedland, A.E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
10. Fung, E. *et al.* A synthetic gene-metabolic oscillator. *Nature* **435**, 118–122 (2005).
11. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
12. Tigges, M., Marquez-Lago, T.T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312 (2009).
13. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.* **25**, 795–801 (2007).
14. Win, M.N. & Smolke, C.D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
15. 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).
16. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. USA* **102**, 3581–3586 (2005).
17. Sohka, T. *et al.* An externally tunable bacterial band-pass filter. *Proc. Natl. Acad. Sci. USA* **106**, 10135–10140 (2009).
18. Bayer, T.S. & Smolke, C.D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.* **23**, 337–343 (2005).

19. Kobayashi, H. *et al.* Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA* **101**, 8414–8419 (2004).
20. 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).
21. You, L., Cox, R.S. III, Weiss, R. & Arnold, F.H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
22. Ellis, T., Wang, X. & Collins, J.J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**, 465–471 (2009).
23. Levskaya, A. *et al.* Synthetic biology: engineering *Escherichia coli* to see light. *Nature* **438**, 441–442 (2005).
24. Levskaya, A., Weiner, O.D., Lim, W.A. & Voigt, C.A. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461**, 997–1001 (2009).
25. Tabor, J.J. *et al.* A synthetic genetic edge detection program. *Cell* **137**, 1272–1281 (2009).
26. 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).
27. Lu, T.K. & Collins, J.J. Dispersing biofilms with engineered enzymatic bacteriophage. *Proc. Natl. Acad. Sci. USA* **104**, 11197–11202 (2007).
28. Lu, T.K. & Collins, J.J. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc. Natl. Acad. Sci. USA* **106**, 4629–4634 (2009).
29. Ro, D.K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
30. Czar, M.J., Cai, Y. & Peccoud, J. Writing DNA with GenoCAD. *Nucleic Acids Res.* **37**, W40–W47 (2009).
31. Guido, N.J. *et al.* A bottom-up approach to gene regulation. *Nature* **439**, 856–860 (2006).
32. Shetty, R.P., Endy, D. & Knight, T.F. Jr. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5 (2008).
33. Carr, P.A. & Church, G.M. Genome engineering. *Nat. Biotechnol.* **27**, 1151–1162 (2009).
34. Wang, H.H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
35. Purnick, P.E. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
36. Lucks, J.B., Qi, L., Whitaker, W.R. & Arkin, A.P. Toward scalable parts families for predictable design of biological circuits. *Curr. Opin. Microbiol.* **11**, 567–573 (2008).
37. Maeder, M.L. *et al.* Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* **31**, 294–301 (2008).
38. Beerli, R.R., Dreier, B. & Barbas, C.F. III. Positive and negative regulation of endogenous genes by designed transcription factors. *Proc. Natl. Acad. Sci. USA* **97**, 1495–1500 (2000).
39. Park, K.S. *et al.* Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat. Biotechnol.* **21**, 1208–1214 (2003).
40. Isaacs, F.J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* **22**, 841–847 (2004).
41. Win, M.N., Liang, J.C. & Smolke, C.D. Frameworks for programming biological function through RNA parts and devices. *Chem. Biol.* **16**, 298–310 (2009).
42. Buchholz, F., Angrand, P.O. & Stewart, A.F. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.* **16**, 657–662 (1998).
43. Kilby, N.J., Snaith, M.R. & Murray, J.A. Site-specific recombinases: tools for genome engineering. *Trends Genet.* **9**, 413–421 (1993).
44. Santoro, S.W. & Schultz, P.G. Directed evolution of the site specificity of Cre recombinase. *Proc. Natl. Acad. Sci. USA* **99**, 4185–4190 (2002).
45. Groth, A.C. & Calos, M.P. Phage integrases: biology and applications. *J. Mol. Biol.* **335**, 667–678 (2004).
46. Kaplan, S., Bren, A., Dekel, E. & Alon, U. The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Mol. Syst. Biol.* **4**, 203 (2008).
47. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. USA* **102**, 12678–12683 (2005).
48. Cox, R.S. III, Surette, M.G. & Elowitz, M.B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3**, 145 (2007).
49. Hammer, K., Mijakovic, I. & Jensen, P.R. Synthetic promoter libraries–tuning of gene expression. *Trends Biotechnol.* **24**, 53–55 (2006).
50. Jensen, P.R. & Hammer, K. Artificial promoters for metabolic optimization. *Biotechnol. Bioeng.* **58**, 191–195 (1998).
51. 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).
52. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
53. Andersen, J.B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246 (1998).
54. Nevozhay, D., Adams, R.M., Murphy, K.F., Josic, K. & Balazsi, G. Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 5123–5128 (2009).
55. Chandran, D., Bergmann, F.T. & Sauro, H.M. TinkerCell: modular CAD tool for synthetic biology. *J Biol Eng* **3**, 19 (2009).
56. Kaznessis, Y.N. Computational methods in synthetic biology. *Biotechnol. J.* **4**, 1392–1405 (2009).
57. Hasty, J., McMillen, D., Isaacs, F. & Collins, J.J. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nat. Rev. Genet.* **2**, 268–279 (2001).
58. Khalil, A.S. *et al.* Single M13 bacteriophage tethering and stretching. *Proc. Natl. Acad. Sci. USA* **104**, 4892–4897 (2007).
59. Svoboda, K. & Block, S.M. Biological applications of optical forces. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 247–285 (1994).
60. Bustamante, C., Bryant, Z. & Smith, S.B. Ten years of tension: single-molecule DNA mechanics. *Nature* **421**, 423–427 (2003).
61. Khalil, A.S. *et al.* Kinesin's cover-neck bundle folds forward to generate force. *Proc. Natl. Acad. Sci. USA* **105**, 19247–19252 (2008).
62. Neuman, K.C. & Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **5**, 491–505 (2008).
63. Lee, S.K. *et al.* Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. *Appl. Environ. Microbiol.* **73**, 5711–5715 (2007).
64. Gulati, S. *et al.* Opportunities for microfluidic technologies in synthetic biology. *J. R. Soc. Interface* **6** Suppl 4, S493–S506 (2009).
65. Bennett, M.R. & Hasty, J. Microfluidic devices for measuring gene network dynamics in single cells. *Nat. Rev. Genet.* **10**, 628–638 (2009).
66. Simpson, M.L., Cox, C.D. & Sayler, G.S. Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl. Acad. Sci. USA* **100**, 4551–4556 (2003).
67. Mettetal, J.T., Muzzey, D., Gomez-Uribe, C. & van Oudenaarden, A. The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. *Science* **319**, 482–484 (2008).
68. Bennett, M.R. *et al.* Metabolic gene regulation in a dynamically changing environment. *Nature* **454**, 1119–1122 (2008).
69. Malmstrom, J. *et al.* Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460**, 762–765 (2009).
70. Summerer, D. *et al.* A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. USA* **103**, 9785–9789 (2006).
71. Wang, J., Xie, J. & Schultz, P.G. A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.* **128**, 8738–8739 (2006).
72. Michalet, X. *et al.* Quantum dots for live cells, *in vivo* imaging, and diagnostics. *Science* **307**, 538–544 (2005).
73. Hamad-Schifferli, K., Schwartz, J.J., Santos, A.T., Zhang, S. & Jacobson, J.M. Remote electronic control of DNA hybridization through inductive coupling to an attached metal nanocrystal antenna. *Nature* **415**, 152–155 (2002).
74. Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
75. Misra, N. *et al.* Bioelectronic silicon nanowire devices using functional membrane proteins. *Proc. Natl. Acad. Sci. USA* **106**, 13780–13784 (2009).
76. Weber, W. & Fussenegger, M. Engineering of synthetic mammalian gene networks. *Chem. Biol.* **16**, 287–297 (2009).
77. Gibson, D.G. *et al.* Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215–1220 (2008).
78. Glass, J.I. *et al.* Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. USA* **103**, 425–430 (2006).
79. Lartigue, C. *et al.* Genome transplantation in bacteria: changing one species to another. *Science* **317**, 632–638 (2007).
80. Lartigue, C. *et al.* Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693–1696 (2009).
81. Carrera, J., Rodrigo, G. & Jaramillo, A. Towards the automated engineering of a synthetic genome. *Mol. Biosyst.* **5**, 733–743 (2009).
82. Rackham, O. & Chin, J.W. A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol.* **1**, 159–166 (2005).
83. Wang, K., Neumann, H., Peak-Chew, S.Y. & Chin, J.W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **25**, 770–777 (2007).
84. An, W. & Chin, J.W. Synthesis of orthogonal transcription-translation networks. *Proc. Natl. Acad. Sci. USA* **106**, 8477–8482 (2009).
85. Wang, Q., Parrish, A.R. & Wang, L. Expanding the genetic code for biological studies. *Chem. Biol.* **16**, 323–336 (2009).
86. Drinnenberg, I.A. *et al.* RNAi in Budding Yeast. *Science* **326**, 544–550 (2009).
87. Yoon, Y.G. & Koob, M.D. Efficient cloning and engineering of entire mitochondrial genomes in *Escherichia coli* and transfer into transcriptionally active mitochondria. *Nucleic Acids Res.* **31**, 1407–1415 (2003).
88. Austin, D.W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439**, 608–611 (2006).
89. Mar, D.J., Chow, C.C., Gerstner, W., Adams, R.W. & Collins, J.J. Noise shaping in populations of coupled model neurons. *Proc. Natl. Acad. Sci. USA* **96**, 10450–10455 (1999).
90. McGinness, K.E., Baker, T.A. & Sauer, R.T. Engineering controllable protein degradation. *Mol. Cell* **22**, 701–707 (2006).
91. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G. & Wandless, T.J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995–1004 (2006).
92. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
93. Macarthur, B.D., Ma'ayan, A. & Lemischka, I.R. Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* **10**, 672–681 (2009).

94. Lu, T., Ferry, M., Weiss, R. & Hasty, J. A molecular noise generator. *Phys. Biol.* **5**, 036006 (2008).
95. Blake, W.J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865 (2006).
96. Discher, D.E., Mooney, D.J. & Zandstra, P.W. Growth factors, matrices, and forces combine and control stem cells. *Science* **324**, 1673–1677 (2009).
97. Fernando, C.T. *et al.* Molecular circuits for associative learning in single-celled organisms. *J. R. Soc. Interface* **6**, 463–469 (2009).
98. Fritz, G., Buchler, N.E., Hwa, T. & Gerland, U. Designing sequential transcription logic: a simple genetic circuit for conditional memory. *Syst. Synth. Biol.* **1**, 89–98 (2007).
99. Tagkopoulos, I., Liu, Y.C. & Tavazoie, S. Predictive behavior within microbial genetic networks. *Science* **320**, 1313–1317 (2008).
100. Mitchell, A. *et al.* Adaptive prediction of environmental changes by microorganisms. *Nature* **460**, 220–224 (2009).
101. Lee, D.K., Itti, L., Koch, C. & Braun, J. Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.* **2**, 375–381 (1999).
102. Brakmann, S. & Grzeszik, S. An error-prone T7 RNA polymerase mutant generated by directed evolution. *ChemBioChem* **2**, 212–219 (2001).
103. Yeh, B.J., Rutigliano, R.J., Deb, A., Bar-Sagi, D. & Lim, W.A. Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. *Nature* **447**, 596–600 (2007).
104. Dueber, J.E., Mirsky, E.A. & Lim, W.A. Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat. Biotechnol.* **25**, 660–662 (2007).
105. 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).
106. Goldberg, S.D., Derr, P., DeGrado, W.F. & Goulian, M. Engineered single- and multicell chemotaxis pathways in *E. coli*. *Mol. Syst. Biol.* **5**, 283 (2009).
107. Li, L. & Lindquist, S. Creating a protein-based element of inheritance. *Science* **287**, 661–664 (2000).
108. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146–158 (2009).
109. Bhalla, U.S., Ram, P.T. & Iyengar, R. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science* **297**, 1018–1023 (2002).
110. Tsai, T.Y. *et al.* Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* **321**, 126–129 (2008).
111. Stavreva, D.A. *et al.* Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nat. Cell Biol.* **11**, 1093–1102 (2009).
112. McMillen, D., Kopell, N., Hasty, J. & Collins, J.J. Synchronizing genetic relaxation oscillators by intercell signaling. *Proc. Natl. Acad. Sci. USA* **99**, 679–684 (2002).
113. Garcia-Ojalvo, J., Elowitz, M.B. & Strogatz, S.H. Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proc. Natl. Acad. Sci. USA* **101**, 10955–10960 (2004).
114. Sarpeshkar, R. & O'Halloran, M. Scalable hybrid computation with spikes. *Neural Comput.* **14**, 2003–2038 (2002).
115. Molin, S. *et al.* Suicidal genetic elements and their use in biological containment of bacteria. *Annu. Rev. Microbiol.* **47**, 139–166 (1993).
116. Looger, L.L., Dwyer, M.A., Smith, J.J. & Hellinga, H.W. Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185–190 (2003).
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. Skerker, J.M. *et al.* Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043–1054 (2008).
119. Xu, J. & Lavan, D.A. Designing artificial cells to harness the biological ion concentration gradient. *Nat. Nanotechnol.* **3**, 666–670 (2008).
120. Jogler, C. & Schuler, D. Genomics, genetics, and cell biology of magnetosome formation. *Annu. Rev. Microbiol.* **63**, 501–521 (2009).
121. Booth, I.R., Edwards, M.D., Black, S., Schumann, U. & Miller, S. Mechanosensitive channels in bacteria: signs of closure? *Nat. Rev. Microbiol.* **5**, 431–440 (2007).
122. Falke, J.J., Bass, R.B., Butler, S.L., Chervitz, S.A. & Danielson, M.A. The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu. Rev. Cell Dev. Biol.* **13**, 457–512 (1997).
123. Gao, Z., Tseng, C.H., Strober, B.E., Pei, Z. & Blaser, M.J. Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. *PLoS ONE* **3**, e2719 (2008).
124. Grice, E.A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* **324**, 1190–1192 (2009).
125. Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R. & Gordon, J.I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* **6**, 776–788 (2008).
126. Turnbaugh, P.J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
127. Turnbaugh, P.J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
128. Steidler, L. *et al.* Treatment of murine colitis by Lactococcus lactis secreting interleukin-10. *Science* **289**, 1352–1355 (2000).
129. Braat, H. *et al.* A phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease. *Clin. Gastroenterol. Hepatol.* **4**, 754–759 (2006).
130. Wei, M.Q., Mengesha, A., Good, D. & Anne, J. Bacterial targeted tumour therapy-dawn of a new era. *Cancer Lett.* **259**, 16–27 (2008).
131. von Maltzahn, G. *et al.* Nanoparticle self-assembly gated by logical proteolytic triggers. *J. Am. Chem. Soc.* **129**, 6064–6065 (2007).
132. Stephanopoulos, G. Challenges in engineering microbes for biofuels production. *Science* **315**, 801–804 (2007).
133. Teule, F. *et al.* A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning. *Nat. Protoc.* **4**, 341–355 (2009).
134. Slotta, U. *et al.* Spider silk and amyloid fibrils: a structural comparison. *Macromol. Biosci.* **7**, 183–188 (2007).
135. Rammensee, S., Slotta, U., Scheibel, T. & Bausch, A.R. Assembly mechanism of recombinant spider silk proteins. *Proc. Natl. Acad. Sci. USA* **105**, 6590–6595 (2008).
136. Widmaier, D.M. *et al.* Engineering the Salmonella type III secretion system to export spider silk monomers. *Mol. Syst. Biol.* **5**, 309 (2009).
137. Choi, J.H. & Lee, S.Y. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **64**, 625–635 (2004).
138. Klein-Marcuschamer, D. & Stephanopoulos, G. Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains. *Proc. Natl. Acad. Sci. USA* **105**, 2319–2324 (2008).
139. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).

nature biotechnology

REVIEW

Genome engineering

Peter A Carr[1] & George M Church[2]

For more than 50 years, those engineering genetic material have pursued increasingly challenging targets. During that time, the tools and resources available to the genetic engineer have grown to encompass new extremes of both scale and precision, opening up new opportunities in genome engineering. Today, our capacity to generate larger *de novo* assemblies of DNA is increasing at a rapid pace (with concomitant decreases in manufacturing cost). We are also witnessing potent demonstrations of the power of merging randomness and selection with engineering approaches targeting large numbers of specific sites within genomes. These developments promise genetic engineering with unprecedented levels of design originality and offer new avenues to expand both our understanding of the biological world and the diversity of applications for societal benefit.

Our capacity to understand and employ living systems has been intimately enmeshed with our ability to manipulate and test the instructive molecules. The ancient manipulation and testing of billion-base-pair DNA systems is evident in the diversity of dog breeds (spanning 3 logs in mass) and agricultural species relative to their wild ancestors. Moving in the direction of specific genetic control, the awesome power of merging chemistry with biology in the 1960s was evident in the use of synthetic oligonucleotides (oligos) to elucidate the fundamentals of the genetic code[1,2] and in the 1970s to produce the first synthetic gene[3] and first synthetic gene functionally tested *in vivo*[4].

Since that time, benchmarks in the capacity to synthesize, manipulate and analyze DNA constructs have been achieved at exponential scales, in a manner reminiscent of Moore's Law[5] for improvement in integrated circuit density. **Figure 1** displays milestones in the *de novo* synthesis of DNA, from the first dinucleotide, dTdT, reported in 1955 by Michelson and Todd[6], to the recent construction of a compact microbial genome (*Mycoplasma genitalium*; 582,970 bp)[7]. **Figure 2** charts the improvement over time in the efficiency of DNA sequencing (in base pairs per dollar) as well as synthesis—both oligos and double-stranded DNA—trends that have also been noted for their exponential behavior[8–10]. The tendency of some of these trends to increase in rate has been called the 'law of accelerating returns', emphasizing that this acceleration can go beyond even normal exponential growth[10] (inflected upward on a log-linear graph).

Nevertheless, a paradoxical gap exists between our ability to synthesize and our ability to design valuable novel constructs. We can now produce oligos at 100 kbp/dollar and sequence DNA at 1 Mbp/dollar, but final gene-length DNA constructs are 2 bp/dollar. Getting a novel DNA construct to work as intended is a nontrivial process. Even modest deviations from natural genes cannot be taken for granted as functional, and must be tested thoroughly; much more so for *de novo* designs (in contrast, consider that combinatorial libraries of oligos can be constructed at 10^{14} bp/dollar[11], although the density of these pools for a given function can be low).

[1]Massachusetts Institute of Technology, Media Lab, Center for Bits and Atoms, Massachusetts, Cambridge, USA. [2]Harvard Medical School, Department of Genetics, Massachusetts, Cambridge, USA. Correspondence should be addressed to P.A.C. (carr@media.mit.edu)

Published online 9 December 2009; doi:10.1038/nbt1590

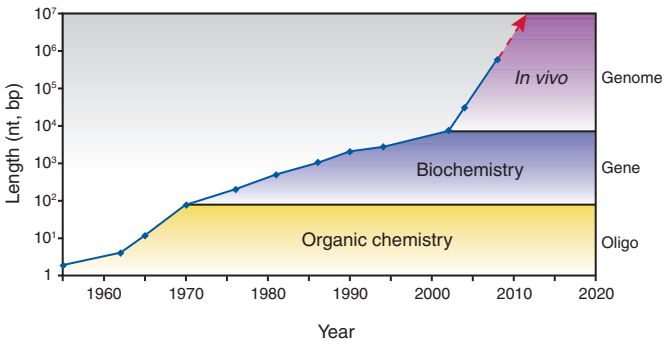


Figure 1 Milestones in the sizes of *de novo* synthesized DNA. Length is nucleotides (nt) for oligos before 1970, base pairs (bp) for double-stranded DNA from 1970 on. *In vitro* biochemical processing steps enabled the leap from oligos to genes, and *in vivo* processing steps (multiple cycles of cloning, sequencing and assembly) made possible the leap from genes to genomes. Future extensions of these tiers may include complex microbial communities or tissue organization. Data for this graph can be found in **Supplementary Table 1**.

Thus, although specific design techniques are improving, we expect that two of biology's special tools—variation and selection—will maintain an important role in the engineer's repertoire.

In discussing the idea of genome engineering, we apply this working definition: extensive and intentional genetic modification of a replicating system for a specific purpose. We leave the terms 'extensive' and 'replicating system' purposefully broad. On a practical level, we include work on viruses and bacteriophage, such as factoring a phage genome into many separate parts, and reorganizing and testing the new combinations[12]. More examples are given in **Figure 3**. We exclude smaller replicative units, such as viroids (as small as 220 nucleotides (nt) in length). The tools and research interests of genome engineering also overlap those of genome-scale engineering. For this latter term, we refer to engineering of genetic systems on a similar scale (e.g., hundreds of genetic components or more) but not integrated into a single replicating (typically cellular) system. Examples of genome-scale engineering include producing and characterizing hundreds of different versions

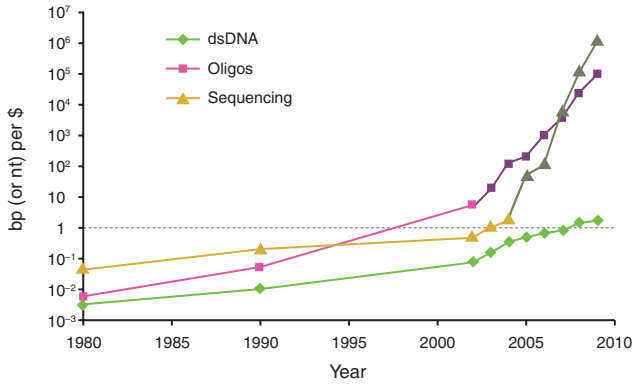


Figure 2 Efficiency trends in synthesis and sequencing over the past 30 years (base pairs per dollar). Double-stranded DNA synthesis ('gene synthesis') while improving rapidly, seems to lag behind the other two trends. The accelerated improvement in sequencing and oligo synthesis this past decade has been predicated on new miniaturization technologies (next-generation sequencing and microarray synthesis, respectively) to where the critical events take place on surface features measured in μm^2. These transitions in technology are noted by a change to a darker line color. Commercial gene synthesis relies on both oligo synthesis (building blocks) and sequencing (verification and error control) but has yet to take effective advantage of these miniaturized formats. Some proofs of principle have been demonstrated[41,51,113]. Data for this graph are detailed in **Supplementary Table 2**.

of a gene *in vitro*, or saturating a genome with single-gene knockouts (thousands of separate strains each with one modification)[13,14].

Genetic engineering as applied over the past several decades has most often employed small numbers of specific components (e.g., a single promoter and ribosome binding site coupled to a protein-encoding gene). Over the past decade, advanced designs have been engineered using larger numbers of components and with more complex interdependencies between them (see ref. 15). Several examples discussed below refer to current work at these scales (e.g., a dozen components), which in turn point the way toward future designs that may approach the genome scale.

Thus, genome engineering is genetic engineering applied to genomes (or at least large portions thereof). The tools used for this purpose are often those developed for smaller-scale genetic engineering, and applied in high-throughput fashion. In addition, genome engineering requires new technology specifically suited to that scale. For example, *de novo* construction of DNA molecules of up to a few thousand base pairs has relied on organic chemical and biochemical procedures. To generate an entire microbial genome, however, requires new methods for combining those smaller synthetic pieces (as detailed in ref. 7).

Genome engineering is in its infancy. The new techniques that have enabled initial work are modest compared with the needs for more tools at all stages: design, DNA construction and manipulation, implementation and testing, and debugging. Similarly, although potential applications are enticing, they are largely unproven at this point in time. As we discuss both these ideas and current progress, we begin with the motivations for expanding current gene and gene systems work to the genome scale, along with some goals that can only be achieved by dramatic engineering (or reengineering) of genomes.

Motivations for genome engineering

What are the factors that will continue driving DNA engineering toward increasingly larger and more complex designs? There is interplay between motivating applications and the technical advances, which enable larger scales while reducing costs. The pursuit of challenging goals leads to new technology and the availability of that new technology encourages more ambitious pursuits. So why do we build genetic systems? Put another way, if you could design and build a genome, what would you want to make?

Build to understand. The Richard Feynman quote "what I cannot create I do not understand"[16] is a favorite among synthetic biologists—and for good reason. Endy[17] has pointed out that for some, synthetic biology is the pursuit of comprehending biological systems by trying to engineer them. (And we defer to that reference for greater exposition on the term 'synthetic biology'.) Much of the history of genetic engineering has been for the sake of understanding the molecular workings of life, frequently at the level of small numbers of parts (e.g., putting the coding sequence for a protein in a new genetic context such as a plasmid for easy manipulation and study). The complexity of such designs is increasing[15,18]. For example, genetic circuits recently have been constructed to produce pattern formation in microbial communities[19]—a model system for studying the basic principles influencing developmental patterns in higher organisms. Furthermore, a central goal of the *M. genitalium* genome synthesis has been to produce a construction technology to examine minimal gene sets required for life[20].

Build for production. Living systems produce a staggering array of products tailored to human needs, including foodstuffs, materials and clothing. Recent years have seen substantial progress in metabolic engineering of microbes—combining, modifying and tuning many genes from different organisms for the sake of producing medicines[21] and biofuels[22]. At the genome level, there is much interest in engineering a cellular 'chassis' for the optimal performance of such metabolic systems, involving large numbers of modifications to a microbial genome.

Build for protection. Genetic systems have also been designed to harness microbes as biosensors for various types of threats[23,24] and bioremediation[25]. Designs are currently in development for systems that allow microbes to hunt and destroy cancer cells[26,27] and instruct one's own cells to minimize the risk of septic shock[28]. An example of genome-wide engineering in this area would be the production of organisms with fundamentally altered codon usage—'orthogonal' genomes incapable of correctly translating genetic messages from other organisms and vice versa. At the scale of microbial genomes, this feature could prevent an engineered laboratory strain from using acquired genes to improve its fitness (e.g., antibiotic resistance genes) and from donating its specially engineered features to wild organisms. Plant genomes (e.g., crops) with this feature would be resistant to many wild pathogens (and uniquely susceptible to designed 'watchdog' pathogens). They would also be incapable of outcrossing with wild strains or conventional crops.

Build to creatively explore. An excellent array of explorations can be found at the website of the International Genetically Engineered Machines (iGEM) competition (http://www.igem.org/). These projects stand out as the accomplishments of interdisciplinary teams of undergraduates, operating in a time frame (months) conventionally considered brief for these types of efforts. There are too many intriguing applications to list here, but they include: (i) programming cells to communicate their growth state by emitting different odors; (ii) employing microbes as a photographic print medium; and (iii) many examples of genetically encoded logic and computation. Although the individual projects often fall into one or more of the above categories of understanding, production or protection, the entire undertaking serves as an experiment in the education and motivation of a new generation of synthetic biologists. In doing so, the students seize the opportunity to explore such questions as, How can I program a cell?

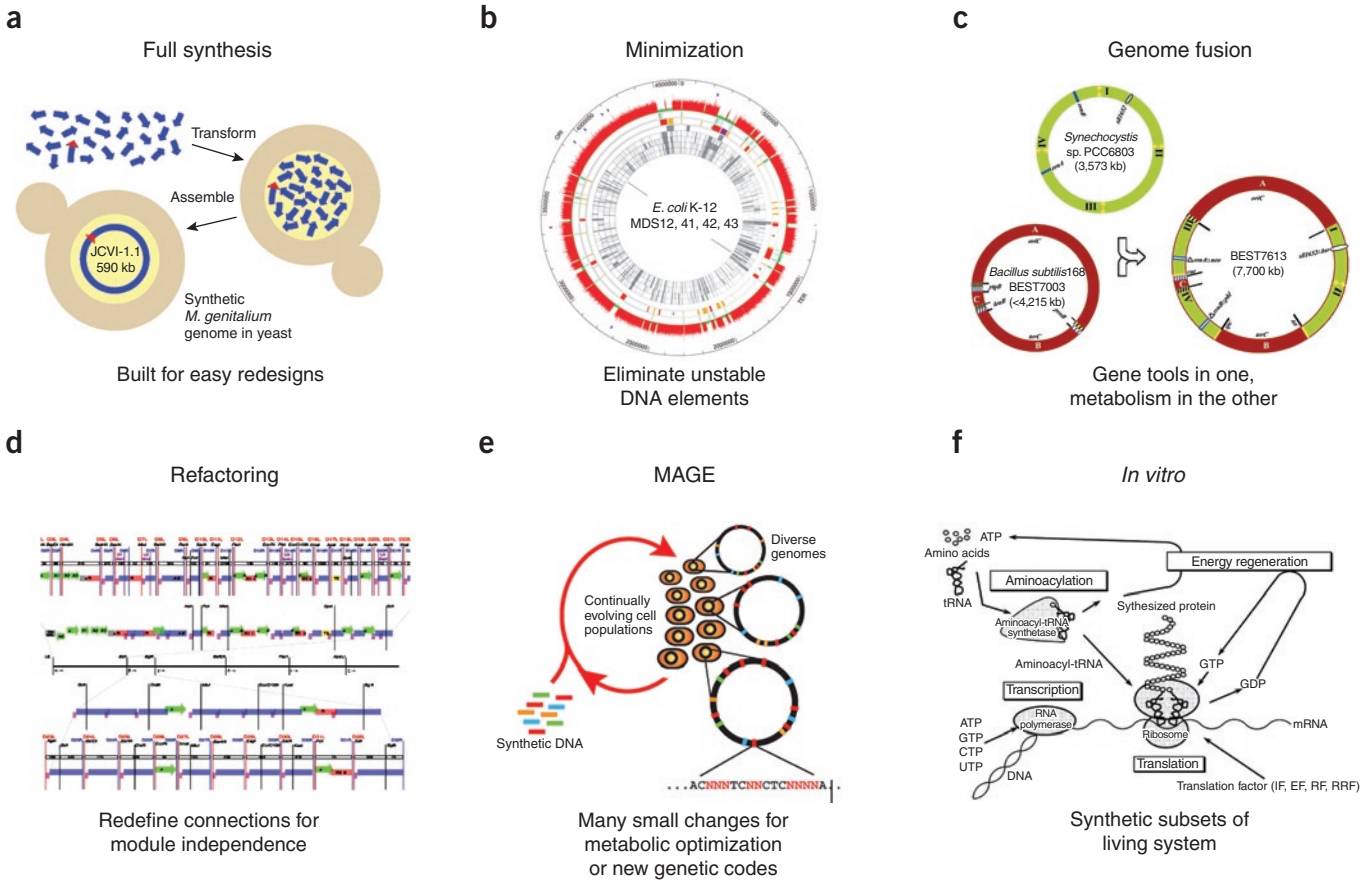


Figure 3 Examples of engineering at the genome scale. (**a**) Gibson *et al.*[61] constructed the first all-synthetic microbial genome from commercially produced DNA cassettes. (**b**) Posfai *et al.*[76] deleted many large segments of the *E. coli* genome to eliminate unstable DNA elements. (**c**) Itaya *et al.*[66] transferred the majority of one (archaeal) genome into another (prokaryote) genome. (**d**) Chan *et al.*[12] decomposed the T7 bacteriophage genome into many reconfigurable modules. (**e**) Wang *et al.*[38] demonstrated a technique for making large numbers of targeted changes to a genome. (**f**) Shimizu *et al.*[114] developed a purified translation system useful for *in vitro* prototyping of genetic functions without requiring moving genes into living cells. Note that manipulations of large DNA segments > 100 kilobase pairs in **a**, **b** and **c** relied heavily on *in vivo* recombination-based techniques.

In partnership with the Registry of Standard Biological Parts[29,30] this work also helps tackle the question of how effectively biological systems can be engineered with composed, standardized and characterized genetic components. Wrestling with these questions is essential if we are to consider designing genetic systems the size of genomes.

Regardless of purpose, most projects in gene and genome engineering share a common set of tools and overall organization principles. In considering the accomplishments, challenges and opportunities of genome engineering, we examine four basic phases of an engineering project, applied here to genomes and other complex genetic systems: design, construction, implementation and/or testing and debugging (troubleshooting).

The design of genetic systems

Although **Figure 1** emphasizes benchmarks achieved in genetic construct size, an even more significant focus should be on engineered function. **Figure 4** compares the scale of a genetic engineering project (*x*-axis, in base pairs) to the proportion of that scale that was designed *de novo* (**Fig. 4a**) and the number of 'design units' manipulated (**Fig. 4b**). No one or two metrics are expected to unify such a broad range of designs and investigations. And although a portion of them can be said to have maximized some metric as a goal (genes deleted, proportion designed, degree of reorganization or synthesis scale), many also have no such goal in mind. Degree of difficulty, the nature of the specific functions, complexity of new configurations and number of steps in an assembly hierarchy are certainly among the terms worthy of consideration.

Nevertheless, we see these projects falling into three broad categories of genetic design:

1. Design of small protein folds (up to 100% new sequence) and design of enzymatic activity (modifying scaffolds to 10–20% new sequence).

2. Design of genetic devices using naturally derived parts. These tend to display little *de novo* designed sequence; instead, new functions are derived from new configurations of existing parts. These have been well reviewed recently[15,18].

3. Manipulation of genomes by constructing, deleting and to some extent reorganizing components. These tend to be proof-of-principle reports pushing the limits of scale—often asking, How much of this can the cell tolerate?—but not of design. This statement is not a criticism, but an observation that genome engineering is in its infancy.

Figures 1, 2 and **4** together also illustrate an underlying principle: just as current DNA sequencing capacity dwarfs DNA synthesis capacity,

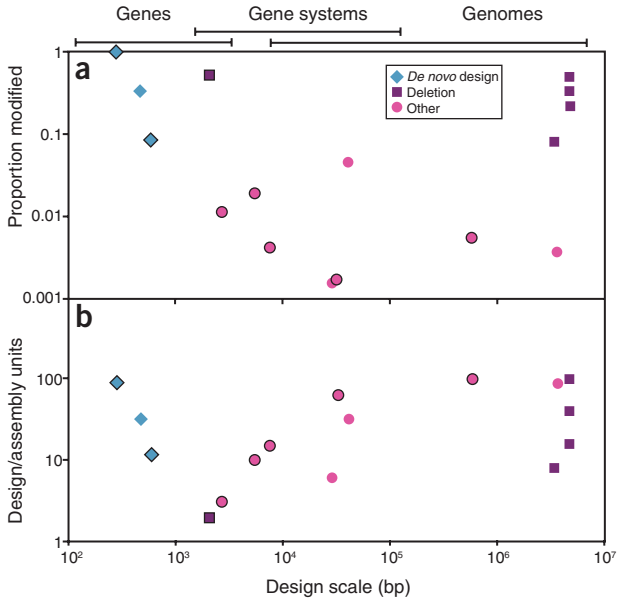


Figure 4 Current projects in genetic engineering exhibit a trade-off between the scale that can be manipulated and the scale at which one can effectively design. Projects are plotted according to (**a**) proportion modified or (**b**) number of design or assembly units. Symbols for constructs produced by *de novo* DNA synthesis are outlined in black. At the largest scales, the extent of modification has been essentially the manipulation of large natural DNA segments (synthesis, deletion, repositioning or cloning). However, the techniques developed to perform these operations set the stage for unprecedented new engineering efforts. Data for this graph are given in detail in **Supplementary Table 3**.

so DNA synthesis dwarfs current capacity for functional design and debugging. If the scale of available synthesis can be considered the size of the canvas on which we may paint, the available choices of brushes and colors are still rather modest.

The recent accomplishment by Gibson *et al.*[7] at the J. Craig Venter Institute (JCVI; Rockville, MD, USA) illustrates the cutting edge of the field. The synthesis and assembly of a 582-kb pair *M. genitalium* genome exceeded by tenfold the size of any previously published *de novo* DNA construct (but did not reduce the cost per base pair). The extent to which this genome was reengineered, however, was small, primarily a handful of DNA watermarks—intended to show that the construct truly is synthetic. And even with these slight changes, getting the product to function proved challenging. Nevertheless, the choice of minimal modification seems especially prudent as the JCVI group seeks to 'boot' the genomic software in a fully operational cytoplasm and debug the ensuing design/assembly process. A failure to run the genetic operating system at this stage does not distinguish between problems with design and problems with the general production and 'booting' methods (see below). As this assembly technology becomes more robust, putting such synthetic capacity into the hands of genetic engineers will generate enticing new opportunities for design.

Standards, parts and design frameworks. In recent years, those in the synthetic biology community have championed the need for a standardized system of genetic parts, with the hopes of enabling sophisticated genetic systems design[17,21,31–35]. A comparison is drawn with progress made last century in electronic design: standardization of parts, such as transistors and resistors, allowed mass production, generalized design and abstraction hierarchies. Such a hierarchy builds from the bottom up, so that at each level of abstraction a specialist may take advantage of foundational work from more fundamental levels. One engineer may design single parts, the next a device based on such parts and a third 'software' using such devices. Integration of an advanced design framework based on this idea requires specialists at each level, as well as generalists broadly versed in the overall design system[36].

In the initial stages of synthetic biology, design has been closely linked to physical assembly. For example with BioBricks—the first major standard implemented—assembly is kept general and independent of specific parts through the use of a restriction-ligation scheme. Although this places some sequence limits on the part boundaries and requires keeping the restriction sites themselves out of the part sequences, the flexible framework has been employed to great effect. The value of the overall concept is underlined by the development of at least five alternative assembly standards[37]. The long-term expectation in this area is that increasingly available DNA synthesis will make some of the current assembly restrictions unnecessary, and that new or modified standards will develop to take advantage of these resources.

Designs with standardized genetic parts may involve on the order of 10–20 parts—modest compared with the scale of a genome—but quite complex compared with most other genetic engineering. It is hoped that the use of such standards, coupled with vigorous characterization, will pave the way for new levels of design complexity. As this type of genetic programming approaches the scale of genomes, cloning contexts will of necessity shift from an emphasis on plasmids, to bacterial and yeast artificial chromosomes, to the primary chromosome(s) of the strain being engineered.

The interplay of design and randomness. Relative to most other fields of engineering, genome engineering has two huge potential advantages. One is the preexistence of highly evolved modules, which have some of the properties of careful design (albeit initially lacking specification sheets and without guarantees of interoperability or lack of cross-talk). The second advantage consists of the capacity to harness present-day (lab-scale) evolution and integrate the targeting of combinatorial changes genome-wide[38,39].

One general—and powerful—category of genetic engineering focuses on improving (or in some cases originating) function without a specific genetic design and instead takes a broader approach of directed evolution. A great body of successful metabolic engineering has benefited by applying this principle. Directed evolution has also been applied to the optimization of synthetic gene circuits[40]. Future breakthroughs will probably focus on the ability to design and select from large collections of semi-synthetic DNA, with major challenges including the collecting and designing of biosensors[41] and developing more complex selection criteria (e.g., involving cellular counters[42]). Biosensing can be implemented using a second cell that requires the sensed molecule for growth (syntrophy)[41,43]. Biosensors can also be obtained from allosteric regulatory proteins and RNA (riboswitches)[44]. These can be evolved *in vitro* or *in vivo* to new specificities.

Computer-aided design tools (CAD). Once natural enzymatic and regulatory modules are adapted, refined and measured, they can be combined—at the drawing console—with a high degree of abstraction (ideally with intuitive graphics) while increasingly sophisticated computational methods handle 'lower level' steps. CAD is required at levels ranging from high-level design and simulation tools for synthetic biology[45] down to the detailed layout and sequences of oligos needed for multiplex assembly of genes or genomes[46–48]. The need for CAD tools spans two extremes of design: first, combinatorial genetic

modifications that enable genome engineering with functional selection in metabolic engineering, where exploring all combinations is feasible (e.g., *cis* regulation of dozens of genes or more[38]); and second, sequence-based screening, where the number of changes to be made is too large, selections are lacking or combinations are not needed (e.g., genome-wide codon conversion in *Escherichia coli*, where, for example, all TAG stop codons are to be converted to TAA).

CAD tools are also needed to generate metabolic and signaling pathways, including processes not yet found in nature. Looking forward, a key goal will be integrating and automating the various aspects from protein design[49] to compatibility of standards and intellectual property. Purnick and Weiss[18] give a detailed listing of current computational tools for design and analysis of genetic networks—many of these demonstrate features extensible to the genome scale, which will require handling hundreds to thousands of design components. The CLOTHO software platform[50] is one example of an environment meant to be extensible to diverse design needs at different scales.

The construction of synthetic genetic material

At the simplest synthesis scale for DNA, single oligos are very affordable and available commercially on rapid time scales. For a pair of PCR primers, the time and cost of synthesis are more or less the same as the time and cost of shipping (frequently, next day shipping). Even so, for

Box 1 Joining DNA

DNA fragments can be joined in essentially one of four ways: chemical coupling, ligation, polymerization and recombination. These are summarized below.

Chemical coupling. Organic chemical synthesis of oligos proceeds by stepwise addition of single nucleotide bases to a growing chain (**Fig. 5**). The extensible end of this base (typically a 5′ hydroxyl group) is protected from further reaction by a protecting group, which is removed for the next cycle. The majority of reaction failures are also terminated by addition of a capping group to halt further chain extension. This highly optimized chemistry can provide oligos with an average stepwise yield of 99% or higher, enabling the production of oligos up to 200 units in length (and on some occasions longer[83]). Phosphoramidite chemistry dominates current synthetic methods, though alternative chemistries have also been used to great utility[84] and new developments have been recently reported[85]. This stage of DNA synthesis is also distinct as the only one achieved without a template or complementary sequence (though sequence-independent ligation of larger segments for this purpose is conceivable). Instead, the single nucleotide building blocks are built into specific strings by choices designated at each step of the serial assembly.

Figure 5 Chemical synthesis of DNA. Nucleotide bases (purple circles) are added sequentially to the 5′ end of the growing chain. Yellow arrowhead indicates the 3′ end.

Ligation. At the heart of nearly all synthetic gene-sized construction is self-assembly by means of programmed complementary base-pair interactions. After the specific association of two or more strands, the next step in producing larger pieces typically follows one of two enzymatic courses: ligation by a DNA ligase (**Fig. 6**), or oligo extension by a DNA polymerase (**Fig. 7**). The first gene syntheses employed ligation of oligos[3,86] and some newer protocols employ ligases as well[54]. Many protocols for assembling larger constructs also rely on ligation. Some of these have used short specific overhangs of 2–4 nt generated by restriction enzymes as the means of association[66], whereas Gibson *et al.*[7] generated long overhangs using the 3′ to 5′ exonuclease activity of DNA polymerases.

Figure 6 Ligation. DNA ligase makes backbone phosphate bonds (purple) connect strands of DNA (yellow).

Polymerization. Although polymerases had been well-studied long before, the introduction of the polymerase chain reaction, PCR[87], paved the way for a new set of gene synthesis protocols[88,89] (**Fig. 7**). Polymerase-based protocols employ pairs of oligos which anneal and are extended, each oligo serving as both primer and template. The typical reaction is set up to employ a pool of oligos with several of these pairings occurring simultaneously in a thermocycled reaction, essentially growing progressively longer intermediates until the full-length product is obtained. The many variations on this theme have been well reviewed elsewhere[90–92]. PCR-based overlap-extension methods can be used to generate fairly large constructs (e.g., 15 kbp by Tian *et al.*[46]), but because the upper limits of long PCR may be ~50 kbp, these approaches seem unlikely to yield larger genomes by themselves. This does not exclude the possibility of alternative methods for genome assembly employing highly processive strand-displacing polymerases in a nonthermocycled *in vitro* context.

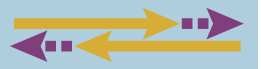


Figure 7 DNA joining by polymerization. Overlapping pairs of oligos (yellow) that anneal serve as both primer and template for extension (purple) by DNA polymerase (in direction of arrows).

Recombination. Recombination methods have been employed both *in vitro* and *in vivo* for the assembly of DNA constructs (**Fig. 8**). A well-known *in vitro* example is the Gateway system (Invitrogen; Carlsbad, CA, USA), which uses phage λ site-specific recombination enzymes for both cloning and higher order assemblies. The other common site-specific recombination system is Cre-*loxP*. Homologous recombination systems have been used for manipulating quite large pieces of DNA, including double-stranded linear replacement[93], double-stranded circle-c integration[94] in *E. coli* and *Bacillus subtilis*[66], and single-stranded-oligo invasion of replication[95]. Although generally used to manipulate one piece of DNA at a time, Gibson *et al.*[61] recently demonstrated the simultaneous recombination of 25 linear DNAs ~22 kbp each in yeast. An advantage of homologous recombination approaches is that no exogenous sequences are required for targeting, giving the possibility for scar-free assemblies.

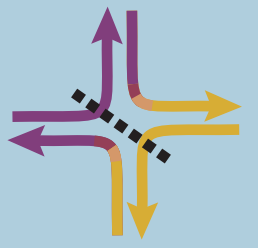


Figure 8 DNA joining by recombination. Two DNA duplexes (yellow, purple) are brought together to form a four-stranded junction. When resolved across the dotted line, new hybrid DNA duplexes result.

large synthesis projects, these costs can be considerable (e.g., 1 million base pairs of double-stranded DNA would currently cost $200,000 or more for the oligos alone before assembly, assuming synthesis at the 10 or 25 nanomole scale).

Gene-scale synthesis is also becoming highly commoditized—commercial synthetic gene providers crossed the 'buck a base' threshold ($1 per base pair) some time ago, and are currently near half that price. The price of such synthetic DNA continues to drop, with trends noted as comparable to Moore's Law, dropping a factor of 1.5 per year[8–10] (**Fig. 2**). It seems plausible that in a few (perhaps 3–5) years, commercial gene synthesis could reach the same level of convenience as for synthetic oligos: a cost and time on par with overnight shipping. When this condition is met, much of the work currently done to manipulate DNA in research labs will be outsourced. Instead of cloning into vectors stored in those labs, custom or standard vectors could simply be resynthesized on demand. To enable this flexible design structure, the synthesis community may employ intellectual property distributions comparable to VLSI (very large-scale integration) library licensing in microchip manufacture[21].

Reaching this tipping point will likely depend on emerging technologies for highly parallelized and miniaturized synthesis[46,51,52]. Bypassing or dramatically modifying current time- and cost-intensive steps, such as cloning, could also be required. High-quality error correction methods[46,53,54] may also allow some applications to proceed without conventional cloning and sequencing, or *in vitro* single-molecule cloning may be adopted[55].

The technology of synthesis and assembly. A great variety of specific protocols exist for generating DNA constructs of different sizes. Although there are too many individual techniques to discuss in detail here, they are most easily presented by factoring their particulars as combinations of a few common core elements, regardless of synthesis scale. The broad categories of 'DNA joining', 'assembly organization' and 'error control' are detailed in **Box 1** (**Figs. 5–8**), **Box 2** (**Figs. 9–12**) and **Box 3** (**Fig. 13**), respectively. Certain combinations of these elements are especially popular, but others represent untapped potential, such as chemical coupling of large constructs.

In addition, the environmental context in which these procedures are implemented has in many ways defined the limits of synthesis scale. **Figure 1** denotes the three tiers of contexts that enable synthesis of increasingly larger DNA targets. Each builds on the one below: oligo synthesis is performed via organic chemical reactions, oligos are assembled into genes via biochemical reactions and genomes are produced by manipulating gene length constructs taking advantage of one or more *in vivo* processing steps. It is worth discussing each of these steps in more detail.

The first enabling technology for the assembly of genes was that of oligo synthesis[6,56]. An excellent short history of this type of organic chemistry has been written by Hogrefe[57]. Synthetic oligos are the building blocks for larger pieces of genetic material. Although a large proportion of oligos are short (~20 nt) and used for processes such as PCR and DNA sequencing, those used for gene synthesis are longer, typically 30 nt or more.

The addition of an *in vitro* biochemical step was a crucial advance for the report of the first synthetic gene in 1970 by Khorana and coworkers[3]. The authors used the newly characterized enzyme T4 ligase to link oligos of 8 to 20 nt in length, generating the structural gene for a 77-bp yeast alanine tRNA. Such processes have been used to assemble products up to several thousand base pairs in length[58,59]. The synthetic product at this stage may variably be referred to as a gene, a synthon, a chunk, a cluster, a cassette, a segment or a part. Though these definitions overlap, they are not all equivalent. In this context, the term 'gene' often refers to a protein reading frame, possibly with additions relating to transcription and translation. A part connotes function, and may be as small as, for example, a promoter, or as large as a complex genetic device assembled from several smaller parts (as per discussion of standardized parts above). The other terms are more general and may include multiple genes or a fraction of one. Regardless of name, these constructs are generally pieces of double-stranded DNA, assembled from two or more oligos. The construction process is very often referred to as gene synthesis.

The first synthetic genome reported was that of poliovirus, published in 2002 (ref. 60). An important aspect of that synthesis was the use of processing in living systems. Though the final destination of synthetic genes before that time was also typically an *in vivo* context, the distinction made in the top tier of **Figure 1** is for processes for which *in vivo* handling was a crucial and fundamental intermediate assembly step. Thus Cello *et al.*[60] first synthesized subsets of the full-length viral genome and cloned these separately into plasmids, which were subsequently used to transform living cells. The resulting clones were sequenced and perfect clones were selected, where possible. In the absence of a perfect clone for a given segment, site-directed mutagenesis was performed to repair the DNA. Once error-free clones were obtained, propagating these and larger assemblies *in vivo* ensured a minimal rate of introduction of new errors, a strategy used in all the large assemblies discussed here.

More recent methods have relied on DNA recombination to perform assembly of very large segments of DNA *in vivo*, with yeast proving especially apt for this purpose. The JCVI team assembling the *M. genitalium* genome employed native recombination mechanisms to produce their full-length product, demonstrating that more laborious *in vitro* handling was unnecessary at some earlier steps[61]. There also exists potential to take advantage of organisms with more extensive recombination capacity, such as *Deinococcus radiodurans*, which can reassemble its own genome after extreme fragmentation[62].

As increasingly larger syntheses are attempted, the fragility of long (genome-length) DNA strands is expected to become a more challenging issue. Using cells to perform not only assembly and amplification but also DNA transfer is likely to become routine for assemblies a million base pairs and larger. Immediate opportunities are apparent in conjugative transfer of DNA between bacterial cells[63] and yeast mating and recombination[64,65].

Figure 14 displays two recent examples of large-scale DNA construction, characterized by choices of assembly technology. In addition to specific choices for joining, organization and error control (**Boxes 1–3**), a degree of parallelization and tuning are inherent in most of these processes, although emphasized more in some than in others. The different stages can be considered essentially modular—methods applied at one stage (e.g., oligo synthesis) need not be tightly coupled to the next (e.g., gene synthesis). Some approaches are better suited to specific stages, but not necessarily limited to them. For example, oligo synthesis chemistry is fairly standardized around serial condensation of phosphoramidite monomers in organic solvents. Gibson *et al.*[7] have noted that the large-scale assembly method they pioneered need not be limited to applications of *de novo* synthesis but should perform equally well for DNA extracted directly from natural sources. This is true for many of the methods detailed here. Similarly, techniques that have been applied to the large-scale manipulation of extracted natural DNA[12,64,66,67] are also worth considering for genome-scale DNA construction. The largest such construct so far is the 10 Mbp minichromosome of Kuroiwa and coworkers[68].

Automation. A central feature of efficient synthetic DNA production is automation and scale-up. DNA synthesis companies generally employ fluid-handling robots and moderately high density (96-, 384- or 1,536-well) plate formats common to the biotech industry.

These approaches are sufficient for high-throughput production on the order of megabase pairs per month. A detailed example of a production pipeline has been published by Hellinga and coworkers[69].

Another approach to high-throughput DNA production is microfluidic processing. Specific advantages to this approach include the following: first, minimization of reagent and consumable use; second, less dependence on expensive robotics; and third, direct coupling to high-density microarray-fabricated sources of oligo building blocks. Regarding this last advantage, we note the potential represented by high-density arrays that contain more genetic information—as oligos—in a few square centimeters (many megabase pairs) than any commercial gene synthesis provider currently assembles in one month. Early reports using such oligos for gene synthesis have removed oligos from array surfaces and manipulated them in macroscopic volumes (e.g., 10–20 μl), frequently requiring parallel amplification of all oligos in a pool[46,51]. Microfluidic devices present a unique opportunity to instead confine these oligos in small volumes, obtaining useful nanomolar concentrations from less than femtomolar oligo yields[52]. Spatially separating the large numbers of array oligos into many small compartments (nanoliters) for parallel syntheses is also expected to reduce the complexity of diverse oligo pools (several thousand sequences or more) to manageable levels. Applications for this purpose are currently still in their early stages[52]. Such devices have been forecast as a principal enabling technology for dramatically pushing down future consumer costs[70]. Other avenues will exploit growing libraries of prefabricated parts or genomes requiring merely hundreds of changes as enabled by the multiplex automated genome engineering (MAGE) approach from our groups[38].

Implementation—'booting' a designed genetic system

After construction or extreme modification of a genome, another special challenge remains: the DNA software must be 'booted'. Booting refers to the pulling oneself up by one's bootstraps, or more recently 'booting' a computer, wherein the software is loaded into 'blank' computer hardware[71]. How is one to get an entire genome running? One

Box 2 Assembly organization

Several different organizational schemes are available for assembling pieces of DNA into larger fragments. These can be categorized in assembly in series, by hierarchy, in parallel or by pooling. These are described below.

Serial. This simply refers to adding one unit at each stage of a synthetic process (**Fig. 9**). Organic chemical syntheses of oligos intrinsically use this approach: the choice of nucleotide added (A, C, G or T) at each stage determines how specific sequences are constructed in the absence of any DNA template. Serial assembly has been employed at every scale of DNA assembly, including the large-scale cloning of one genome (3.5 Mbp) into another species of cell[66]. One advantage of serial processes is control[96]. Even in cases where higher throughput methods may be preferred, serial construction can prove the fall-back option for difficult projects, such as the synthesis of low-complexity or repeat-intensive sequences.

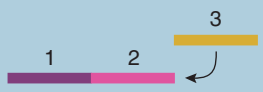


Figure 9 DNA assembly in series. Single subunits (whether single nucleotides or large DNA cassettes) are added one at a time sequentially.

Parallel. It is important to recognize a degree of parallelization inherent to some steps of a large-scale DNA construction effort (**Fig. 11**). All the oligos synthesized for producing a gene or a genome are expected to be produced in parallel. The same is true for assembling several DNA cassettes on the way to a multigene construct. Microarrays are an exceptional example of parallelizing the serial process of oligo synthesis. Oligo microarrays can be synthesized *in situ*, typically on a few cm^2 of surface, with complexities as high as over one million different specific oligo sequences.

Figure 11 DNA assembly in parallel. Most large synthesis projects require some degree of parallel processing.

Hierarchical. These schemes provide a potent balance of throughput and control. Pieces of DNA are joined in multiple stages, frequently combining in pairs at each stage (**Fig. 10**). For example, eight pieces joined as pairs produces four larger assemblies; joining these as pairs produces two even larger assemblies, and combining these yields the final construct. Employing this hierarchy requires three stages of assembly compared with seven stages for a serial approach. The advantage grows considerably at larger scales (more pieces) as N pieces can be combined with on the order of $\log_2 N$ stages (versus N–1 for a serial approach). An increase in size by a factor of ten requires only a few more stages. A version of this strategy was employed in the initial *M. genitalium* genome synthesis, with combinations of two to four DNA segments at each stage[7]. (See below, however for a pooled version.) Another advantage of hierarchical assembly strategies is that the intermediates produced can be helpful for debugging problems in complex designs. This challenge is also expected to become increasingly difficult at the longer synthesis scales.

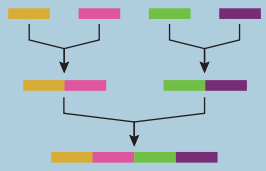


Figure 10 Hierarchical DNA assembly. Segments are joined together in subsets, producing successively larger constructs at each stage.

Pooling. Performing several joining reactions in the same mixture has proven extremely advantageous to improving the efficiency of DNA construction (**Fig. 12**). At the level of gene synthesis, both ligase-based and polymerase-based assemblies are often performed with pools of oligos. See Khorana *et al.*[86] as well as Dillon and Rosen[89] for early examples. One gene synthesis protocol, thermodynamically balanced inside-out (TBIO), combines advantages of both serial and pooled strategies[97]: oligos to make a DNA segment are combined in one pool and extended via a thermocycled polymerase reaction, but the arrangement of oligos allows only incremental growth of the product at each step. Pooled assembly reactions have been performed with groups of >200 oligos to produce a 5.3-kbp phage genome[59], and a pool of ~600 oligos has been used to assemble 21 separate genes that were later hierarchically combined to yield a 15-kbp product[46]. Pools can also be used *in vivo*[38,61].

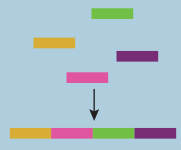


Figure 12 Pooling approaches. Multiple DNA segments are joined in a single reaction.

Box 3 Error control

Errors in synthetic DNA at any length scale need to be considered carefully. Even an error rate as low as 1 in 10,000 bp[53] can be a major concern if the product of interest is of that scale (10^4) or larger. Two major types of mistakes are worth considering: failure to assemble (global error) and mutations in an assembled product (local error). However, far more detailed information is available on the latter. Although there are many options, error control is not explicitly required at every stage. For example, some single-gene syntheses may simply sequence a small number of clones of the gene product to find one that has no errors.

Select. When it can be arranged, selection for function or viability can dramatically reduce errors in the surviving clones. Several examples have shown utility for single genes[98], gene systems (e.g., plasmids[58,99]) and a small phage genome[59]. A more general form of this concept involves fusing synthetic open-reading frames (protein-encoding or not) in-frame to a downstream selectable gene[69,100]. Most deletions in the synthetic genes should then give rise to frameshifts so that the host cell does not produce the downstream gene it needs to survive. As deletions (especially point deletions) can be the most common defect in chemically synthesized DNA[53,69,101] and are generally the most deleterious to function, this approach can improve the quality of a construct substantially when the desired product is a single protein reading frame. Beyond this are selections for proper folding (**Fig. 13**, 1) and solubility[102,103].

Tune. Most stages of DNA construction rely on some degree of optimization to minimize the opportunities for flawed pieces to occur. Examples of tuning include the extensive optimization that has accompanied commercial oligo synthesis (average stepwise yields in excess of 99%), use of stringent annealing temperatures to favor joining of oligos without mismatches[54] and selecting the most high-fidelity polymerases for amplification. Because commercial oligo manufacture has generally been optimized for other applications, tuning this organic synthesis specifically for the purpose of gene synthesis is desirable. For example, oligos from some providers lead to single-base deletions as the primary error[53,101], whereas others lead mainly to point substitutions[104]. Selection of such parameters as reagent concentrations and reaction times are likely to lead to these differences. All oligo syntheses are influenced by the degree to which undesired trace water is present during coupling reactions, as well as the age of the phosphoramidite reagents. Cerrina and coworkers[105] have demonstrated the utility of optimizing *in situ* oligo synthesis in microarrays specifically for the purpose of gene synthesis.

Repair. This category includes all manner of approaches that modify a DNA site containing an error (**Fig. 13**, 2). One such example used enzymes that cut at the site of an error (in the form of a DNA mismatch) coupled to exonuclease activity to degrade the defective sequence, and subsequent resynthesis by polymerases[106]. Many applications have used information from DNA sequencing to fix flawed clones through site-directed mutagenesis[60,107]. Note that in the former case no specific knowledge of the errors is involved, whereas in the latter it is an absolute requirement. There are a number of *in vivo* repair pathways that have not yet been adapted for synthetic DNA production. Thus, we might expect to see more applications of this type in the near future.

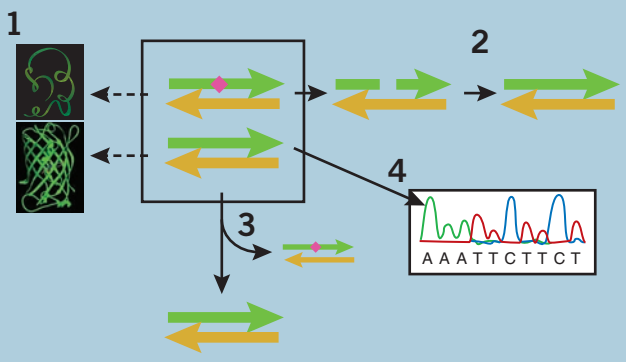


Figure 13 Error control. Two copies of a DNA assembly (green and yellow strands) are shown. One copy contains an error, such as a single-base substitution or indel (shown as a red diamond). Approaches to reduce the prevalence of such error-containing DNA include (1) selecting for clones that encode properly folded proteins that, for example, fluoresce, (2) DNA repair via nuclease excision and polymerase resynthesis or site-directed mutagenesis, (3) purification to remove defective DNA or (4) sequencing to identify error-free clones.

Purify. This category refers to methods for removing undesired species from the set of DNA molecules (**Fig. 13**, 3). Purification of oligos can be performed before gene synthesis to improve either the reliability of the process. Purification to remove defective oligos also has the potential to reduce mutations in the final assembled gene. Because one of the most common errors observed is a single-base deletion, purification of an oligo of length N must be stringent enough to remove defective oligos of length $N-1$. Hybridization-based purification of oligos has also demonstrated a dramatic improvement in error rate[46]. At the level of gene-sized pieces of DNA, the use of mismatch binding proteins has proved effective for separation of mismatched (error-containing) species[53,108]. Cleavage of mismatch duplexes has also been accomplished with endonucleases, followed by electrophoretic separation[109] or selective degradation[54]. This latter method possesses the additional advantage of all-fluid handling steps (*in vitro* biochemistry) without the need for additional separations.

Sequence. DNA sequencing is the gold standard for ascertaining the quality of a synthetic construct (**Fig. 13**, 4). For single-gene assemblies, sequencing is often the final stage of picking a winning clone. For most of the larger syntheses reported[7,46,60,104], it has been expedient to clone and sequence intermediate fragments, often of length 400–600 bp. One advantage of this size is that with error rates typical for commercial oligos, an error-free specimen can be identified after sequencing only a few clones. Also, typical read-lengths for conventional Sanger sequencing are slightly longer than this range. Thus, sequencing cloned constructs of this size can be performed using primers generic to the vector instead of specific to the construct. Going forward, the integration of second-generation (high-density microarray) synthesis[46] and sequencing[110] may require multiplex tagging and/or selective release from oligo microarrays, as well as *in vitro* molecular cloning[55,111,112] where single molecules are amplified by PCR to produce clone-like isolates, which are then sequenced.

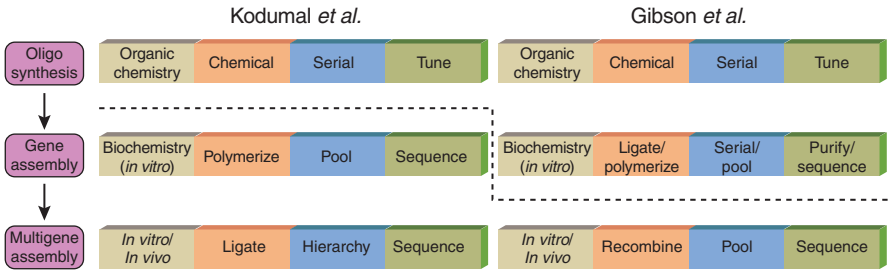


Figure 14 The general process of assembling large genetic constructs. Two recent examples[61,104] are diagrammed. Three major tiers are shown, indicating the different stages for 1, the synthesis of oligos; 2, the assembly of oligos into larger double-stranded synthons (usually in the 0.5–5 kbp size range and frequently a gene); and 3, assembly of these units into larger constructs. Colored boxes correspond to choices made regarding the assembly environment (gold), joining mode (orange, see **Box 1**), assembly organization (blue, see **Box 2**) and error control (green, see **Box 3**). A dotted line indicates when the project flow crosses over from commercially provided services—oligos for Kodumal *et al.*[104] and 5–7 kbp cassettes for Gibson *et al.*[61]. Thus, for the latter report, some gene assembly elements likely vary between the three vendors employed. '*In vitro/vivo*' refers to the toggling back and forth between cellular and test tube (aqueous) environments at stages in the processing.

may consider the possibility of creating cellular 'ghost' cells, with transcription and translation machinery, but no genome of their own. These could be generated through cell division mutants, by internal digestion of the host genomic DNA or by reassembly of membrane and cytoplasmic fractions. The synthetic genome could then be transformed into these cell-like compartments. The JCVI team has reported a related technique of transforming one type of cell with an extracted donor genome (with the host genome originally intact, but later selected against or possibly digested)[72]. Nevertheless, because the hardware environment of living systems is frequently redefined by the resident genome, the degree of designed modification (**Fig. 3a**) raises compatibility issues—including codon usage, restriction and/or modification systems[73], chromosome stability[68] and regulatory incompatibilities. In contrast, booting a synthetic bacteriophage or virus genome poses much less difficulty—though it is by no means trivial—typically using the type of cell host the wild-type virus is compatible with, ensuring fairly optimal compatibility[12,67,74].

A second approach is to incrementally alter an existing genome while the cell continues to operate. (This requires genome compatibility from one stage to the next, though not necessarily between the original and final genome states.) This strategy is a continuum of 'traditional' genetic engineering and much akin to altering an operating system while a computer is running; it has proven useful in *E. coli* genome-scale deletion studies[75,76]. In some cases, the incoming genome could stay largely silent during *in vivo* transfers and assembly. Having the core set of new transcription, translation and replication functions under dual, inducible control would allow them to be switched on and produced first with the host machinery, later assuming the dominant role in the cell. The two-genome fusion of Itaya and coworkers[66] employed the first part of this strategy, with the incoming genome segments largely dormant. Such 'running patch' methods are also proving efficient in attempts to reprogram the genetic code of *E. coli*, with the goal of enabling nonnatural amino acid applications and blocking effective horizontal gene transfer (F.J. Isaacs & P.A.C. *et al.*, unpublished data).

A third approach would be complete breakdown into *in vitro* modules, which also permits a radical degree of redesign and debugging[77]. Taking such rebooting to an extreme in terms of degree of modification would be a mirror-image genome[78], where every stereocenter in every biomolecule would be inverted relative to life as we know it. Such a system would be incompatible with any existing cytoplasm, and would require true bootstrapping from a minimal set of biochemical functions (replication, transcription and translation). Nevertheless, even this grand challenge for vitalism and the accidental nature of life's chirality leans heavily on our knowledge of working living systems. Constructing a genome entirely from *de novo* designs or selections from true random-polymers (a la polynucleotide aptamer libraries) is considerably further off.

Troubleshooting—debugging the bugs

The various approaches shown in **Figure 3** also illustrate the dual robustness and fragility of living systems. Vast amounts of a genome can be completely deleted without apparent harm to the organism and even yield improved performance[75,76]. At the same time, very modest changes expected to be functionally invisible can cause reduced fitness[60] and single-point mutations can easily be fatal. Whereas a remaining challenge in genome engineering is to improve our ability to design more robustly, designs at these scales should also place an emphasis on planned troubleshooting.

Biological complexity represents a special challenge for genome engineers. Across the different fields of engineering, many kinds of design may have conflicts between the working parts—components that do not connect as intended, or which in combination display unexpected behaviors. But for biological systems especially, the background environment is still very incompletely understood when contrasted with other disciplines, such as electronics design. Though a given genome sequence may be known, the functions of many predicted proteins typically remain unknown and the relationships between known functions incompletely mapped. The interactions between a given designed genetic system and its cellular environment may display both general components (e.g., drains on cell resources, such as ATP and ribosome translation capacity) and specific components (e.g., undesired action of a designed DNA-binding protein on host genes). Various technologies in development have the potential to reduce this complexity, such as routing protein synthesis for the engineered system via an orthogonal ribosome[79] or running a genetic circuit in an existing organelle, such as a mitochondrion or engineered cellular vesicle. Nevertheless, many types of designs will not necessarily be amenable to such isolation.

Just as the value of design and assembly hierarchies has been emphasized above, hierarchical debugging strategies will greatly facilitate successful implementation of designed genetic systems. There are two relevant hierarchies to consider. The first mirrors that used for design. All the separate genetic parts of a designed system should be tested singly in parallel, or in as simple a representation as possible. Where possible, combinations of simple parts into larger units should be performed along lines of linked function, so that these combinations can also be tested en route to the final assembly. Parts assembly strategies such as for BioBricks are intended

to facilitate this process. In contrast, an all-or-nothing assembly and test strategy leaves the designer with little hint as to what went wrong.

The second hierarchy is that of the testing environment. *In silico* (virtual, simulation), *in vitro* and *in vivo* contexts all have a role to play. Numerical simulations involving ordinary differential equations have been used to test many genetic circuit designs. *In vitro* tests, such as with transcription/translation mixes, have the potential to quickly profile simple parts and test devices in the absence of complicating factors from the whole cell. Thus, failures at this testing level are more likely to reveal fundamental flaws in the parts themselves. Neither virtual nor *in vitro* simulation can replace *in vivo* testing, as they cannot effectively represent the complexity of the biological environment. Even so, problems revealed at these earlier testing stages are likely to indicate real concerns for the *in vivo* context—and with the right resources in place will typically be much faster to test. Large-scale, rapid and cost-effective DNA synthesis will be an enabling technology for any troubleshooting that involves modifying or making new parts for a design. Thus, transitions back and forth between the drawing table and the laboratory can be kept minimal.

Perspectives

We have commented on technological advances that will enhance molecular engineers' capacity to design and build at increasingly larger genetic scales. There are a number of research goals in this area likely to be achieved in the next several years that merit additional comment.

One of these is all *in vitro* processing of large-scale syntheses, particularly when coupled to protein synthesis and functional assays—a sharp departure from the noted trend of increasingly biological processing. These syntheses may be on scales up to megabase pairs, probably exploiting parallelism rather than stitching DNA fragments together into genome lengths. One genome-scale application would be synthesizing and testing hundreds of versions of proteins designed around a specific function, a scale-up in complexity (and scale-down in time and cost per gene) of recent work such as that from the Baker group[49]. Another use would be the profiling of many genes for evaluation in constructing *in vivo* biosynthetic pathways. Starting from sequence data, cellulases from 100 or more organisms could be constructed and compared for performance in biofuel feedstock production, or enzyme components mixed and matched to optimize terpenoid production for pharmaceutical biomanufacture. A third version of these *in vitro* applications would be to start from DNA sequences obtained as clinical data (e.g., the entire genetic diversity of a specific HIV patient's viral load) and resynthesize the corresponding genes and proteins to test for compatibility with choices of drug regimens (or evaluate a new drug). This degree of personalized medicine, coupled with expected advances in DNA sequencing will be facilitated by microfluidic integration of oligo synthesis, gene synthesis[52], translation of genes to proteins and assaying[80]. The level of integration possible (sequence data in, assay data out) will also serve to decouple physical sample acquisition from the experimental molecules, which in turn will be decoupled from the data's final destination (that is, the clinician and patient). Centralized high-performance resources could thus serve a world-wide community with rapid response on the order of a day.

These forms of *in vitro* rapid prototyping have potential for evaluating not just single-gene designs, but also more complex systems—genetic circuits, metabolic pathways and even genomes. Such approaches will not represent the complexity such designs will face in their final *in vivo* settings. Rather, they provide the opportunity to characterize performance in a defined and adjustable setting (e.g., chemical environment) for single components and specified combinations of those components.

As the scale of synthesis and assembly continues to grow, new methods will also be developed to deal with the challenges of large, more fragile, genome-lengths of DNA. 'All-biological' handling of DNA transfer and recombination events seems especially likely, such as with bacterial conjugation, yeast mating and mammalian cell fusions. Furthermore, although a majority of efforts in genome engineering have focused on single-celled organisms, emerging tools for large-scale genetic manipulation in higher organisms[68,81] are also expected to prove of great use. It would not be surprising if the next tier extending the trends of **Figure 1** relies on using cellular communities to execute designs that exceed the complexity of a single genome.

By way of summary, consider the utility of a microorganism with a reengineered genome that combines the following features: (i) removal of DNA elements that contribute to genetic instability, such as insertion sequences[76] and phase variation systems; (ii) restructuring of the genetic code to ensure no cross-compatibility with other organisms[82]; (iii) simplification of the genetic code to allow easy inclusion of new, nonnatural amino acids (F.J. Isaacs & P.A.C. *et al.*, unpublished data); (iv) removal of metabolic pathways that drain the resources of the cell and are not needed in a laboratory or biomanufacturing capacity[75]; and (v) incorporation of systems that allow rapid and efficient tuning of genetic components, taking advantage of directed evolution[38]. These properties (and others) would contribute to the production of a cellular chassis that would be the starting point for a wide range of genetic programming applications. The 'genetic isolation' of a unique genetic code would give a safer context to perform advanced bioengineering—unable to make use of exogenous genes that encode toxicity factors or antibiotic resistance, and unable to effectively donate its own special genetic features to wild strains. Increased genetic stability would provide a more consistent engineering environment, with genetic variation occurring primarily only where directed. Removal of unneeded components that use up cellular resources provides the opportunity to direct more of these resources for producing useful compounds, leading to higher yields. We expect many different versions (and species) of such reengineered strains to be of great utility, with some in highly specialized application roles, and others serving for broad general use.

We now find ourselves at an intriguing turning point. The current scale of *de novo* synthesis and reuse of engineered genetic parts seems to be leading directly to new modes of design and exploration. At the gene level, many simple gene modifications, such as cloning and mutagenesis, are being replaced with automated synthesis, assembly and even characterization. This transition will also allow entry into the field of designers who need not be experts in traditional DNA manipulation techniques. Some goals will be accessible only by genome-wide methods, such as the reformatting of the genetic code by altering tens of thousands of native codon assignments, chirality or pH/thermal stabilities genome-wide. In the longer term, the scaling of genome engineering will lead us toward engineering synthetic ecosystems, multicellular developmental systems (including human) and general programmable matter.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The authors wish to acknowledge helpful discussions with J. Jacobson, A. Forster, B. Chow, members of the Church and Jacobson laboratories and the comments of anonymous reviewers. Grant support has been provided by Department of Energy-GTL, National Science Foundation (SynBERC, Center for Bits and Atoms, and Genes and Genomes Systems Cluster) and Defense Advanced Research Projects Agency programmable matter programs.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturebiotechnology/.

1. Khorana, H.G. Nucleic acid synthesis in the study of the genetic code. in *Nobel Lectures: Physiology or Medicine (1963–1970)* 341–369 (Elsevier Science, New York; 1972).

2. Nirenberg, M. The genetic code. in *Nobel Lectures: Physiology or Medicine (1963–1970)* 372–395 (Elsevier Science, New York; 1968).
3. Agarwal, K.L. *et al.* Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227**, 27–34 (1968).
4. Ryan, M.J., Brown, E.L., Sekiya, T., Kupper, H. & Khorana, H.G. Total synthesis of a tyrosine suppressor tRNA gene. XVIII. Biological activity and transcription, in vitro, of the cloned gene. *J. Biol. Chem.* **254**, 5817–5826 (1979).
5. Moore, G.E.P. Cramming more components onto integrated circuits. *Electronics* **38**, 114–117 (1965).
6. Michelson, A.M. & Todd, A.R. Nucleotides Part XXXII. Synthesis of a dithymidine dinucleotide containing a 3′:5′-internucleotidic linkage. *J. Chem. Soc.* 2632–2638 (1955).
7. Gibson, D. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
8. Carlson, R. The pace and proliferation of biological technologies. *Biosecur. Bioterror.* **1**, 203–214 (2003).
9. Shendure, J., Mitra, R.D., Varma, C. & Church, G.M. Advanced sequencing technologies: methods and goals. *Nat. Rev. Genet.* **5**, 335–344 (2004).
10. Kurzweil, R. The 21st century: a confluence of accelerating revolutions. Keynote address at 8th Annual Foresight Conference, Bethesda, Maryland, November 3, 2000. http://www.kurzweilai.net/meme/frame.html?main=/articles/art0184.html (2000).
11. Wilson, D.S. & Szostak, J.W. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611–647 (1999).
12. Chan, L.Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).
13. Hutchison, C.A. *et al.* Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* **286**, 2165–2169 (1999).
14. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
15. Lu, T.K., Khalil, A.S. & Collins, J.J. *Nat. Biotechnol.* **27**, 1139–1150 (2009).
16. Hawking, S.W. *The Universe in a Nutshell* (Bantam Books, New York, 2001).
17. Endy, D. Foundations for engineering biology. *Nature* **438**, 449–453 (2005).
18. Purnick, P.E. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (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).
20. Glass, J.I. *et al.* Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. USA* **103**, 425–430 (2006).
21. Baker, D. *et al.* Engineering life: building a fab for biology. *Sci. Am.* **294**, 44–51 (2006).
22. Rude, M.A. & Schirmer, A. New microbial fuels: a biotech perspective. *Curr. Opin. Microbiol.* **12**, 274–281 (2009).
23. Rider, T.H. *et al.* A B cell-based sensor for rapid identification of pathogens. *Science* **301**, 213–215 (2003).
24. Kobayashi, H. *et al.* Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA* **101**, 8414–8419 (2004).
25. 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).
26. 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).
27. Cheong, I. *et al.* A bacterial protein enhances the release and efficacy of liposomal cancer drugs. *Science* **314**, 1308–1311 (2006).
28. Ciglic, M. *et al.* Engineered human cells: say no to sepsis. *IET Synth. Biol.* **1**, 13–16 (2007).
29. Kelly, J.R. *et al.* Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.* **3**, 4 (2009).
30. Kosuri, S., Kelly, J.R. & Endy, D. TABASCO: A single molecule, base-pair resolved gene expression simulator. *BMC Bioinformatics* **8**, 480 (2007).
31. McDaniel, R. & Weiss, R. Advances in synthetic biology: on the path from prototypes to applications. *Curr. Opin. Biotechnol.* **16**, 476–483 (2005).
32. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **26**, 787–793 (2008).
33. Lucks, J.B., Qi, L., Whitaker, W.R. & Arkin, A.P. Toward scalable parts families for predictable design of biological circuits. *Curr. Opin. Microbiol.* **11**, 567–573 (2008).
34. Heinemann, M. & Panke, S. Synthetic biology–putting engineering into biology. *Bioinformatics* **22**, 2790–2799 (2006).
35. Ellis, T., Wang, X. & Collins, J.J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**, 465–471 (2009).
36. Knight, T. Idempotent vector design for standard assembly of Biobricks. *Dspace@MIT* http://hdl.handle.net/1721.1/21168 (2003).
37. Biobrick Formats. http://bbf.openwetware.org/Standards/Technical/Formats.html.
38. Wang, H.H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
39. Santos, C.N. & Stephanopoulos, G. Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr. Opin. Chem. Biol.* **12**, 168–176 (2008).
40. Haseltine, E.L. & Arnold, F.H. Implications of rewiring bacterial quorum sensing. *Appl. Environ. Microbiol.* **74**, 437–445 (2008).
41. Pfleger, B.F., Pitera, D.J., Newman, J.D., Martin, V.J. & Keasling, J.D. Microbial sensors for small molecules: development of a mevalonate biosensor. *Metab. Eng.* **9**, 30–38 (2007).
42. Friedland, A.E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
43. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
44. Link, K.H. & Breaker, R.R. Engineering ligand-responsive gene-control elements: lessons learned from natural riboswitches. *Gene Ther.* **16**, 1189–1201 (2009).
45. Computational Tools. http://openwetware.org/wiki/Computational_Tools.
46. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
47. Python Synthetic Biology libraries. http://bitbucket.org/chapmanb/synbio/.
48. Czar, M.J., Cai, Y. & Peccoud, J. Writing DNA with GenoCAD. *Nucleic Acids Res.* **37**, W40–47 (2009).
49. Jiang, L. *et al.* De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).
50. CLOTHO. http://2008.igem.org/Team:UC_Berkeley_Tools (2008).
51. Richmond, K.E. *et al.* Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.* **32**, 5011–5018 (2004).
52. Kong, D.S., Carr, P.A., Chen, L., Zhang, S. & Jacobson, J.M. Parallel gene synthesis in a microfluidic device. *Nucleic Acids Res.* **35**, e61 (2007).
53. Carr, P.A. *et al.* Protein-mediated error correction for de novo DNA synthesis. *Nucleic Acids Res.* **32**, e162 (2004).
54. Bang, D. & Church, G.M. Gene synthesis by circular assembly amplification. *Nat. Methods* **5**, 37–39 (2008).
55. Ben Yehezkel, T. *et al.* De novo DNA synthesis using single molecule PCR. *Nucleic Acids Res.* **36**, e107 (2008).
56. Tener, G.M., Gilham, P.T., Razzell, W.E., Turner, A.F. & Khorana, H.G. Studies on the chemical synthesis and enzymatic degradation of desoxyribo-oligonucleotides. *Ann. NY Acad. Sci.* **81**, 757–775 (1959).
57. Hogrefe, R. A short history of oligonucleotide synthesis. http://www.trilinkbiotech.com/tech/oligo_history.pdf
58. Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. & Heyneker, H.L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
59. Smith, H.O., Hutchison, C.A. III, Pfannkoch, C. & Venter, J.C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445 (2003).
60. Cello, J., Paul, A.V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016–1018 (2002).
61. Gibson, D.G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. USA* **105**, 20404–20409 (2008).
62. Zahradka, K. *et al.* Reassembly of shattered chromosomes in *Deinococcus radiodurans*. *Nature* **443**, 569–573 (2006).
63. Gemski, P. Jr., Wohlhieter, J.A. & Baron, L.S. Chromosome transfer between *Escherichia coli* HFR strains and Proteus mirabilis. *Proc. Natl. Acad. Sci. USA* **58**, 1461–1467 (1967).
64. Larin, Z., Taylor, S.S. & Tyler-Smith, C. A method for linking yeast artificial chromosomes. *Nucleic Acids Res.* **24**, 4192–4196 (1996).
65. Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.* **3**, 371–377 (2008).
66. Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc. Natl. Acad. Sci. USA* **102**, 15971–15976 (2005).
67. Donaldson, E.F. *et al.* Systematic assembly of a full-length infectious clone of human coronavirus NL63. *J. Virol.* **82**, 11948–11957 (2008).
68. Kuroiwa, Y. *et al.* Manipulation of human minichromosomes to carry greater than megabase-sized chromosome inserts. *Nat. Biotechnol.* **18**, 1086–1090 (2000).
69. Cox, J.C., Lape, J., Sayed, M.A. & Hellinga, H.W. Protein fabrication automation. *Protein Sci.* **16**, 379–390 (2007).
70. Newcomb, J., Carlson, R. & Aldrich, S.C. *Genome Synthesis and Design Futures: Implications for the US Economy* (Bio Economic Research Associates, Cambridge, MA, USA, 2007).
71. http://en.wikipedia.org/wiki/Bootstrapping.
72. Lartigue, C. *et al.* Genome transplantation in bacteria: changing one species to another. *Science* **317**, 632–638 (2007).
73. Lartigue, C. *et al.* Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* (2009).
74. Tumpey, T.M. *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310**, 77–80 (2005).
75. Mizoguchi, H., Sawano, Y., Kato, J. & Mori, H. Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Res.* **15**, 277–284 (2008).
76. Posfai, G. *et al.* Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044–1046 (2006).
77. Forster, A.C. & Church, G.M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2**, 45 (2006).
78. Forster, A.C. & Church, G.M. Synthetic biology projects in vitro. *Genome Res.* **17**, 1–6 (2007).
79. Rackham, O. & Chin, J.W. A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol.* **1**, 159–166 (2005).

80. Maerkl, S.J. & Quake, S.R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).
81. Ivics, Z. *et al.* Transposon-mediated genome manipulation in vertebrates. *Nat. Methods* **6**, 415–422 (2009).
82. Church, G.M. Safeguarding biology. *Seed* **20**, 84–86 (2009).
83. Ciccarelli, R.B., Gunyuzlu, P., Huang, J., Scott, C. & Oakes, F.T. Construction of synthetic genes using PCR after automated DNA synthesis of their entire top and bottom strands. *Nucleic Acids Res.* **19**, 6007–6013 (1991).
84. Reese, C.B. The chemical synthesis of oligo- and poly-nucleotides: a personal commentary. *Tetrahedron* **58**, 8893–8920 (2002).
85. Sierzchala, A.B. *et al.* Solid-phase oligodeoxynucleotide synthesis: a two-step cycle using peroxy anion deprotection. *J. Am. Chem. Soc.* **125**, 13427–13441 (2003).
86. Khorana, H.G. *et al.* Total synthesis of the structural gene for the precursor of a tyrosine suppressor transfer RNA from *Escherichia coli*. 1. General introduction. *J. Biol. Chem.* **251**, 565–570 (1976).
87. Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 263–273 (1986).
88. Prodromou, C. & Pearl, L.H. Recursive PCR: a novel technique for total gene synthesis. *Protein Eng.* **5**, 827–829 (1992).
89. Dillon, P.J. & Rosen, C.A. A rapid method for the construction of synthetic genes using the polymerase chain reaction. *Biotechniques* **9**, 298–300 (1990).
90. Stewart, L. & Burgin, A.B. Whole gene synthesis: a Gene-O-Matic future. *Front. Drug Des. Discov.* **1**, 297–341 (2005).
91. Czar, M.J., Anderson, J.C., Bader, J.S. & Peccoud, J. Gene synthesis demystified. *Trends Biotechnol.* **27**, 63–72 (2009).
92. Xiong, A.S. *et al.* Chemical gene synthesis: strategies, softwares, error corrections, and applications. *FEMS Microbiol. Rev.* **32**, 522–540 (2008).
93. Kolisnychenko, V. *et al.* Engineering a reduced *Escherichia coli* genome. *Genome Res.* **12**, 640–647 (2002).
94. Link, A.J., Phillips, D. & Church, G.M. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J. Bacteriol.* **179**, 6228–6237 (1997).
95. Li, X.T. *et al.* Identification of factors influencing strand bias in oligonucleotide-mediated recombination in *Escherichia coli*. *Nucleic Acids Res.* **31**, 6674–6687 (2003).
96. Van den Brulle, J. *et al.* A novel solid phase technology for high-throughput gene synthesis. *Biotechniques* **45**, 340–343 (2008).
97. Gao, X., Yo, P., Keith, A., Ragan, T.J. & Harris, T.K. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **31**, e143 (2003).
98. Withers, S.T., Gottlieb, S.S., Lieu, B., Newman, J.D. & Keasling, J.D. Identification of isopentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. *Appl. Environ. Microbiol.* **73**, 6277–6283 (2007).
99. Mandecki, W., Hayden, M.A., Shallcross, M.A. & Stotland, E. A totally synthetic plasmid for general cloning, gene expression and mutagenesis in *Escherichia coli*. *Gene* **94**, 103–107 (1990).
100. Davis, C.A. & Benzer, S. Generation of cDNA expression libraries enriched for in-frame sequences. *Proc. Natl. Acad. Sci. USA* **94**, 2128–2132 (1997).
101. Hoover, D.M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002).
102. Drew, D. *et al.* A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci.* **14**, 2011–2017 (2005).
103. Cabantous, S. & Waldo, G.S. *In vivo* and *in vitro* protein solubility assays using split GFP. *Nat. Methods* **3**, 845–854 (2006).
104. Kodumal, S.J. *et al.* Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. USA* **101**, 15573–15578 (2004).
105. Kim, C. *et al.* Progress in gene assembly from a MAS-driven DNA microarray. *Microelectron. Eng.* **83**, 1613–1616 (2006).
106. Fuhrmann, M., Oertel, W., Berthold, P. & Hegemann, P. Removal of mismatched bases from synthetic genes by enzymatic mismatch cleavage. *Nucleic Acids Res.* **33**, e58 (2005).
107. Marsic, D., Hughes, R.C., Byrne-Steele, M.L. & Ng, J.D. PCR-based gene synthesis to produce recombinant proteins for crystallization. *BMC Biotechnol.* **8**, 44 (2008).
108. Binkowski, B.F., Richmond, K.E., Kaysen, J., Sussman, M.R. & Belshaw, P.J. Correcting errors in synthetic DNA through consensus shuffling. *Nucleic Acids Res.* **33**, e55 (2005).
109. Young, L. & Dong, Q. Two-step total gene synthesis method. *Nucleic Acids Res.* **32**, e59 (2004).
110. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145 (2008).
111. Zhang, K. *et al.* Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* **24**, 680–686 (2006).
112. Hutchison, C.A. III, Smith, H.O., Pfannkoch, C. & Venter, J.C. Cell-free cloning using phi29 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **102**, 17332–17336 (2005).
113. Zhou, X. *et al.* Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res.* **32**, 5409–5417 (2004).
114. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* **36**, 299–304 (2005).

nature biotechnology

REVIEW

Synthetic viruses: a new opportunity to understand and prevent viral disease

Eckard Wimmer[1], Steffen Mueller[1], Terrence M Tumpey[2] & Jeffery K Taubenberger[3]

Rapid progress in DNA synthesis and sequencing is spearheading the deliberate, large-scale genetic alteration of organisms. These new advances in DNA manipulation have been extended to the level of whole-genome synthesis, as evident from the synthesis of poliovirus, from the resurrection of the extinct 1918 strain of influenza virus and of human endogenous retroviruses and from the restructuring of the phage T7 genome. The largest DNA synthesized so far is the 582,970 base pair genome of *Mycoplasma genitalium*, although, as yet, this synthetic DNA has not been 'booted' to life. As genome synthesis is independent of a natural template, it allows modification of the structure and function of a virus's genetic information to an extent that was hitherto impossible. The common goal of this new strategy is to further our understanding of an organism's properties, particularly its pathogenic armory if it causes disease in humans, and to make use of this new information to protect from, or treat, human viral disease. Although only a few applications of virus synthesis have been described as yet, key recent findings have been the resurrection of the 1918 influenza virus and the generation of codon- and codon pair–deoptimized polioviruses.

Unprecedented progress in synthesis and sequence analysis of DNA lies at the heart of the recent transformation of molecular biology and the emergence of the field termed synthetic biology. Sequencing a DNA in the megabase (Mb) range is no longer a daunting undertaking and, applying the most advanced technology, can be accomplished within less than a week. DNA synthesis has not yet advanced to the efficiency of DNA sequencing, but synthesizing DNA of 8–30 kilobase pairs (kbp)—the genome size of most RNA viruses and many DNA viruses—can be accomplished easily and is largely a matter of available resources.

It is not surprising, therefore, that the *de novo* synthesis of viral genomes in the absence of a natural template has found its way into studies of viruses, although this branch of virology is still in its infancy. Chemical synthesis of viral genomes provides a new and powerful tool for studying the function and expression of viral genes, as well as their pathogenic potential. This method is particularly useful if the natural viral template is not available. It also allows the genetic modification of viral genomes on a scale that would be impossible to achieve by conventional molecular biology methods.

In this Review, we summarize briefly the recent advances in DNA synthesis and sequencing and their impact on virology. Specifically, we describe the *de novo* synthesis of viruses in the absence of a natural template with the aim of resurrecting viruses using archaevirology, identifying viruses that cause human diseases after zoonotic infections, reconstructing viral genomes to unravel complex biological systems ('refactoring' genomes) or recoding viral genetic information for the production of vaccine candidates.

Synthetic biology is the design and construction of new biological entities, such as enzymes, genetic circuits and cells, or the redesign of existing biological systems[1]. Such changes exceed those introduced previously into biological systems by methods of classic molecular engineering. Synthetic biology depends on the collaboration of specialists from different disciplines, as it requires knowledge in molecular biology, computer science, engineering, mathematics, physics and chemistry. Of the papers discussed in this short Review, it is only the research efforts to refactor the genome of a bacteriophage[2] or to recode RNA viruses[3–5] that belong to the category of synthetic biology. It can be predicted with certainty, however, that this will rapidly change in the coming years. In the future, large-scale changes will be introduced into numerous viruses, allowing the creation of redesigned particles that can provide new insights into biology or the design of new vectors that can prevent or cure infectious diseases, cure genetic deficiencies by delivering genes or treat cancer through oncolytic mechanisms, to name but a few applications.

Nucleic acid synthesis and sequencing

In 1828, Friedrich Wöhler synthesized urea from inorganic sources[6], striking a heavy blow to the doctrine of vitalism[7]. The chemical DNA was discovered in 1869 (ref. 8), but it took decades to solve the structural configuration of polynucleotides[9,10]. In keeping with their tradition, chemists began to synthesize DNA as soon as DNA structures had been published. The most ambitious of such early ventures was Khorana's synthesis of a 75-base-pair (bp) double-stranded DNA that encoded the nucleotide sequence of yeast $tRNA^{Ala}$, published in 1970 (ref. 11). This was followed by the chemical and enzymatic synthesis of the first man-made functional gene, the 207-bp DNA of *Escherichia coli* tyrosine suppressor tRNA[12].

These early landmarks consumed enormous resources, in the case of $tRNA^{Ala}$ some "20 man-years of effort"[13]. In the 1980s, however, DNA synthesis went through a rapid transformation, with the introduction of

[1]Department of Molecular Genetics and Microbiology, Stony Brook University, Stony Brook, New York, USA. [2]Influenza Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA. [3]Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA. Correspondence should be addressed to E.W. (ewimmer@ms.cc.sunysb.edu).

Published online 9 December 2009; doi:10.1038/nbt.1593

novel activated nucleosides that allowed fully automated 3′-to-5′ synthesis of oligodeoxynucleotides (oligos) on solid supports[13,14]. In particular, phosphoramidites (that is, nucleotides that carry protective groups on the reactive hydroxyl and phosphate groups of the ribose and the amine of the base) have been the building blocks of choice. During the past 20 years, numerous DNA synthesis companies have been established in response to an exploding demand for oligos (~15–80 nucleotides (nt)) that are used for genetic analyses, PCR, diagnostic assays, sequence determination or other procedures. The turnaround time for an order of a 75-bp DNA, corresponding to yeast $tRNA^{Ala}$, with extra base pairs at each end encoding restriction sites for subcloning, is currently less than 1 week—a fraction of the time and effort expended originally in Khorana's laboratory.

The assembly of larger DNA segments representing genes or entire genomes, however, is still tedious and costly, even today. It requires many oligos that must be purified, because their chemical synthesis is error prone (none of the successive chemical reactions during 3′-to-5′ chain elongation proceeds at 100%). For this reason, the building blocks for the assembly of large polynucleotides are generally no longer than 40–80 nt. Different approaches have been used to assemble oligos into large polynucleotides, although all have in common the processes of enzymatic chain elongation and/or ligation of hybridized overlapping oligos[14,15]. Examples are the 2.7-kbp plasmid containing the β-lactamase gene[16] and the 4,917-kbp gene encoding the merozoite surface protein (MSP-1) of *Plasmodium falciparum*[17]. Currently, synthesizing genes or genomes is most cost efficient when done in part by commercial facilities, where the cost per base pair of finished and sequence-confirmed DNA is now as low $0.39 (E.W. and S.M., based on information obtained from an informal web survey).

Work in one of our groups (E.W. and colleagues)[18] led to the first chemical synthesis of a DNA (7,500 bp) corresponding to the entire genome of an infectious organism, poliovirus, published in 2002. At the time of its publication, the poliovirus-specific DNA was the largest DNA ever synthesized. This milestone was subsequently dwarfed in scale by the synthesis of the 582,970-bp genome of *Mycoplasma genitalium* in 2008 (ref. 19). Although this synthetic bacterial genome has not yet been 'booted' to life, the assembly of such a large DNA molecule bears witness to the vast possibilities that DNA synthesis will ultimately offer in engineering bacteria or viruses.

Although the mechanics of constructing genes or genomes from oligos is being refined, DNA synthesis is making rapid progress, so that it is likely to fundamentally change research in molecular biology[14]. In 2004, Tian *et al.*[20] published a massively parallel microchip-based DNA synthesis approach that they predicted "might increase yields in oligo synthesis from 9 bp per dollar to 20 kbp per dollar." Once this or related strategies have matured and reach commercialization, the synthesis of small viral DNA genomes (for example, the 3,215-bp genome of hepatitis B virus (HBV)) could be accomplished for less than $100. At so low a cost, who would then construct an HBV mutant by such classic methods as site-directed mutagenesis? All current gene synthesis methods, either practiced or just conceived, still depend on relatively short oligos as their basic building blocks. But further progress in synthetic biology will require accurate synthesis of long, continuous DNA sequences.

Synthesizing large DNA molecules would be of only limited value if new methods of DNA sequencing had not kept pace with the advances in synthesis. In fact, the advances in DNA sequencing have dwarfed current DNA synthesis technology.

The first sequence of a naturally occurring polynucleotide, yeast $tRNA^{Ala}$, was deciphered by R. Holley and colleagues[21] in 1965. Initiated in 1958, the most difficult task of this project was to isolate from 140 kg of bakers' yeast 1 g of highly purified $tRNA^{Ala}$, whose 76 ribonucleotides were then sequenced in 2.5 years (the sequence was later revised slightly)[22]. Since then, however, two phases of technological innovation in sequencing have led to rapid progress[23].

The first phase was based on generating radioactive, sequence-specific fragments of DNA and separating them by PAGE[24,25]. Sanger's method of producing fragments enzymatically by chain termination with dideoxynucleoside triphosphates proved to be more practical than the chemical method of Maxam and Gilbert. Subsequently, gels were replaced by capillaries, and radioactive labels by four-color fluorescence; the process was automated and streamlined, but the underlying principle of the dideoxy method remains, to this day, the most widely used platform of DNA sequencing.

The second phase, still in its infancy, falls under the rubric of a single paradigm, termed 'cyclic array sequencing'. "Cyclic array platforms achieve low cost by simultaneously decoding a two-dimensional array bearing millions (potentially billions) of distinct sequence features"[23]. Such instruments, with slightly different technologies, are already commercially available from companies. Other methods, such as single-molecule sequencing, sequencing by microelectrophoresis, sequencing by mass spectrometry or sequencing by squeezing DNA through tiny nanopores (reviewed in ref. 23) are being tested, but have yet to mature into commercially useful techniques.

The strong progress in sequencing technologies is evident in the reduction in time and costs of human genome projects. The sequence of the 'inaugural human genomes' (3×10^9 bp), published in 2001 (refs. 26–28), was determined over a period of roughly 10 years at a cost of $3 billion—and it was incomplete[27]. In contrast, the complete sequence of Jim Watson's genome was determined in 4 months at a cost of less than $1 million[29]. Currently, the price has dropped further to below $50,000 (ref. 30), and there is reason to believe that the number of solved human sequences will exceed 1,000 in the near future.

Virology in the era of gene synthesis

Viruses store their genetic information in DNA or RNA. Total-genome synthesis of a viral genome seemed likely to occur first with one of the small DNA viruses; the protocol seemed straightforward: simply transfect the synthetic DNA into suitable host cells and assay the emerging virus. In fact, the first chemical whole-genome synthesis was performed with poliovirus, an RNA virus.

How do RNA viruses fit into the world of DNA synthesis and DNA sequencing? The answer is 'reverse genetics'. In their landmark paper of 1978, Weissmann and colleagues[31] converted the RNA genome (4,127 nt) of the RNA phage Qβ into double-stranded DNA with the aid of reverse transcriptase, an enzyme (of retroviruses) that transcribes RNA into DNA. The virus-specific double-stranded DNA (cDNA), which was embedded into a plasmid, yielded authentic Qβ phage following transfection into bacteria. At the time, the authors concluded that the viral cDNA "would allow genetic manipulations that cannot be carried out at the RNA level"[31,32], an understatement that revolutionized molecular biology of RNA viruses. Three years later, Racaniello and Baltimore[33] repeated this experiment with poliovirus. Again, the virion RNA, embedded as cDNA into a plasmid, yielded authentic poliovirus in very poor yield when transfected into HeLa cells, a human cancer cell line that is optimal for poliovirus proliferation.

Depending on the nature of the RNA virus (either positive-strand viruses, whose genome is of the same polarity as mRNA, or negative-strand viruses, whose genome is of the opposite polarity to that of mRNA), virus-specific cDNAs can now be readily prepared and used by different strategies to regenerate the parental RNA virus in high yield. The utility of reverse genetics was quickly recognized and, not surprisingly, it has now been developed for member viruses of nearly every known RNA virus family (for example, rabies virus[34], respiratory

syncytial virus[35], influenza A virus[36,37], measles virus[38], Ebola virus[39] and bunyavirus[40]. Reverse genetics systems have also been recently achieved for members of the *Reoviridae* (viruses with a segmented double-stranded genome—for example, rotavirus[41]—using a helper virus–driven reverse genetics procedure).

Synthesis of viral cDNA with reverse transcriptase requires, of course, naturally occurring virion RNA as template. An alternative is the chemical synthesis of cDNA, which, of course, requires knowledge of the viral genome sequence. The first replicating structure that was synthesized from sequence information was a replicon of the hepatitis C virus that lacked the genes for the structural proteins[42], and the first synthesis of a complete viral genome was that of poliovirus[18]. Currently, 2,361 complete viral genome sequences have been deposited in the Viral Genome Resource (http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid=10239), ready to be downloaded and investigated further. Thus, there are huge resources of information available in the virus field, waiting to enter studies that we may broadly term synthetic virology.

There are, of course, cases in which no complete viral genome sequences are available for chemical synthesis. A notable recent example is the synthesis of the 1918 'Spanish' influenza pandemic virus, which caused the most severe influenza pandemic in history. Although the pandemic virus was not isolated at the time, work in one of our laboratories (J.K.T. and colleagues)[43–49] deciphered the genome sequence using influenza viral RNA fragments, <100 nucleotides in length, that were preserved in the tissues of victims of the 1918 pandemic. In addition, Hahn and colleagues successfully synthesized chimpanzee retrovirus simian immunodeficiency virus cpz (SIVcpz)[50], the natural reservoir of HIV-1 and another case in which chemical synthesis was the only means of obtaining cDNA to generate infectious virus. The resurrection of an infectious retrovirus by whole-genome synthesis[51] of a consensus sequence of ancient remnants endogenous to the human genome also illustrates the potential of this approach in archaevirology. As with the synthesis of SIVcpz, the total synthesis of an infectious recombinant bat severe acute respiratory syndrome (SARS)-like coronavirus cDNA (29.7 kbp) was also aimed at studying mechanisms of *trans*-species infection and zoonosis[52].

In the following sections, we discuss in more detail the complete synthesis of several different viruses, either in the absence of natural template or using reverse genetics. We then go on to describe several applications of synthetic virology, such as the large-scale recoding of the viral genomes for the production of attenuated vaccine candidates or the use of refactoring (that is, the synthesis of portions of a genome) to facilitate the elucidation of individual gene functions.

Whole-genome synthesis of poliovirus

In 2002, one of our groups (E.W. and colleagues)[18] published the cell-free chemical biochemical synthesis of poliovirus type 1, Mahoney (PV1(M) in the absence of a natural template. This work caught global attention, high praise, ridicule and fierce condemnation[53]. Apart from providing a 'proof of principle', the experiment signaled a new era in biology—that is, the chemical synthesis of organisms as an approach to investigating gene function and pathogenicity by allowing large-scale changes in a genome of interest. The extent of such alterations (for example, large changes in genome architecture and gene structure) cannot be achieved with the 'traditional' methods in molecular biology[54].

The synthetic polio cDNA contained 27 intentional nucleotide changes that were placed across the genome to serve as genetic markers (watermarks). When grown in HeLa cells, an efficient tissue culture system for poliovirus proliferation, the synthetic virus (denoted sPV1(M)) showed no phenotypic changes compared with the wild-type PV1(M). However, when injected intracerebrally into *CD155* tg mice, which are transgenic for the poliovirus receptor, CD155, the median lethal dose (LD_{50}) was five orders of magnitude higher than that of the wild-type virus (LD_{50} values of 10^2 and 10^7 for the wild-type virus and sPV, respectively; E.W. and colleagues)[18]. Unexpectedly, the genetic locus for the enormous attenuation of sPV(M) was a single A residue at position 102 of the genome, located in the 5′ nontranslated region (5′ NTR) of the genome at a site long thought to serve as simply a spacer between two highly structured regions[54]. This unexpected result led us[55] to develop a highly attenuated oncolytic poliovirus.

The synthesis of poliovirus did not require living cells. Subsequent to its chemical synthesis, the cDNA was transcribed *in vitro* into infectious viral RNA (E.W. and colleagues)[56] that, in turn, yielded infectious sPV1(M) upon incubation in an extract of non-infected HeLa cells (E.W. and colleagues)[57]. For the chemist, therefore, poliovirus is nothing more than a chemical. When the virus enters a cell, however, it has a program for survival. It will subvert cellular compartments and turn them into viral factories, in which it will proliferate subject to the evolutionary laws—heredity, genetic variation, selection toward fitness, evolution into different species and so on. That is, poliovirus obeys the same rules that apply to living entities[53]. One could even argue that poliovirus has sex in the infected cell, as it readily recombines with sibling progeny or with related viruses should they co-infect the same cell (E.W. and colleagues)[58]. This fascinating dual nature of viruses as nonliving and living entities[53,59–63]—that is, an existence as chemicals with a life cycle—has been largely ignored in response to the chemical and biochemical synthesis of poliovirus, which was published in 2002.

Finally, it should be noted that the synthesis of poliovirus also confirmed the accuracy of the genome sequence. This may be considered utterly superfluous, as the sequencing of PV1(M), the first of any lytic animal RNA virus, was originally carried out by two different methods[64,65] and confirmed subsequently in numerous genetic analyses. But chemical synthesis is clearly useful in providing confirmation of sequence and will prove useful going forward in the proofreading of larger genomic sequences[14,18,66].

Whole-genome synthesis of 1918 'Spanish' influenza virus

Unlike the strain of poliovirus type 1 whose synthesis was described above, the virus causing the 'Spanish' influenza pandemic in 1918–1919 was not isolated at the time of the outbreak, and thus its reconstruction using gene synthesis and reverse genetics technology first required characterization of the viral genome using archaevirology. The influenza pandemic of 1918–1919 caused up to 50 million deaths worldwide and remains an ominous warning to public health as to the possible impact that a future influenza pandemic could have (J.K.T. and colleagues)[67,68]. Many questions about its origins, its unusual epidemiological features and the basis of its pathogenicity remain unanswered, but interest in the 1918 virus has been prompted by the possible emergence of a future pandemic caused by the H5N1 virus. Understanding how the 1918 pandemic virus emerged and mapping the virulence factors may also help us in preparations for the current H1N1 influenza pandemic.

The effort to determine the complete genomic sequence of the 1918 influenza virus began in 1995, when one of us (J.K.T. and colleagues) initiated a project to recover viral RNA fragments of the 1918 virus from preserved tissues of victims of the pandemic using reverse transcription PCR (RT-PCR)[43]. The genome was completed in 2005 (refs. 44–49; reviewed in ref. 69). The development of reverse genetics technology for influenza viruses in 1999, which allowed the production of infectious virus entirely from plasmid-cloned influenza gene segments without helper virus[36,37,70], makes it possible to produce influenza viruses with specific sequences for research into pathogenesis and molecular virology, as well as for vaccine production. This technology also made possible experiments using infectious viruses that contain 1918 influenza genes. Once

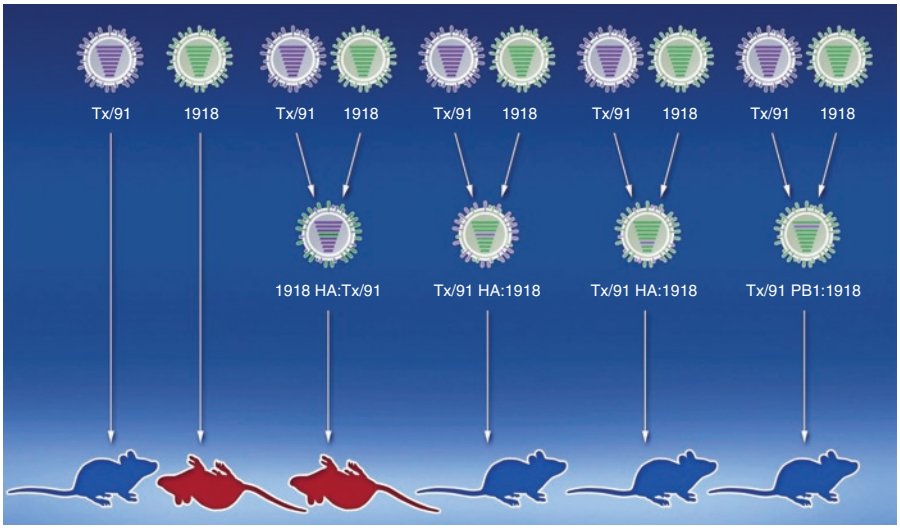


Figure 1 Comparison of lethality in mice infected with select 1918 and modern human H1N1 influenza A (Tx/91) reassortant viruses. BALB/c mice were inoculated intranasally with 10^5 PFU of virus to determine which virus genes of the 1918 virus contributed to virulence. Among all eight gene segments tested individually, the HA gene was the only 1918 virus gene able to confer a virulent phenotype when rescued on the genetic background of Tx/91 H1N1 virus. In the reciprocal experiments, the exchange of most of the individual 1918 influenza virus genes with seasonal influenza Tx/91 virus genes did not alter the virulence of the lethal 1918 virus; however, substitution of the HA, NA or PB1 genes substantially affected the ability of this virus to cause severe disease in mice. Illustration by J. Archer (Centers for Disease Control).

the sequence was determined, the 1918 influenza virus gene segments were synthesized using commercially obtained overlapping oligos and subcloned into plasmids for 'rescue' using reverse genetics[36,37,71]. This was crucial, because sequence analysis alone offered no direct clues to the pathogenicity of the 1918 virus.

Work in our laboratories (J.K.T., T.J.T. and collaborators)[72–78] has shown that, in mice, viral constructs bearing at least the 1918 hemagglutinin (HA) gene in a background of modern, non–mouse-adapted human influenza A virus are all highly pathogenic. Furthermore, expression microarray analysis performed on whole-lung tissue of mice infected with the reconstructed 1918 virus or viral constructs containing at least the 1918 HA and neuraminidase (NA) genes showed marked upregulation of mouse genes involved in apoptosis, tissue injury and oxidative damage[73,75]. Pathology in mice, although reminiscent of some of the acute viral pneumonia pathology seen in 1918 autopsy studies (J.K.T. and Morens, D.M.)[79], is nevertheless distinctive. These findings were unexpected, because the viruses with the 1918 HA gene had not been adapted to mice. Control experiments in which mice were infected with modern human viruses produced limited viral replication and little disease.

A recent study in which single gene segments of the 1918 virus were replaced with those from a recent human H1N1 influenza virus, work in one of our laboratories (T.M.T. and colleagues)[76] has revealed that the HA, NA and polymerase PB1 genes are important for virulence and replication in the mouse system; however, the fully virulent phenotype is observed only with the completely reconstructed virus (**Fig. 1**; T.M.T., J.K.T. and colleagues)[80,76]. The demonstrated role of the HA and PB1 genes in replication efficiency and virulence is particularly interesting because both genes were transferred by reassortment from an avian virus to the then-circulating human influenza virus, to generate the 1957 and 1968 pandemic strains. The acquisition of an avian influenza PB1 gene by reassortment might result in increased transcriptional activity of the RNA-dependent RNA polymerase and increased virus replication efficiency of a new pandemic strain[49,81].

The HA and its binding preference for particular sialic acid (SA)-terminated glycans has also been implicated in efficient transmission of the 1918 virus in ferrets (T.M.T. *et al.*)[82]. It has been generally suspected that a switch in receptor-binding preference that confers efficient transmission among humans would be a necessary step for avian influenza viruses in the generation of a pandemic virus. Notably, we (T.M.T. and colleagues) have found that mutation of two amino acid residues (D190E, D225G) in the HA, which was previously identified as sufficient to switch the receptor-binding preference of parental 1918 HA (α2,6 SA receptor preference) to the avian α2,3 SA receptor preference[44,83], prevented transmission among ferrets without affecting the replication efficiency of the rescued 1918 virus[82]. These findings suggest that changes in receptor binding of avian influenza viruses could potentially move them one step closer to a pandemic phenotype.

The viral genotypic basis of the 1918 pandemic virus's virulence and transmissibility has not yet been fully mapped; however, by making chimeric viruses containing at least one 1918 influenza virus gene segment, and by targeted mutagenesis or gene synthesis, future experiments should help us to determine how this pandemic virus killed and spread so efficiently. Such knowledge may help us to elucidate virulence factors for other influenza viruses such as the 2009 influenza pandemic and, thereby, help us to identify targets for future drug intervention.

Whole-genome syntheses of other RNA viruses

Apart from poliovirus and influenza virus, the complete genomes of several other RNA viruses have recently been chemically synthesized. These include human endogenous retrovirus, HIVcpz and SARS-like coronavirus.

Reconstitution of an infectious, human endogenous retrovirus. Of the 3×10^9 bp hat constitute the human genome, nearly 8% (that is, 2.8×10^8 bp), comprise sequences of retroviral origin[84,85]. After having invaded the chromosome of human germ cells, they were inherited for millennia in a mendelian manner; thus, they are viral fossils, but the function of these remnants in human evolution, physiology and disease remains unclear. Most of the genes or gene fragments are, however, inactive owing to various replication errors during proliferation of the host cells. An exception is the *env* gene, which seems to be conserved because it may have a crucial role in hominoid placental physiology[84,85]. Nevertheless, all of the ancient human retroviruses are degenerate, including the human mouse mammary tumor virus–like 2 provirus (HML-2) of the human endogenous retrovirus (HERV) K proviruses (HERV-K(HML-2)). The latter may have been added to the Old World primate genomes relatively recently in human evolution, but no functional proviruses able to produce infectious particles have been isolated.

To reconstruct a replicating retrovirus that may resemble the ancestor of HERV-K sequences, Lee and Bieniasz designed a consensus genome (9,472 nt) and, using whole-genome synthesis, generated the proviral clone HERV-K_{CON}, which "likely resembles the progenitor of HERV-K(HML-2) variants that entered the human genome within

the last few million years"[51]. In a parallel study, Dewannieux *et al.*[86] also reconstructed infectious HERV-K(HML-2) from a consensus sequence, but they applied site-directed mutagenesis to arrive at an infectious provirus, which they named *Phoenix*.

In both studies, the first human retrovirus of endogenous origin had all the properties of a C-type retrovirus. The infectivity of ancestral retrovirus in various cell types, however, was extremely low, which to some extent dispelled concerns that resuscitating an ancient human infectious virus is inherently risky[87]. Still, studying the pathogenic potential of a virus that probably circulated in the then-human population for millions of years may yield valuable clues as to its impact on human evolution.

Synthesis of HIVcpz—the origin of the HIV-1 pandemic. It was long suspected that chimpanzees provided the natural reservoir for the human immunodeficiency viruses that caused the zoonotic infections responsible for the AIDS pandemic. But because the simian immunodeficiency virus most closely related to HIV-1 (SIVcpz) was found only in animals (*Pan troglodytes troglodytes*) in captivity, direct proof was lacking.

In 2006, Hahn and colleagues[88] provided the first convincing evidence of SIVcpz antibodies and nucleic acid in fecal samples from wild *P. T. troglodytes* in a narrow area in south-eastern Cameroon. However, recovery of replication-competent virus from fecal samples had failed. These authors therefore analyzed virus-specific nucleic acids isolated from the fecal samples and obtained a consensus sequence that, when chemically synthesized, yielded infectious molecular clones of SIVcpz[50]. Analyses of these isolates yielded the important result that "naturally occurring SIVcpz strains already have many of the biological properties required for persistent infections of humans." The authors conclude that "medically important 'SIV isolates' that have thus far eluded investigation… are needed to identify viral determinants that contribute to cross-species transmission and host adaptation"[50].

Synthesis of infectious bat SARS-like coronavirus. In 2002, a new acute respiratory syndrome emerged in China, caused by an unknown infectious agent. By the summer of 2003, the agent had caused disease in 8,427 people, of whom 813 died, and fears of a deadly pandemic spread around the globe. As a result of unprecedented collaborative efforts, led by the World Health Organization (Geneva), the pathogenic agent was rapidly identified as a new coronavirus, named severe acute respiratory syndrome virus coronavirus, or SARS-CoV.

Coronaviruses are plus-strand RNA viruses with the largest-known RNA genome (~30 kb). The properties (genome sequence, cultivation and serology) and pathogenic potential of SARS-CoV were rapidly established, but, intriguingly, in July 2003 SARS-CoV disappeared (http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5228a4.htm) as quickly as it had emerged. This lucky break happened despite the fact that there were no drugs, let alone vaccines, available to treat or prevent SARS infection. Isolation of patients, an old medical practice, has been credited with the fading of the SARS-CoV epidemics.

Fear remained that SARS-CoV might reappear, perhaps more contagious than before. Thus, the source of SARS-CoV became an important issue, as it was suspected that the human agent may have evolved from a zoonotic infection, as true of influenza and HIV. Early evidence implicated the Chinese cat-like mammals known as civets, but overwhelming evidence now suggests that "bats are natural reservoirs of SARS-like coronaviruses"[89].

As yet, there is no known tissue culture system that supports the replication of bat SARS virus, suggesting that it is not infectious in humans[52]. However, this inability to culture the virus also prevents investigation of the mechanism of cross-species transmission from bats to civets to humans (or transmission directly from bats to humans). The nucleotide sequence of the bat SARS virus is known. To determine the possible steps by which the bat SARS-CoV may have adapted to human populations, Denison and colleagues[52] synthesized the 29.7-bp bat SARS virus cDNA. They subsequently succeeded in converting the bat SARS-CoV to an infectious clone by exchanging the region encoding its receptor-binding domain (RBD) with that of the human SARS-CoV. The result is the largest replicating genome to be synthesized so far[52]. The authors conclude that "rational design, synthesis, recovery of hypothetical recombinant virus can be used to investigate mechanisms of *trans* species movement of zoonoses and has great potential to aid in rapid public health responses…"[52].

Whole-genome synthesis of DNA viruses

To date, the complete genome of only one DNA virus—ΦX174—has been assembled by synthesis. Other work has applied DNA synthesis to understanding the structure and function of bacteriophage T7 DNA, but this involved the creation of portions, rather than the entire reconstitution, of the viral genome (see "Refactoring the bacteriophage T7 genome" below).

Eighteen months after the poliovirus synthesis, Smith *et al.*[66] described the *de novo* synthesis of the first DNA virus genome—the 5,386-bp genome of bacteriophage ΦX174[66]. Remarkably, existing methods of DNA synthesis were fine-tuned to complete the genome in 2 weeks. The unedited DNA was then transfected into bacteria, which sorted the good from the bad and produced viable bacteriophages[66]. This elegant work confirmed the general utility of DNA synthesis in assembling the genomes of viruses that was first shown with poliovirus, and it has subsequently spurred further work to make larger DNA assemblies, allowing the synthesis of whole bacterial chromosomes.

Virus attenuation by large-scale recoding

Synthetic biology strives to generate new biological systems that do not exist in nature, primarily for medical or commercial applications. The poliovirus synthesis described above was not intended to be an example and, indeed, hardly falls into the category of synthetic biology, because it resulted in a poliovirus with a nearly identical phenotype as the model wild-type virus. In the following section, we describe experiments that have led to the rational design of vaccine candidates from the poliovirus and influenza viruses.

Upon entry into a host cell, the poliovirus uses its genome as mRNA, the hallmark of all plus-strand RNA viruses[90]. Poliovirus belongs to a large family of human and animal pathogenic viruses, the *Picornaviridae*. These viruses express all of their proteins in the form of a single polypeptide of just over 2,000 amino acids—the polyprotein (**Fig. 2a**). This large precursor polyprotein is co- and post-translationally cleaved by the proteolytic action of two virus proteinases that are, remarkably, embedded in the polyprotein itself (E.W. *et al.*)[91]. A poliovirus polypeptide of 2,209 amino acids can be encoded in about $10^{1,100}$ ways, a number much larger than the number of atoms in the universe. This poses the question of how and why selection has led to one of these possible $10^{1,100}$ sequences that we consider 'wild-type' PV1(M).

It should be noted that poliovirus, like all RNA viruses, is a quasispecies that, in reality, exists in nature as a large swarm of different genotypes[92,93]. This is the consequence of the high error rate of viral RNA synthesis in the absence of proof reading and editing functions. The 'wild-type' sequence in this vast swarm is the genotype that can proliferate most efficiently under the prevailing conditions, where it out-competes all of its related genotypes[92,93].

One of our groups (E.W., S.M. and colleagues)[4,5] has been investigating the effect of genome-scale changes in poliovirus on codon usage. To reduce the complexity of our experiments, we have restricted our investigations to only one-third of the poliovirus genome. This is the

region encoding the P1 capsid precursor. The P1 polyprotein, which consists of 881 amino acids (2,643 nt; **Fig. 2**), can be encoded in 10^{442} ways—still a mind-boggling number.

Codon bias. A major reason for the nearly unlimited possibilities of encoding a protein is the degeneracy in the genetic code (for example, several synonymous codons can specify the same amino acid). However, the preference for a synonymous codon is not the same in *E. coli*, in jellyfish or in human cells; this phenomenon is termed 'codon bias'. For example, in humans, the alanine codon GCC is used four times more frequently than the synonymous codon GCG. The cell's preference of one synonymous codon over another to specify the same amino acid is thought to relate to the abundance of the corresponding cognate tRNAs in the cell. Consequently, rare codons are associated with a suboptimal translation of an mRNA. Codon bias, then, may contribute to the restriction of the abundance of sequences encoding the same protein. Codons used frequently in the jellyfish may be used rarely in human cells, and thus expression of the jellyfish green fluorescent protein (GFP) in human cells is poor unless the codons of the jellyfish gene have been changed to those frequently used in human cells; accordingly, the GFP gene has been 'humanized' to achieve good expression in human cells[94].

To exploit this phenomenon, we (S.M., E.W. and collaborators)[5] have 'dehumanized' the sequence encoding P1 of the poliovirus polyprotein. We chose this segment of the poliovirus genome because we have gathered abundant evidence that the P1 coding sequence does not harbor RNA signals essential for viral proliferation (E.W. and colleagues)[91,95]. For example, (i) the nucleotide sequence of the P1 region can be changed drastically, as long as the amino acid sequence that it encodes is preserved[3–5]; (ii) the P1 coding region can be exchanged with foreign genes (for example, firefly luciferase[96]); or (iii) the P1 coding region can be deleted altogether (in defective interfering particles[97]) without loss of efficient RNA replication. However, changing synonymous codons in the P1 region from frequently used to rarely used codons (that is, 'codon deoptimizing' this segment of viral mRNA) will unbalance the synthesis of the polyprotein without changing its amino acid sequence, resulting in attenuated viruses[3,5].

Because of the existence of a polyprotein in picornaviruses, codon deoptimizing the P1 region (the N-terminal third of the polyprotein) compromises viral replication: fewer ribosomes arrive at the coding region for the essential replication proteins (genomic regions P2 and P3; **Fig. 2a**), and genome replication is thus reduced or shut off altogether. It was not surprising, therefore, that extensive codon deoptimization in virus PV-AB (**Fig. 2b**), harboring 680 changes (out of 2,643 nt) without altering a single encoded amino acid, led to a 'dead' phenotype. However, subcloning individual segments of the recoded P1 segment revived the virus, albeit in attenuated form. Indeed, the subclones are not only inhibited in protein synthesis[5], but their neurovirulence is attenuated in *CD155* tg mice as well[5].

A notable property of the subclones of PV-AB is a marked reduction of their specific infectivity, also observed by Burns and colleagues[3]. Wild-type poliovirus (PV1(M)) has a specific infectivity of one plaque-forming unit (PFU) per 115–130 particles[4,5]; in one of the subclones (PV-AB$^{2470–2954}$),

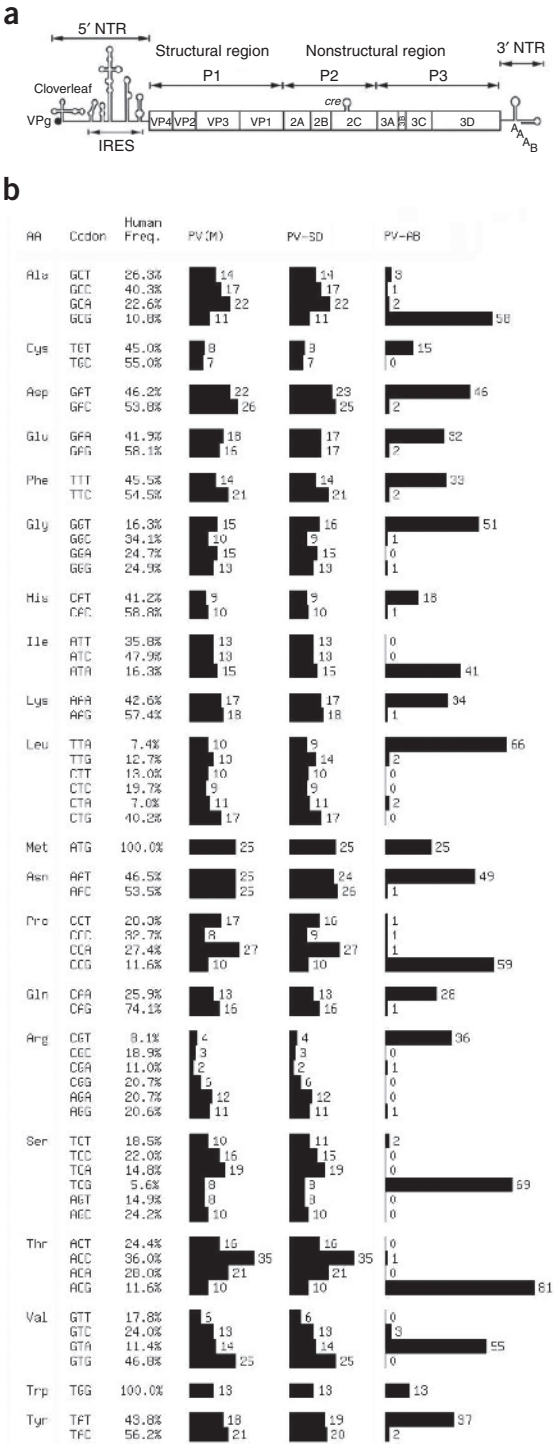


Figure 2 The poliovirus genome and the effect of codon bias. (**a**) Poliovirus genomic RNA[56,91] is of plus-strand polarity (that is, it functions as mRNA in viral replication). It is covalently linked at the 5′ end to the small viral protein VPg (3B of the polyprotein), followed by a long 5′ nontranslated region (5′ NTR), a continuous open reading frame (ORF), a 3′ NTR and poly(A). The 5′ NTR consists of structural elements that control RNA replication (cloverleaf) and translation (the internal ribosomal entry site (IRES)). The ORF encodes the polyprotein, the single translation product of the viral mRNA. The polyprotein is proteolytically processed by viral proteinases 2A^{pro} and 3C/3CD^{pro} into functional proteins, which have been divided into the structural region (P1 and the capsid precursors) and the nonstructural regions P2 and P3 (replication proteins). The ORF is followed by a 3′ NTR, which contributes to the control of RNA synthesis, and poly(A). The P1 coding region has been a target for codon and codon pair deoptimization. (**b**) Codon use statistics in synthetic P1 capsid designs. PV-SD maintains nearly identical codon frequencies compared to wild-type PV1(M), while maximizing codon positional changes within the sequence[5]. In PV-AB capsids, the use of nonpreferred codons was maximized. The length of the bars and the numbers behind each bar indicate the occurrence of each codon in the sequence. As a reference, the normal human synonymous codon frequencies for each amino acid are given. Adapted from S.M. *et al.*[5].

this ratio was reduced to about 1 PFU per 100,000[5] particles. That is, only one plaque can be expected to emerge if 10^5 particles are plated onto a dish of 10^6–10^7 HeLa cells. Once this one virus has succeeded in overcoming the host cell, however, its burst size will be only one order of magnitude lower than that of the wild-type poliovirus. That is, although the dehumanized virus can replicate in HeLa cells, once released it has enormous problems in spreading to other cells. It should be noted that we have analyzed the sequences of the codon-deoptimized viruses for the emergence of higher-order structures that could have impeded replication. No such structures have been found. The properties of the virus with a 'scrambled' P1 region (PV-SD; **Fig. 2b**) are discussed below.

Codon pair bias. It has been known since 1989 that in addition to, and independently of, codon usage, pairs of synonymous codons do not exist in the genome at the frequency that one might expect on the basis of the frequency of the two individual codons that make up the pair. This phenomenon, called 'codon pair bias', was discovered in prokaryotic cells[98] but has since been seen in all other examined species, including humans[99]. For example, given the known codon frequencies in humans, the amino acid pair Ala-Glu is expected to be encoded by GCC GAA and GCA GAG about equally often. In fact, the codon pair GCC GAA is strongly underrepresented, despite containing the most frequent alanine codon, such that it is used only one-seventh as often as GCA GAG[4]. The functional significance of codon pair bias is a mystery, but it can be studied in systems, such as poliovirus, in which large-scale changes of codon pairing are likely to present with phenotypes in viral proliferation.

On the basis of 14,795 annotated (known) human genes, the Wimmer group has calculated a codon pair score, specific for each of the possible 3,721 codon pair combinations, as well the codon pair bias (CPB) for each gene, taking into consideration the codon frequency for each of the paired codons and the frequency of the encoded amino acid pair. In **Figure 3**, the calculated CPB of a human gene is plotted against its amino acid length. Underrepresented codon pairs yield negative scores. Wild-type PV1(M) shows a slightly negative score (CPB = –0.02), but, not surprisingly, it uses codon-pairing corresponding to human genes (**Fig. 3**). Using a custom-made computer optimization algorithm, we have constructed polioviruses whose P1 coding region had either a substantially negative codon pair bias, containing many underrepresented codon pairs (PV-Min, CPB = –0.474), or a substantially positive codon pair bias, containing many overrepresented codon pairs (PV-Max, CPB = +0.246; **Fig. 2**), while retaining the exact set of codons present in the wild-type virus[4] and the same amino acid sequence of P1.

Unexpectedly, transcripts of PV-Min, using underrepresented codon pairs, did not yield virus upon transfection and blind passages. Apparently, the accumulation of hundreds of unfavorable codon pairs led to a dead phenotype ('death by a thousand cuts'). Conversely, various subclones carrying segments of the P1 region of PV-Min cloned into wild-type poliovirus (for example, PV-MinXY or PV-MinZ, with CPB scores of –0.32 and –0.19, respectively; **Fig. 3**) were viable, albeit severely debilitated, as revealed by plaque assays and single-step growth kinetics experiments, and their neurovirulence in *CD155* tg mice was reduced 1,000-fold[4]. Similar to the reduced specific infectivity of subclones of PV-AB, that of the PV-Min subclones, PV-Min XY and PV-MinZ, was also reduced to roughly 1 PFU per 10,000 particles[4]. Moreover, the translational activities of the subclone RNAs were impaired, an observation suggesting a relationship between viability and viral protein synthesis[4,5].

This prompted us to investigate whether PV-Max, the variant in which the P1 coding region carried many overrepresented codon pairs, would grow to titers substantially higher than those of wild-type poliovirus or be more neurovirulent than wild-type virus in *CD155* tg mice. It had neither of these phenotypes. PV-Max, therefore, was not a highly virulent variant virus, an observation that suggests that the end product of evolution of poliovirus had already optimized the encoding of polypeptide P1 (ref. 4) and cannot be 'improved'. For an RNA virus, which exists as a quasi-species, this is not unexpected. The signature of poliovirus is the efficient replication of its small genome, which, under optimal conditions, is driven by an irresistible desire to maintain optimal genome structure; that is, during replication, important replication signals are constantly rebuilt, and unnecessary nucleotide sequences (for example, foreign genes and duplications) are deleted. Indeed, the virus has an inexhaustible arsenal to achieve these goals—by exploiting point mutations, by homologous or illegitimate recombination and even by the acquisition of foreign RNA sequences. These considerations do not mean that the sequence of PV-Max is the only other possible sequence that can express wild-type phenotypes in tissue culture and in *CD155* tg mice. On the contrary, there are probably a huge number of sequences with wild-type phenotypes.

Regardless of the changes in usage of rare codons or underrepresented codon pairs, the product of the translational machinery remains the same; however, the efficiency of protein synthesis may be vastly altered. Thus, no matter how many synonymous changes have been introduced into the genome, a virus synthesized in the infected cell will have the same structure and will encode the same replication proteins as the wild-type virus, but it may be substantially disadvantaged in terms of proliferation. Such a variant of a human pathogenic virus may enter the host by its normal route and replicate poorly, but still allow the host to mount an immune response strong enough to induce lasting protective immunity. In other words, a human virus with altered codon usage or altered codon pair usage could possibly serve as a vaccine. Recoding viral genomes, a process that we call 'synthetic attenuated virus engineering' (SAVE), may be a new and rapid route to discover vaccine candidates and prevent viral disease. Indeed, polioviruses harboring underrepresented codons or underrepresented codon pairs are attenuated in *CD155* tg mice. Infection of these mice with a sublethal dose of codon- or codon pair–deoptimized viruses induced an immune response that protected the animals against a lethal dose of the wild-type virus[4,5].

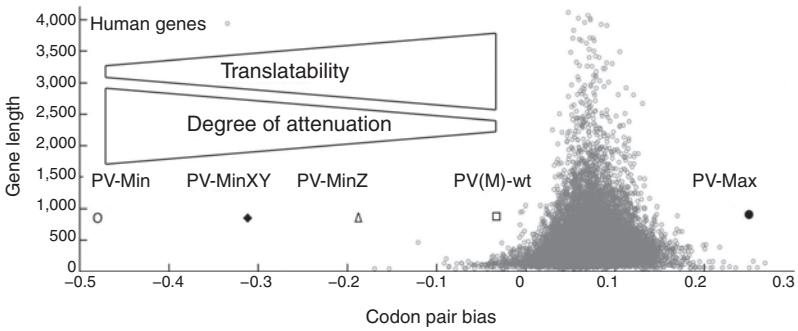


Figure 3 The codon pair bias (CPB) score for each of the 14,795 annotated human genes was calculated[4]. Each dot represents the calculated CPB score of one gene plotted against its amino acid (aa) length. Predominant use of underrepresented codon pairs yields negative CPB scores. Various poliovirus constructs are plotted according to the CPB score of their P1 capsid precursor protein. As the CPB decreases, translatability decreases and the attenuation effect on the virus increases. PV-Min is nonviable; PV-Max expresses replication and virulence phenotypes similar to those of wild-type PV. Adapted from ref. 4.

Box 1 'Dual use' concerns and total synthesis of viruses

As discussed elsewhere[53], the publication of the *de novo* synthesis of poliovirus in the absence of natural template aroused unusually strong and sometimes conflicting responses from different quarters of society. Many of the negative reactions were tainted by fear that the poliovirus synthesis could aid bioterrorism. This was not surprising, as the news of the synthesis reached a public (particularly in the United States) that was highly sensitized to the threat of bioterrorism in the months following the attack on the World Trade Center on September 11, 2001 and the anthrax attacks in Washington and elsewhere in 2001 and 2002. However, these concerns, led to numerous useful public debates about complex issues of biological research, scientific publication and national security[105–109]. Notably, the synthesis of the bacteriophage ΦX174 genome in 2003, which stood out because of its speed of merely 2 weeks[67], or the resurrection of the 1918 'Spanish' influenza virus in 2005 (ref. 80), did not set off the shock waves experienced in 2002, when the poliovirus genome synthesis was published[18]. There may be two reasons for this. First, these later papers were embedded in numerous editorials or, as in the case of the ΦX174 synthesis, in a carefully orchestrated press conference called by the then US Secretary of Energy. These activities were aimed at explaining to the public the significance, particularly the benefit, of the research involving *de novo* virus synthesis. Second, compared to in 2002, the general public was probably better prepared and better educated to accept the new reality of synthetic viruses and their possible consequences[53]. Meanwhile, several publications, of which only a few can be cited[105–109], attest to the serious efforts by the scientific community to define dual-use research and to limit its possible disastrous consequences.

The Wimmer group is now extending the SAVE strategy from poliovirus to the influenza virus because of its enormous importance as a major human pathogen, both in its pandemic and epidemic forms. We also want to know to what extent computer-aided rational design can lead rapidly to vaccine candidates of a virus whose genetic and pathogenic properties are completely different from that of poliovirus. First, influenza virus is a negative-stranded RNA virus with a segmented genome; on entry into the host cell the virion must activate its virus-associated RNA polymerase, which will synthesize genome-complementary RNA for translation and replication. Second, many of the important steps in influenza virus replication occur in the cell nucleus, an environment that is avoided by nearly all other RNA viruses (for example, poliovirus can replicate even in enucleated cells). Third, influenza virus is an enveloped virus that replicates in the respiratory tract. Although this work is still in progress, our results show that, by codon pair deoptimization, it is possible to rapidly construct highly attenuated influenza viruses that, after a single cycle of immunization, protect mice against a lethal dose of wild-type influenza virus (S.M., D. Papamichail, J.R. Coleman, S. Skiena and E.W., unpublished data). Most importantly, the best of the constructed vaccine candidate strains offer a wide margin of safety at a relatively low concentration of inoculating virus (S.M., D. Papamichail, J.R. Coleman, S. Skiena and E.W., unpublished data).

RNA sequences with shuffled codons

So far, we have discussed changing the codon bias (**Fig. 2b**) or codon pair bias (**Fig. 3**) in the P1 region while retaining its amino acid sequence. In a separate approach, the Wimmer group (S.M., E.W. and colleagues)[5] has shuffled synonymous codons of the P1 region according to a specifically designed computer algorithm to maximize the number of nucleotide changes while retaining the existing codons (for example, without changing the codon bias). The resulting P1 sequence was synthesized and cloned into the backbone of poliovirus, yielding PV-SD[5] (**Fig. 2b**). The P1 coding region of PV-SD contained 934 changes out of 2,643 nucleotides; that is, on average, every third nucleotide was different from that of the wild-type sequence. Notably, PV-SD replicated in HeLa cells with wild-type kinetics[5]. Apparently, the positioning of the synonymous codons in PV-SD, after the extensive codon shuffle, did not influence viral protein synthesis, confirming that synonymous mutations have an effect on the virus only when they are specifically directed to lower the codon bias[5] or codon pair bias[4]. An intriguing conclusion from PV-SD is that, at the genome level, RNA viruses are promiscuous with respect to nucleotide changes, as long as these changes do not affect protein function or the rate of protein synthesis. In other words, RNA structures throughout much of the viral coding regions, with notable exceptions, such as *cis*-acting replication elements or encapsidation signals, are probably rather inconsequential.

As pointed out above, a protein of 881 amino acids (the P1 capsid precursor of poliovirus) can be encoded in about 10^{442} different ways. How many of these sequences, if 'cloned' into the present-day poliovirus, would express a wild-type phenotype that would be stable if passaged in HeLa cells, the preferred substrate for poliovirus in the laboratory? We have not passaged PV-SD in HeLa cells for numerous generations in an attempt to observe genetic variation toward the sequence of wild-type PV1(M). Such an experiment may not yield relevant results, because PV1(M) may also express a genetic drift. After all, the natural human cells for poliovirus proliferation are not known, but they reside in the gastrointestinal tract and are obviously very different from HeLa cells, which have been derived from a cervical cancer.

We have used shuffled sequences to test for unknown *cis*-acting RNA elements in the poliovirus genome. The fact that PV-SD replicates with wild-type kinetics provided proof that the P1 region is void of essential *cis*-acting replication elements, which would probably have been destroyed by the large-scale shuffling. Note that other human picornaviruses, such as rhinovirus type 14 (HRV14), do contain such an essential element (*cre* in HRV14—a stem-loop structure of ~50 nt) in P1. Hence, HRV14-SD would have probably been nonviable[100]. The poliovirus equivalent to the HRV14 *cre* maps to the P2 coding region[101]. We expected that scrambling the P2 coding sequence of the poliovirus genome would destroy the *cre* element and kill the virus, which is indeed what we have found (Y. Song, C. Ward, D. Futcher, S. Skiena, E.W. and S.M., unpublished data). Re-establishing the *cre* sequence in P2 by molecular engineering rescues the virus, which indicates that the P2 region does not contain any other important RNA sequences in addition to *cre*. Synonymous scrambling of RNA virus sequences may be an excellent tool in searching for RNA structures that are essential for viral replication.

Refactoring the bacteriophage T7 genome

"Refactoring [is] a process that is typically used to improve the design of legacy computer software"[102]. Endy and colleagues[2] have used this definition to describe their efforts to redesign the DNA bacteriophage T7 genome (39,937 bp[103]) with an aim to test the functions of a set of T7 genes once they are untangled from each other (by removing overlapping gene segments). To this end, they replaced the left 11,515 bp of the wild-type genome with 12,179 bp of synthetic redesigned DNA (available in cassettes α and β) and tested the biological properties of the synthetic DNA when combined with the remainder of the wild-type (WT) genome. Notably, three chimera, α-WT, WT-β-WT and α-β-WT, were viable, but, perhaps not surprisingly to the

investigators, none grew as well as wild-type T7 phage (although no burst size was included in the report)[2].

Phage α-β-WT, called T7.1, represents a redesign of >30% of the T7 genome and, by removing overlaps, the genes in the α-β region could be studied independently, an enormous advantage for genetic analyses. T. F. Knight described the work as "the most compelling example of work in synthetic biology to date"[104]. From a scientific perspective, Endy's work demonstrated that overlapping genetic elements in the T7 genome were, in aggregate, non-essential for phage replication. Until these experiments, the community had been stressing the importance of these features, given that they are conserved across evolutionary distance, but the synthetic approach provided a way of clarifying this issue. From an engineering perspective, this work provided confidence that up to 5% of the DNA sequence of an organism can be changed while still maintaining its viability.

Conclusions

The ability to manipulate the genomes of viruses has long been important as a model for molecular systems, for investigating viral pathogenesis and for the production of viral vaccines. The methods of molecular biology and the utility of reverse genetics allow the rapid production of altered viruses from cloned viral genes, including those that are important for public health. Until recently, these methods have relied on PCR or RT-PCR amplification of templates from the pre-existing virus, followed by subcloning into the appropriate plasmid vectors, with or without mutagenesis, and such techniques will continue to be invaluable for virology and vaccinology. However, the advances in gene synthesis, coupled with the ability to use the techniques described in this Review, have allowed the production of viruses in the absence of available infectious virus. This has implications not only in terms of dual use (**Box 1**), but also for our understanding of evolution and the properties of important pathogens. Moreover, genome synthesis of both DNA and RNA viruses will lead to unprecedented possibilities in modifying naturally occurring genomes, thereby allowing new studies of viral genome architecture, viral gene expression and gene function. The examples presented in this Review are only the beginning of a new era in which genome synthesis is likely to dominate genetic experiments with viruses.

ACKNOWLEDGMENTS

We are indebted to our colleagues who have participated in the work described here and who have in part edited the manuscript, particularly A. Paul and B. Futcher, and we thank J. Shendure, L. Steward and A.B. Burgin for information provided. The work described here was supported partially by US National Institutes of Health (NIH) grants AI075219 and AI15122 and contract N65236 from the US Defense Advanced Research Project Agency to E.W.; and partially by the intramural research program of the NIH and the National Institute of Allergies and Infectious Diseases (NIAID). The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the funding agency.

Published online at http://www.nature.com/naturebiotechnology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Keasling, J. The promise of synthetic biology. *The Bridge* **35**, 18–21 (2005).
2. Chan, L.Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).
3. Burns, C.C. *et al.* Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J. Virol.* **80**, 3259–3272 (2006).
4. Coleman, J.R. *et al.* Virus attenuation by genome-scale changes in codon pair bias. *Science* **320**, 1784–1787 (2008).
5. Mueller, S., Papamichail, D., Coleman, J.R., Skiena, S. & Wimmer, E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J. Virol.* **80**, 9687–9696 (2006).
6. Wöhler, F. Ueber die künstliche Bildung des Harnstoffs. *Ann. Phys.* **12**, 253–256 (1828) (in German).
7. Kinne-Saffran, E. & Kinne, R.K. Vitalism and synthesis of urea. From Friedrich Wohler to Hans A. Krebs. *Am. J. Nephrol.* **19**, 290–294 (1999).
8. Miescher, F. Ueber der chemische Zusammensetzung der Eiterzellen. *Med.-Chem. Unters.* **4**, 441–460 (1871) (in German).
9. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
10. Brown, D.M. & Todd, A.R. in *The Nucleic Acids* Vol. 1 (eds. Chargaff, E. & Davidson, J.N.) 409–430 (Academic, New York, 1955).
11. Agarwal, K.L. *et al.* Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227**, 27–34 (1970).
12. Khorana, H.G. Total synthesis of a gene. *Science* **203**, 614–625 (1979).
13. Caruthers, M.H. Gene synthesis machines: DNA chemistry and its uses. *Science* **230**, 281–285 (1985).
14. Stewart, L. & Burgin, A.B. Whole gene synthesis: a gene-o-matic future. *Front. Drug Des. Disc.* **1**, 297–341 (2005).
15. Sanghvi, Y. A roadmap to the assembly of synthetic DNA from raw materials. in *Working Papers for Synthetic Genomics: Risks and Benefits for Science and Society* (eds. Garfinkel, M.S., Endy, D., Epstein, G.L. & Friedman, R.M.) 17–33 (2007).
16. Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. & Heyneker, H.L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
17. Pan, W. *et al.* Vaccine candidate MSP-1 from *Plasmodium falciparum*: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from *Escherichia coli* and mammalian cells. *Nucleic Acids Res.* **27**, 1094–1103 (1999).
18. Cello, J., Paul, A.V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016–1018 (2002).
19. Gibson, D.G. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
20. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
21. Holley, R.W. *et al.* Structure of a ribonucleic acid. *Science* **147**, 1462–1465 (1965).
22. Penswick, J.R., Martin, R. & Dirheimer, G. Evidence supporting a revised sequence for yeast alanine tRNA. *FEBS Lett.* **50**, 28–31 (1975).
23. Shendure, J.A., Porreca, G.J. & Church, G.M. Overview of DNA sequencing strategies. in *Current Protocols in Molecular Biology.* (ed. Ausubel, F.M. *et al.*) Unit 7.1 (John Wiley and Sons, Hoboken, NJ, USA; 2008).
24. Maxam, A.M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**, 560–564 (1977).
25. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467 (1977).
26. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
27. The International Human Genome Mapping Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
28. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
29. Wheeler, D.A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
30. Pennisi, E. Personal genomics. Number of sequenced human genomes doubles. *Science* **322**, 838 (2008).
31. Taniguchi, T., Palmieri, M. & Weissmann, C. Qβ DNA-containing hybrid plasmids giving rise to Qβ phage formation in the bacterial host. *Nature* **274**, 223–228 (1978).
32. Weissmann, C., Weber, H., Taniguchi, T., Muller, W. & Meyer, F. Reversed genetics: a new approach to the elucidation of structure–function relationship. *Ciba Found. Symp.* **66**, 47–61 (1979).
33. Racaniello, V.R. & Baltimore, D. Cloned poliovirus complementary DNA is infectious in mammalian cells. *Science* **214**, 916–919 (1981).
34. Schnell, M.J., Mebatsion, T. & Conzelmann, K.K. Infectious rabies viruses from cloned cDNA. *EMBO J.* **13**, 4195–4203 (1994).
35. Collins, P.L. *et al.* Production of infectious human respiratory syncytial virus from cloned cDNA confirms an essential role for the transcription elongation factor from the 5′ proximal open reading frame of the M2 mRNA in gene expression and provides a capability for vaccine development. *Proc. Natl. Acad. Sci. USA* **92**, 11563–11567 (1995).
36. Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl. Acad. Sci. USA* **96**, 9345–9350 (1999).
37. Fodor, E. *et al.* Rescue of influenza A virus from recombinant DNA. *J. Virol.* **73**, 9679–9682 (1999).
38. Takeuchi, K., Takeda, M. & Miyajima, N. Toward understanding the pathogenicity of wild-type measles virus by reverse genetics. *Jpn. J. Infect. Dis.* **55**, 143–149 (2002).
39. Neumann, G., Feldmann, H., Watanabe, S., Lukashevich, I. & Kawaoka, Y. Reverse genetics demonstrates that proteolytic processing of the Ebola virus glycoprotein is not essential for replication in cell culture. *J. Virol.* **76**, 406–410 (2002).
40. Överby, A.K., Popov, V., Neve, E.P. & Pettersson, R.F. Generation and analysis of infectious virus-like particles of uukuniemi virus (bunyaviridae): a useful system for studying bunyaviral packaging and budding. *J. Virol.* **80**, 10428–10435 (2006).
41. Komoto, S., Sasaki, J. & Taniguchi, K. Reverse genetics system for introduction of site-specific mutations into the double-stranded RNA genome of infectious rotavirus. *Proc. Natl. Acad. Sci. USA* **103**, 4646–4651 (2006).
42. Blight, K.J., Kolykhalov, A.A. & Rice, C.M. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972–1974 (2000).

43. Taubenberger, J.K., Reid, A.H., Krafft, A.E., Bijwaard, K.E. & Fanning, T.G. Initial genetic characterization of the 1918 "Spanish" influenza virus. *Science* **275**, 1793–1796 (1997).
44. Reid, A.H., Fanning, T.G., Hultin, J.V. & Taubenberger, J.K. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. *Proc. Natl. Acad. Sci. USA* **96**, 1651–1656 (1999).
45. Reid, A.H., Fanning, T.G., Janczewski, T.A. & Taubenberger, J.K. Characterization of the 1918 "Spanish" influenza virus neuraminidase gene. *Proc. Natl. Acad. Sci. USA* **97**, 6785–6790 (2000).
46. Basler, C.F. *et al.* Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. *Proc. Natl. Acad. Sci. USA* **98**, 2746–2751 (2001).
47. Reid, A.H., Fanning, T.G., Janczewski, T.A., McCall, S. & Taubenberger, J.K. Characterization of the 1918 "Spanish" influenza virus matrix gene segment. *J. Virol.* **76**, 10717–10723 (2002).
48. Reid, A.H., Fanning, T.G., Janczewski, T.A., Lourens, R.M. & Taubenberger, J.K. Novel origin of the 1918 pandemic influenza virus nucleoprotein gene. *J. Virol.* **78**, 12462–12470 (2004).
49. Taubenberger, J.K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
50. Takehisa, J. *et al.* Generation of infectious molecular clones of simian immunodeficiency virus from fecal consensus sequences of wild chimpanzees. *J. Virol.* **81**, 7463–7475 (2007).
51. Lee, Y.N. & Bieniasz, P.D. Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.* **3**, e10 (2007).
52. Becker, M.M. *et al.* Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl. Acad. Sci. USA* **105**, 19944–19949 (2008).
53. Wimmer, E. The test-tube synthesis of a chemical called poliovirus. The simple synthesis of a virus has far-reaching societal implications. *EMBO Rep.* **7**, S3–S9 (2006).
54. De Jesus, N., Franco, D., Paul, A., Wimmer, E. & Cello, J. Mutation of a single conserved nucleotide between the cloverleaf and internal ribosome entry site attenuates poliovirus neurovirulence. *J. Virol.* **79**, 14235–14243 (2005).
55. Toyoda, H., Yin, J., Mueller, S., Wimmer, E. & Cello, J. Oncolytic treatment and cure of neuroblastoma by a novel attenuated poliovirus in a novel poliovirus-susceptible animal model. *Cancer Res.* **67**, 2857–2864 (2007).
56. van der Werf, S., Bradley, J., Wimmer, E., Studier, F.W. & Dunn, J.J. Synthesis of infectious poliovirus RNA by purified T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **83**, 2330–2334 (1986).
57. Molla, A., Paul, A.V. & Wimmer, E. Cell-free, *de novo* synthesis of poliovirus. *Science* **254**, 1647–1651 (1991).
58. Jiang, P. *et al.* Evidence for emergence of diverse polioviruses from C-cluster coxsackie A viruses and implications for global poliovirus eradication. *Proc. Natl. Acad. Sci. USA* **104**, 9457–9462 (2007).
59. Claverie, J.M. Viruses take center stage in cellular evolution. *Genome Biol.* **7**, 110 (2006).
60. Villarreal, L.P. Are viruses alive? *Sci. Am.* **291**, 100–105 (2004).
61. Ryan, F.P. Viruses as symbionts. *Symbiosis* **44**, 11–21 (2007).
62. Koonin, E.V., Senkevich, T.G. & Dolja, V.V. Compelling reasons why viruses are relevant for the origin of cells. *Nat. Rev. Microbiol.* **7**, 615 (2009).
63. Claverie, J.M. & Ogata, H. Ten good reasons not to exclude viruses from the evolutionary picture. *Nat. Rev. Microbiol.* **7**, 615 (2009).
64. Kitamura, N. *et al.* Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature* **291**, 547–553 (1981).
65. Racaniello, V.R. & Baltimore, D. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proc. Natl. Acad. Sci. USA* **78**, 4887–4891 (1981).
66. Smith, H.O., Hutchison, C.A., III, Pfannkoch, C. & Venter, J.C. Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445 (2003).
67. Taubenberger, J.K. & Morens, D.M. 1918 Influenza: the mother of all pandemics. *Emerg. Infect. Dis.* **12**, 15–22 (2006).
68. Taubenberger, J.K., Morens, D.M. & Fauci, A.S. The next influenza pandemic: can it be predicted? *J. Am. Med. Assoc.* **297**, 2025–2027 (2007).
69. Taubenberger, J.K., Hultin, J.V. & Morens, D.M. Discovery and characterization of the 1918 pandemic influenza virus in historical context. *Antivir. Ther.* **12**, 581–591 (2007).
70. Pekosz, A., He, B. & Lamb, R.A. Reverse genetics of negative-strand RNA viruses: closing the circle. *Proc. Natl. Acad. Sci. USA* **96**, 8804–8806 (1999).
71. Hoffmann, E., Neumann, G., Hobom, G., Webster, R.G. & Kawaoka, Y. "Ambisense" approach for the generation of influenza A virus: vRNA and mRNA synthesis from one template. *Virology* **267**, 310–317 (2000).
72. Tumpey, T.M. *et al.* Existing antivirals are effective against influenza viruses with genes from the 1918 pandemic virus. *Proc. Natl. Acad. Sci. USA* **99**, 13849–13854 (2002).
73. Kash, J.C. *et al.* Global host immune response: pathogenesis and transcriptional profiling of type A influenza viruses expressing the hemagglutinin and neuraminidase genes from the 1918 pandemic virus. *J. Virol.* **78**, 9499–9511 (2004).
74. Kobasa, D. *et al.* Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. *Nature* **431**, 703–707 (2004).
75. Kash, J.C. *et al.* Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* **443**, 578–581 (2006).
76. Pappas, C. *et al.* Single gene reassortants identify a critical role for PB1, HA, and NA in the high virulence of the 1918 pandemic influenza virus. *Proc. Natl. Acad. Sci. USA* **105**, 3064–3069 (2008).
77. Tumpey, T.M. *et al.* Pathogenicity and immunogenicity of influenza viruses with genes from the 1918 pandemic virus. *Proc. Natl. Acad. Sci. USA* **101**, 3166–3171 (2004).
78. Tumpey, T.M. *et al.* Pathogenicity of influenza viruses with genes from the 1918 pandemic virus: functional roles of alveolar macrophages and neutrophils in limiting virus replication and mortality in mice. *J. Virol.* **79**, 14933–14944 (2005).
79. Taubenberger, J.K. & Morens, D.M. The pathology of influenza virus infections. *Annu. Rev. Pathol.* **3**, 499–522 (2008).
80. Tumpey, T.M. *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310**, 77–80 (2005).
81. Naffakh, N., Massin, P., Escriou, N., Crescenzo-Chaigne, B. & van der Werf, S. Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. *J. Gen. Virol.* **81**, 1283–1291 (2000).
82. Tumpey, T.M. *et al.* A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* **315**, 655–659 (2007).
83. Stevens, J. *et al.* Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.* **355**, 1143–1155 (2006).
84. Ryan, F.P. Human endogenous retroviruses in health and disease: a symbiotic perspective. *J. R. Soc. Med.* **97**, 560–565 (2004).
85. Bannert, N. & Kurth, R. Retroelements and the human genome: new perspectives on an old relation. *Proc. Natl. Acad. Sci. USA* **101**, S14572–S14579 (2004).
86. Dewannieux, M. *et al.* Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res.* **16**, 1548–1556 (2006).
87. Enserink, M. Viral Fossil brought back to life. *Science NOW* **1101**, 4 (2006).
88. Keele, B.F. *et al.* Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523–526 (2006).
89. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
90. Baltimore, D. Expression of animal virus genomes. *Bacteriol. Rev.* **35**, 235–241 (1971).
91. Wimmer, E., Hellen, C.U. & Cao, X. Genetics of poliovirus. *Annu. Rev. Genet.* **27**, 353–436 (1993).
92. Holland, J. *et al.* Rapid evolution of RNA genomes. *Science* **215**, 1577–1585 (1982).
93. Eigen, M. Viral quasispecies. *Sci. Am.* **269**, 42–49 (1993).
94. Zolotukhin, S., Potter, M., Hauswirth, W.W., Guy, J. & Muzyczka, N. A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. *J. Virol.* **70**, 4646–4654 (1996).
95. Gromeier, M., Wimmer, E. & Gorbalenya, A.E. Genetics, pathogenesis and evolution of picornaviruses. in *Origin and Evolution of Viruses* (eds. Domingo, E., Webster, R.G. & Holland, J.J.) 287–343 (Academic, New York, 1999).
96. Porter, D.C. *et al.* Demonstration of the specificity of poliovirus encapsidation using a novel replicon which encodes enzymatically active firefly luciferase. *Virology* **243**, 1–11 (1998).
97. Kuge, S., Saito, I. & Nomoto, A. Primary structure of poliovirus defective-interfering particle genomes and possible generation mechanisms of the particles. *J. Mol. Biol.* **192**, 473–487 (1986).
98. Gutman, G.A. & Hatfield, G.W. Nonrandom utilization of codon pairs in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **86**, 3699–3703 (1989).
99. Moura, G. *et al.* Large scale comparative codon-pair context analysis unveils general rules that fine-tune evolution of mRNA primary structure. *PLoS ONE* **2**, e847 (2007).
100. McKnight, K.L. & Lemon, S.M. Capsid coding sequence is required for efficient replication of human rhinovirus 14 RNA. *J. Virol.* **70**, 1941–1952 (1996).
101. Goodfellow, I. *et al.* Identification of a *cis*-acting replication element within the poliovirus coding region. *J. Virol.* **74**, 4590–4600 (2000).
102. Fowler, M., Beck, K., Brant, J., Opdyke, W. & Roberts, D. *Refactoring: Improving the Design of Existing Code* (Addison-Wesley, Boston, 1999).
103. Dunn, J.J. & Studier, F.W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166**, 477–535 (1983).
104. Knight, T.F. Engineering novel life. *Mol. Syst. Biol.* **1**, 2005 0020 (2005).
105. Atlas, R. *et al.* Statement on the consideration of biodefence and biosecurity. *Nature* **421**, 771 (2003).
106. Atlas, R. *et al.* Statement on scientific publication and security. *Science* **299**, 1149 (2003).
107. National Research Council. *Biotechnology in an Age of Terrorism* (The National Academies Press, Washington, DC, 2004).
108. Bügl, H. *et al.* DNA synthesis and biological security. *Nat. Biotechnol.* **25**, 627–629 (2007).
109. Garfinkel, M., Endy, D., Epstein, G. & Friedman, R. Synthetic genomics: options for governance. *Biosecur. Bioterror.* **5**, 359–362 (2007).