ELSEVIER

Genetic parts to program bacteria

Christopher A Voigt

Genetic engineering is entering a new era, where microorganisms can be programmed using synthetic constructs of DNA encoding logic and operational commands. A toolbox of modular genetic parts is being developed, comprised of cell-based environmental sensors and genetic circuits. Systems have already been designed to be interconnected with each other and interfaced with the control of cellular processes. Engineering theory will provide a predictive framework to design operational multicomponent systems. On the basis of these developments, increasingly complex cellular machines are being constructed to build specialty chemicals, weave biomaterials, and to deliver therapeutics.

Addresses
Biophysics and Chemistry & Chemical Biology, Department of Pharmaceutical Chemistry, University of California San Francisco, QB3 Box 2540, 1700 4th Street, San Francisco, CA 94158, USA

Corresponding author: Voigt, Christopher A (cavoigt@picasso.ucsf.edu)

Current Opinion in Biotechnology 2006, **17**:548–557

This review comes from a themed issue on
Tissue and cell engineering
Edited by James L Sherley

Available online 15th September 2006

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2006.09.001

Introduction

The genome contains commands dictating how cells eat, reproduce, communicate, move and interact with their environment. Cells can be programmed by introducing synthetic DNA containing new commands that instruct the cell to perform a set of artificial tasks in series or in parallel. These programs consist of multiple genes and regulatory elements that function as a system composed of sensors, circuits and converters to control biological responses.

A rudimentary language is emerging to genetically program bacteria [1]. Sensors have been developed that respond to small molecules, light and temperature [2,3••]. Genetic circuits are available that function as inverters, logic gates, pulse generators, band pass filters and oscillators [4,5,6•,7]. Sender and receiver components enable cells to communicate [8••]. Based on these genetic parts, strains of bacteria have been developed that can communicate to form two-dimensional patterns [8••], control their population density [9••], synthesize antimalarial and cancer-fighting drugs [10,11], and attack malignant cells in response to environmental cues present in a tumor [12••] (Figure 1).

The analogy with electronic parts is useful in constructing genetic circuits that perform signal processing tasks. However, the analogy is less applicable for the design of systems composed of many parts. Genetic parts have problems with interference — where one part inadvertently affects another part — because their functions are carried out by molecular interactions and reactions that occur in the same confined space of the cell. This imposes the restriction that a particular genetic circuit can only be used once in a design. Thus, the language to program cells is going to require redundancy, or breadth, in the available parts.

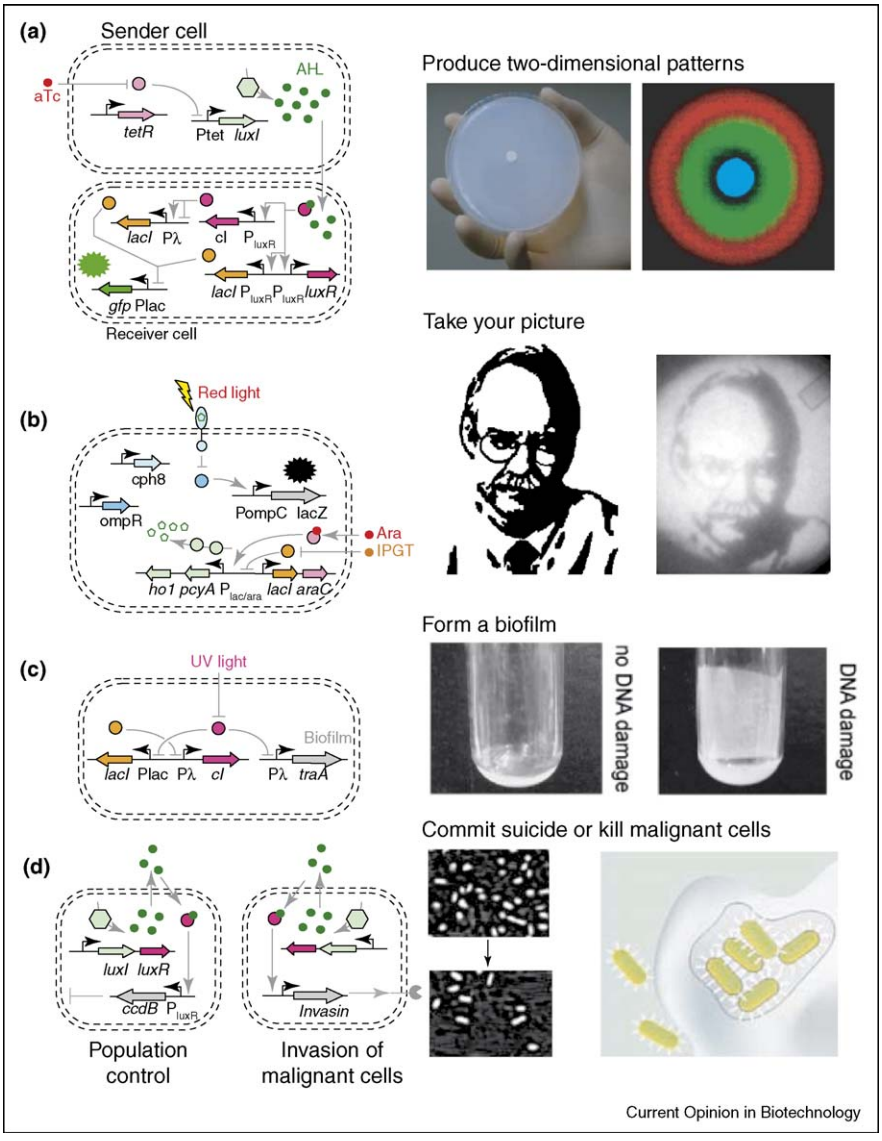
A second problem is that cells are alive. They eat, grow, avoid stress and evolve. Bacteria undergo remarkable changes in cell state as a function of their growth stage. The cell volume, metabolism, membrane composition, and global regulators change in response to the growth media and cell state. All of these factors can impact the function of synthetic sensors and circuits. Some are more fragile than others and recent designs have attempted to build genetic parts whose function is as detached from the cell state as possible. Also, evolution can effectively 'break' a synthetic part by introducing mutations over many generations.

This review has been written to introduce readers to the most robust genetic parts that have been reused in different designs. They have been loosely divided into three categories (Table 1 and Supplementary material). Sensors encompass all means by which information is received by the cell. Genetic circuits represent how information is processed and decisions made. Actuators describe how the circuits and sensors can be used to control processes in the cell. The sequences and performance characteristics for many of these parts are available at the Massachusetts Institute of Technology (MIT) Registry of Standard Biological Parts (http://parts.mit.edu). When given, the part number refers to the Registry numbering system. When available, the transfer function of a genetic part is provided. This is an empirical measurement that describes how the output changes as a function of the input [4,13] (Box 1). The focus of this review is on bacteria, although there has been much recent work in eukaryotes [14].

Sensors and inputs

Cell-based sensors can be used to identify a microenvironment, to direct communication between cells or to

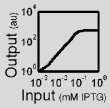
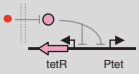
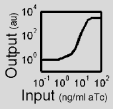
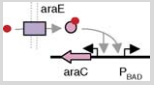
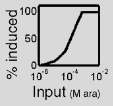
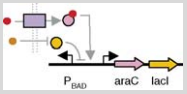
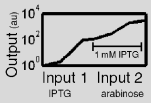
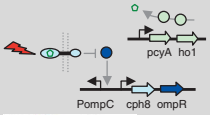
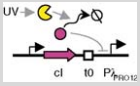
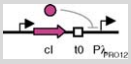
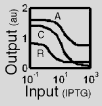
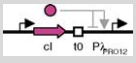
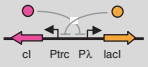
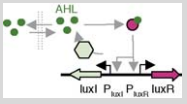
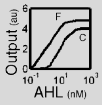
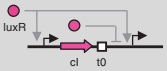
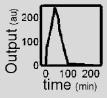
Figure 1



Using genetic parts to program bacteria. Programmed bacteria can **(a)** autonomously form spatial patterns [8••], **(b)** record images of light [3••], **(c)** form a biofilm in response to UV light [35••], and **(d)** commit suicide (left-hand panel) or kill tumor cells (right-hand panel) after reaching a critical population density [9••,12••]. Each design involves the linkage of cellular sensors to the control of biological processes, mediated by genetic circuits. In (a), spatial patterns are formed by using a quorum sensing system to program the communication between bacteria. The enzyme LuxI produces a small molecule (green dots) that diffuses through the cell membrane. Once the molecule accumulates to a sufficient concentration in the media, it binds to a regulatory protein (LuxR). This regulatory protein is then connected to a pulse generator, which controls the expression of green fluorescent protein. Thus, cells only turn green at an intermediate concentration of the signal. This forms rings of gene expression (green, red) around the source of the signal (blue dot). In (b), bacterial photography was achieved using a light-sensing sensor from a cyanobacterial two-component system. The protein domain that responds to light (light blue) was fused to a signal transduction domain from *E. coli* (dark blue). In addition, the metabolic enzymes (green) that produce the required chromophore (pentagons) were included. The output of the light sensor was connected to the expression of an enzyme that turns the media black. In (c), the toggle switch (yellow, magenta) was used to control the expression of a protein that causes the bacteria to form a biofilm. One of the repressors in the toggle switch is sensitive to UV. Thus, in the presence of UV light, the bacteria will form a biofilm. In (d), two similar quorum sensing systems are used to control different responses as outputs. On the left, a gene is controlled that causes the cell to commit suicide (*ccdB*). Once the cell density reaches a critical threshold, the cells begin to die. On the right, a gene is controlled that causes *E. coli* to invade malignant cells (*invasin*). This gene is only turned on when there is a high concentration of bacteria. (Note that the same parts are reused in different designs and appear in Table 1 or Supplementary material.)

Table 1

DNA parts for programming bacteria.

Name	Genes[a]	Performance	Notes	References
Sensors – small molecule inducers				
Lac			• Graded population induction • LacI can exist in the genome or on a plasmid	[64]
Tet			• Intermediate induction difficult • *tetR* can exist in the genome or on a plasmid	[64]
Ara			• All-or-none response • 300-fold induction • Strain must transport (*araE*), but not metabolize (Δ*araBCD*) arabinose • Sensitive to glucose	[65]
Ara–lac			• Dual control by arabinose and IPTG • Many mutants/detailed parameters available	[64,66]
Sensors – environmental inducers				
Light[b]			• Tenfold induction/high basal activity • Responds to red light • Requires *E. coli* RU1012 [16] • cyA/ho1 make the chromophore PCB • cph8 is chimera with EnvZ	[3••]
UV[c]			• The wild-type cI repressor is proteolyzed in response to UV • Turns on after a UV dose of 5 J/m^2	[67]
Genetic circuits – switches and logic				
Inverter			• Transfer function shown with IPTG input driving the cI repressor • Can be built with other repressors (e.g. TetR and LacI) • Mutants available with different transfer functions (e.g. A, C, R)[e] • Can also amplify a signal	[4,68•]
Biphasic switch			• The promoter has multiple cI binding sites with varying affinity that either activate or repress transcription • Pλ off at both low and high input	[33•]
Toggle switch			• All-or-none response • Inputs either small molecules or promoters that are linked to either repressor • Bistable; hysteresis in switch • Epigenetic memory	[34]
Cell–cell communication[d]			• Mutants available with different transfer functions (e.g. F, C)[e] • Parallel communication using *lasIR* and *rhlLR* [68•] • LuxR can also repress promoters [69]	[6•]
Genetic circuits – dynamic responses				
Pulse generator			• The response can be controlled by changing the cI rbs • Is an incoherent feedforward regulatory motif [41]	[6•]
Actuators				
Suicide			• Kills bacterium when expressed	[9••]
Biofilm			• Induces biofilm formation	[35••]
Adhesion/invasion			• Causes *E. coli* to adhere to and invade mammalian cells expressing β1-integrins, including malignant cells	[12••]

au, arbitrary units; ara, arabinose; AHL, actylhomoserinelactone; aTc, anhydrotetracycline; IPTG, isopropyl-β-D-thiogalactopyranoside; PCB, phycocyanobilin; rbs, ribosome-binding site. [a]The following notation is used to describe the genetic parts. Large colored arrows and the corresponding colored circles represent genes and their encoded proteins. The black arrows are promoters. Gray arrows represent activating interactions and lines with blunt ends are repressing interactions. The double dotted line represents the cell membrane. Dots without black borders are small molecules. Small-molecule transporters are shown as purple boxes in the membrane. [b]The green pentagon represents the synthesis of the required chromophore PCB. The red lightning bolt shows the activation of the light sensor with red light. [c]UV light activates RecA, which leads to the proteolytic cleavage (yellow) of cI. [d]LuxI (green hexagon) is an enzyme that produces the quorum signal AHL (green dots). [e]The notation (A, C, R) and (F,C) represent genetic mutants that have different transfer functions. See references for details.

Box 1 The transfer function.

The transfer function is an empirical measurement that describes how the output changes as a function of the input [4,13]. For example, for a light sensor, the transfer function might be the level of gene expression as a function of light intensity [3••]. When the transfer functions are known, they can be used to predict how multiple connected parts will behave as a system.

There is a stochastic component to the transfer function, as was demonstrated beautifully by Elowitz and co-workers [41••]. This arises from cell-to-cell variation, fluctuations owing to small numbers of molecules, and noise in transcription and translation [47,70]. When multiple parts are connected in series, the noise from an upstream part can influence all of the downstream processes [43••]. Stochastic effects represent a challenge to the assembly of parts, especially those which have multiple parts operating in series. Formal mathematics are being developed to incorporate the stochastic component into the transfer function [71•].

Ideally, all of the transfer functions would be measured using a standardized strain, genetic system and reporter. This is often not the case, so these functions are now only a qualitative guide and cannot yet be used to quantitatively predict how multiple circuits will function in series or the interference between circuits operating in parallel. There is currently an effort to standardize the measurement of part performance, which will ultimately enable computer-aided design.

make the system respond to external commands. Sensors can be wired to turn genes on or off or to directly influence cellular behavior; for example, the direction in which it is swimming. There are four major classes of sensors that are commonly engineered: cytoplasmic regulatory proteins, two-component systems, regulatory RNAs, and environment-responsive promoters (Table 1 and Supplementary material).

Cytoplasmic regulatory proteins

Inducible systems allow specific genes to be turned on by adding a small molecule to the growth media. Typically, the inducer passes through the cell membrane and binds to a cytoplasmic regulatory protein. This either turns on an activator or turns off a repressor, leading to the activation or derepression of a promoter, respectively. There are many variations of these systems that are appropriate for different applications.

There are several key parameters that describe the transfer function of an inducible system. First, there is the dynamic range of the induction. This is the difference between the basal activity in the absence of inducer and the maximally induced state. The form of the transfer function is also important; for example, some systems are strongly cooperative, making it difficult to obtain intermediate ranges of expression.

Inducible systems can also have different population-level behaviors. All of the cells in a population can behave identically and expression increases in a graded manner as a function of inducer concentration. By contrast, some systems produce an all-or-none response, where a percentage of the population turn on and a greater fraction of the cells express the gene as more inducer is added.

Two-component systems

Two-component systems represent the most prevalent natural sensing motif in prokaryotes [15]. Bacteria contain many two-component systems that are simultaneously expressed (*Escherichia coli* has 32) and respond to different stimuli, such as light, temperature, touch, metals, metabolites and chemicals. The canonical system consists of a membrane-bound sensor which, when stimulated, phosphorylates a response regulator. The phosphorylated regulator then binds to promoters to activate or repress gene expression.

The intracellular parts of two-component systems are homologous, with similar structures and mechanisms. This homology can be exploited to rewire the system by genetically fusing the extracellular sensing domain to a new, heterologous intracellular signal transduction domain [16,17]. This was recently demonstrated through the construction of a synthetic sensor that gives *E. coli* the ability to see light [3••]. The extracellular domain of a cyanobacterial light sensor was fused to a signal transduction domain from *E. coli*, which was used to control the expression of a gene that produces a black pigment. This strain can record images of light projected at a two-dimensional lawn of growing bacteria (Figure 1).

A modified two-component system has been used to sense metabolic changes [18]. Using a strain where the cognate sensor (NRII) has been knocked out, the NRI protein was used to sense acetyl phosphate, which changes in response to glucose flux. This sensor was used to optimize the production of lycopene by diverting the carbon flux away from the toxic byproduct, acetate. The system has also been used to maximize protein production [19].

The bacterial chemotaxis sensing apparatus has a similar organization to a two-component system. Chimeras have been made between the extracellular domain of tar (a chemosensory protein) and the intracellular domain of EnvZ (a two-component sensor), such that chemo-attractants can be used to regulate gene expression [17,20]. The inverse chimeras can also be made, where the extracellular portion of a two-component system is fused to the intracellular portion of tar (see Supplementary material). This enables cells to move towards the signal received by the two-component systems; for example, a Nar–tar chimera produced cells that move towards nitrate [21]. Mutations affecting the tar ligand-binding site have also been shown to direct chemotaxis towards alternate amino acids [22].

Maltose-binding protein (MBP) is a periplasmic protein that interacts with tar to direct chemotaxis towards maltose. Hellinga and co-workers have used computational protein design to reengineer the ligand-binding pocket of MBP to bind to unnatural chemicals, such as trinitrotoluene (TNT) [2], L-lactate, Zn^{2+} [23], and a nerve agent analog [24•]. When the engineered MBPs are co-expressed with the tar–EnvZ chimera, gene expression can be controlled by these chemicals.

Environment-responsive promoters

Cells change their patterns of gene expression in response to different environmental conditions. There are several conditions, such as pH, temperature, oxygen concentration or UV light, for which bacteria have existing sensing systems. Promoters that turn on under these conditions can be used as sensory inputs. The transcription factors acting on a promoter do not have to be known. For example, promoters identified through microarray experiments, where little else is known about their activation, could be used as inputs. The transfer function of a promoter can be characterized like an inducible system, where gene expression is determined as a function of the input (e.g. temperature or pH).

There are several examples where an environment-inducible promoter has been used as an input for a genetic circuit. Recently, Voigt and co-workers used an anaerobic inducible promoter to create a bacterium that can invade malignant cells in the low-oxygen microenvironment of a tumor [12••]. Promoters involved in the bacterial heat- and cold-shock responses have also been used as temperature sensors [25]. A common problem is that the dynamic range of a promoter does not match the range required to obtain an inducible phenotype. Methods to overcome this problem are described at the end of this review.

RNA aptamers

An aptamer is a small RNA molecule that changes conformation when bound to a protein, peptide or small molecule [26•]. Aptamers can be rationally fused to elements that regulate translation, such as antisense RNAs that inhibit translation by interfering with a ribosome-binding site [27••]. Translation only occurs in the presence of a ligand. Aptamers can also be used to control the activation of other regulatory mRNA motifs, such as ribozymes, which can regulate genes by cutting a target transcript [28••].

RNA regulators are easily engineered as compared with their protein counterparts and can frequently transcend organismal boundaries [29]. In addition, they can potentially regulate any gene through the specification of base-pairing, and their performance characteristics can be easily fine-tuned. This enables rational design to change the form and threshold of the transfer function [27••].

Genetic circuits

Genetic circuits enable cells to process input signals, make logical decisions, implement memory and to communicate with each other (Table 1 and Supplementary material). A variety of synthetic genetic circuits have been constructed to mimic the information-processing capability of electronic circuits, such as inverters, logic gates, pulse generators and oscillators. These circuits can be used to convert a sensor output into a biological response and used to program the cell to perform a series of coordinated tasks.

Switches

Genetic switches are used to turn on gene expression once an input has crossed a threshold required for activation. A switch can also be used as an intermediary to connect the output of a sensor to control a biological response. A switch can be constructed using transcriptional activators or repressors or using post-transcriptional mechanisms, such as DNA-modifying enzymes [30•] or riboregulators [31••] (Supplementary material). Switches are characterized by similar performance parameters as inducible systems: the activation threshold, the cooperativity of the transition, and the cell-to-cell variation.

An inverter is a switch that produces a reciprocal response. When the input to the inverter is on, then the output is off, and vice versa. Weiss and co-workers [4] built an inverter by linking an input promoter to the expression of a repressor, which then turned off a downstream promoter. The transfer function of this inverter can be varied by using directed evolution to modify the genetic control elements or by using different repressors [4] (http://parts.mit.edu; e.g. BBa_Q01121).

Biphasic switches combine positive and negative regulation so that they are only turned on by a small band of input. A promoter can be made biphasic by introducing a binding site where a regulator can behave as an activator and one where it behaves like a repressor. When the regulator has a higher affinity for the first site, then small concentrations induce transcription and larger concentrations repress it. The CI promoter from phage λ naturally contains this type of regulation and has been used in synthetic applications [32,33•].

A toggle switch has been constructed using two repressors that cross-regulate each other's promoter [34]. The system can exist in two states, where one or the other repressor is fully expressed. The switch can be flipped between states by changing the activity of a repressor, either by directly modifying the protein or by altering its expression. For example, the toggle switch has been linked to quorum sensing and a UV-sensitive repressor has been used as an input [35]. A toggle switch can also act like a memory device. Once the switch has been latched into one state, a large perturbation is required to switch it

into the other state. This introduces irreversibility into the switch.

Riboswitches regulate gene expression by blocking translation [36]. The addition of a hairpin to the 5′-end of an mRNA transcript that overlaps the ribosome-binding site will efficiently prevent ribosome binding [31••]. This hairpin can be disrupted by the expression of a small regulatory RNA. This exposes the ribosome-binding site and activates expression. The transfer function of this circuit can be tuned by making nucleotide substitutions to vary the binding free energy competition between the internal hairpin and the small RNA.

Logic gates

Logic gates are the building blocks of digital circuits. They apply a computational operation to convert one or more inputs into a single output. For example, the output of an AND gate is only on when both inputs are on. By contrast, an OR gate is on if either (or both) inputs are on. Many forms of logic have been demonstrated in biological circuits, including NOT, AND, OR and NOR (NOT OR) gates [7,37]. For truly digital logic, the inputs and outputs are either on or off (1 or 0). Biological logic is often fuzzy, where intermediate levels of induction are possible [38].

Two-input logic gates have been built based on the interactions between synthetic rRNA and mRNA. Rackham and Chin [39••] used selection to obtain orthogonal rRNA–mRNA pairs that would only result in protein function when both were expressed. Orthogonal pairs do not cross-react with their endogenous counterparts, which makes them ideal substrates for logic. This was demonstrated by using different topologies of orthogonal pairs to construct OR and AND gates [40].

Dynamic circuits

The circuit functions described in the previous two sections are defined by their steady-steady state input-output response. Circuits can also generate a dynamic response; for example, by functioning as a pulse generator or oscillator. A challenge in the design of dynamic circuits is to make them robust to environmental conditions and to reduce cell-to-cell variation.

Cascades are a common motif in natural regulatory networks. A transcriptional cascade is formed when a chain of transcription factors concurrently regulates each other. Cascades can be used to temporally order the expression of proteins, which is important in metabolic pathways and processes involving self-assembly. Synthetic two- and three-stage cascades have been constructed and analyzed extensively [41••,42••]. In addition to a temporal ordering, each stage of a cascade modifies the transfer function and stochastic characteristics of the overall circuit [43••].

A pulse generator can be formed by an incoherent feedforward motif [6•,44]. A feedforward motif occurs when the input signal is split by turning on an intermediate transcription factor that, together with the input, regulates a downstream process [45]. An incoherent motif occurs when the intermediate regulator and the input have opposite downstream effects [44]; for example, when the input activates a repressor and together they influence the output promoter. This can form a pulse generator when the repressor is turned on slowly and strongly affects the downstream promoter.

A coherent feedforward motif can function as a time delay [46]. This motif occurs when both the input and intermediate regulator have the same effect on a downstream promoter. This delay can act as a filter, where short pulses of inputs do not lead to the activation of the circuit.

Several synthetic oscillators have been constructed by combining a ring of three repressors, a repressor and an activator, and incorporating parts that sense metabolic flux [47,48,49••]. The period of these oscillators varies from 40 min to 10 h. All of the synthetic oscillators are uncoupled and a population rapidly desynchronizes after a few oscillations.

Cell–cell communication

Bacteria use chemical signals to communicate with each other. In their natural setting, these quorum systems are used to detect the cell density and to distinguish between species [50]. Genetic circuits participating in quorum sensing have sender and receiver components. The signal is sent by an enzymatically synthesized small molecule that can freely diffuse through the cell membrane. When the quorum molecule accumulates beyond a threshold, it binds to and activates a regulatory protein (either cytoplasmic or a two-component system). Quorum sensing has been used to program cell–cell communication between bacteria [1,6•,8••,9••,12••].

Different species of bacteria communicate using variations of the quorum molecule. If these systems function independently (i.e. do not cross-react), then they provide multiple channels by which bacterial communication can be programmed [51•,52••]. Interference occurs when the quorum molecule from different species binds non-specifically to the regulator protein. Arnold and co-workers have used directed evolution to change the specificity of the regulator towards different quorum molecules [51•,52••]. To avoid the interference problem entirely, synthetic systems have been developed that use metabolic products as the quorum signal [53].

Actuators: controlling cells

Cellular behavior can be controlled by using the output from a synthetic genetic circuit to drive a natural or transgenic response (Table 1 and Supplementary

material). Implementing synthetic control over a biological process is one of the least explored areas in synthetic biology. Research has to be done to determine how to decouple these systems from their natural inputs and to ultimately control their function with synthetic circuits. Sometimes this is as simple as knocking out a transcription factor from the genome and then placing it under the control of a new circuit. For example, Collins and co-workers linked a UV-responsive toggle switch to a gene that induces biofilm formation (Figure 1) [35••]. In the presence of UV light, this strain will form a visible biofilm. Similarly, Voigt and co-workers used a cell–cell communication circuit to drive the expression of a protein that enables *E. coli* to adhere to and invade cancer cells [12••]. Control over cellular movement has also been implemented by placing a single gene under inducible control [54].

In other instances, linking synthetic sensing and logic to cellular behavior is significantly more challenging. Often the dynamic range of the circuitry does not match the physiological response, either producing a constitutively on or off phenotype. Regulatory feedback can also complicate the control. Finally, processes that rely heavily on core cellular processes, such as central metabolism, often generate unintended consequences when perturbed. Obtaining synthetic control over a complicated, multigene function might require deconstruction of the natural regulation and the use of synthetic regulation to control the entire system. A step towards this goal was recently demonstrated by refactoring and synthesizing a version of T7 bacteriophage, which was engineered to contain simplified regulation [55••].

Debugging and tuning complex systems

Operationally, it is difficult to connect parts to obtain a functioning system. Two parts cannot be connected in series when their timing and dynamic range do not match. To overcome this problem, it is necessary to tune the performance characteristics of one of the parts. Sometimes, it is possible to rationally mutate a part by replacing an operator or ribosome-binding site with an alternative that is expected to fix the problem [6•]. A growing database of parameterized genetic parts (http://parts.mit.edu) is making this approach more accessible. A successful alternative has been to use directed evolution, where random mutagenesis is either applied across an entire part [4] or used to target a specific region [12••,56•].

Ultimately, it might be possible to use computer-aided design tools to accelerate the construction of a system. Before this can happen, it is necessary to first obtain a sufficiently large toolbox of standardized and parameterized parts and then to develop simplified theoretical techniques to understand how these parts will function together. This theory may require a combination of detailed biochemical data to characterize the inner workings of a part with higher level, empirical relationships that can be used to engineer the linkage between parts. Considering promoter regulation, statistical mechanics can be used to link the transfer function to the thermodynamics of transcription factor binding [57•]. Computational methods can also be used to predict which regulatory elements are the most fruitful targets for combinatorial mutagenesis [56•].

Conclusions

DNA synthesis and cloning technology have far outstripped our design capacity. Chemical synthesis can now routinely build 50 kilobases of contiguous DNA at a reasonable cost [58]. Improved transformation techniques enable us to put this much synthetic DNA into a cell [59••]. Sequencing the whole genome of a microbe is becoming faster and cheaper [60]. New microfluidics technologies are being developed that could facilitate the rapid determination of transfer functions at the level of individual cells [61••]. Despite all of these advances, we are not at the point where we can design an integrated, working system on the 50 kilobase scale. Increasing the complexity of the designs will require improvements in four areas: the construction of new robust parts that can be easily interchanged; increased understanding of how to routinely wire parts in series; the development of new theory and computer design methods; and standardized data sharing between laboratories.

The focus of this review has been on the individual genetic parts that can be combined to create more complex systems. These parts can be used to encode when, where, how much, and under what conditions genes are expressed. They can be used to drive the expression of transgenic genes, to manipulate metabolism or to drive natural processes in the cell. Cells can be programmed to perform a series of tasks. For example, genetic circuits can be used to control the expression of a series of metabolic proteins at the precise time and amount required for the maximal synthesis of a drug or energetic compound, while diverting the flux from competing pathways [10,11,18]. Bacteria could weave complex materials by temporally and spatially controlling the expression of biopolymer and modifying proteins. Cells could be used as therapeutics, programmed to find and fix medical problems in the body [12••,62••,63]. Entire multicomponent machines and organelles (e.g. photosystems, secretions systems, pili and metabolic pathways) could be transferred between species and placed under complete synthetic control. These projects represent a revolution in genetic engineering, where cells are programmed to undertake large and complex tasks.

Acknowledgements

The author thanks Elizabeth Clarke, Jeff Tabor, J Christopher Anderson, and Drew Endy for critical comments. CAV is supported by NIH PN2-EY016546, NSF BES-0547647, the Sloan Foundation, the Pew Scholars Program, and the Sandler Family.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.copbio.2006.09.001.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Endy D: **Foundations for engineering biology**. *Nature* 2005, **438**:449-453.

2. Looger LL, Dwyer MA, Smith JJ, Hellinga HW: **Computational design of receptor and sensor proteins with novel functions**. *Nature* 2003, **423**:185-190.

3. •• Levskaya A, Chevalier AA, Tabor JJ, Simpson ZB, Lavery LA, Levy M, Davidson EA, Souras A, Ellington AD, Marcotte EM, Voigt CA: **Engineering *E. coli* to see light**. *Nature* 2005, **438**:441-442.

The authors recombine a cyanobacterial phytochrome with an *E. coli* signal transduction system to create a strain of bacteria that can record a two-dimensional image of projected light.

4. Yokobayashi Y, Weiss R, Arnold FH: **Directed evolution of a genetic circuit**. *Proc Natl Acad Sci USA* 2002, **99**:16587-16591.

5. Elowitz MB, Leibler S: **A synthetic oscillatory network of transcriptional regulators**. *Nature* 2000, **403**:335-338.

6. • Basu S, Mehreja R, Thiberge S, Chen M-T, Weiss R: **Spatiotemporal control of gene expression with pulse-generating networks**. *Proc Natl Acad Sci USA* 2004, **101**:6355-6360.

A pulse generator is developed that uses a quorum signal as an input. See also [8••].

7. Guet CC, Elowitz MB, Hsing W, Leibler S: **Combinatorial synthesis of genetic networks**. *Science* 2002, **296**:1466-1470.

8. •• Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R: **A synthetic multicellular system for programmed pattern formation**. *Nature* 2005, **434**:1130-1134.

The *Vibrio fischeri luxIR* circuit is cut into pieces and used to create 'sender' and 'receiver' cells that communicate to create beautiful two-dimensional patterns.

9. •• You L, Cox RS, Weiss R, Arnold FH: **Programmed population control by cell-cell communication and regulated killing**. *Nature* 2004, **428**:868-871.

The *luxIR* quorum circuit is connected to the *ccdB* cell suicide output. This enabled the programming of cell death after the population surpasses a critical density.

10. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD: **Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids**. *Nat Biotechnol* 2003, **27**:796-802.

11. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C: **Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli***. *Science* 2001, **291**:1790-1792.

12. •• Anderson JC, Clarke EJ, Arkin AP, Voigt CA: **Environmentally controlled invasion of cancer cells by engineered bacteria**. *J Mol Biol* 2006, **355**:619-627.

Sensors that respond to the environmental cues in a tumor are linked to the expression of a protein that directs the bacteria to adhere to and invade malignant cells.

13. Weiss R, Homsy GE, Knight TR: **Towards *in vivo* digital circuits**. *DIMACS Workshop on Evolution as Computation* 1999; **1**: 1–18.

14. Chin JW: **Modular approaches to expanding the functions of living matter**. *Nat Chem Biol* 2006, **2**:304-311.

15. Hoch JA, Silhavy TJ: *Two-Component Signal Transduction*. Washington: ASM Press; 1995.

16. Utsumi R, Brissette RE, Rampersaud A, Forst SA, Oosawa K, Inouye M: **Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate**. *Science* 1989, **245**:1246-1249.

17. Kwon O, Georgellis D, Lin ECC: **Rotational on-off switching of a hybrid membrane sensor kinase Tar-ArcB in *Escherichia coli***. *J Biol Chem* 2003, **278**:13192-13195.

18. Farmer WR, Liao JC: **Improving lycopene production in *Escherichia coli* by engineering metabolic control**. *Nat Biotechnol* 2000, **18**:533-537.

19. Farmer WR, Liao JC: **Acetate-inducible protein overexpression from *glnA*p2 promoter of *Escherichia coli***. *Biotechnol Bioeng* 2001, **75**:504-509.

20. Michalodimitrakis KM, Sourjik V, Serrano L: **Plasticity in amino acid sensing of the chimeric receptor Taz**. *Mol Microbiol* 2005, **58**:257-266.

21. Ward SM, Delgado A, Gunsalus RP, Manson MD: **A NarX-Tar chimera mediates repellent chemotaxis to nitrate and nitrite**. *Mol Microbiol* 2002, **44**:709-719.

22. Derr P, Boder E, Goulian M: **Changing the specificity of a bacterial chemoreceptor**. *J Mol Biol* 2006, **355**:923-932.

23. Dwyer MA, Looger LL, Hellinga HW: **Computational design of a Zn^{2+} receptor that controls bacterial gene expression**. *Proc Natl Acad Sci USA* 2003, **100**:11255-11260.

24. • Allert M, Rizk SS, Looger LL, Hellinga HW: **Computational design of receptors for an organophosphate surrogate of the nerve agent soman**. *Proc Natl Acad Sci USA* 2004, **101**:7907-7912.

The ligand pocket of maltose-binding protein is redesigned to bind to a hydrolytic product of the nerve agent sarin.

25. Mujacic M, Cooper KW, Baneyx F: **Cold-inducible cloning vectors for low-temperature protein expression in *Escherichia coli*: application to the production of a toxic and proteolytically sensitive fusion protein**. *Gene* 1999, **238**:325-332.

26. • Lee JF, Hesselberth JR, Meyers LA, Ellington AD: **Aptamer database**. *Nucleic Acids Res* 2004, **32**:D95-D100.

Database of all engineered aptamers: http://aptamer.icmb.utexas.edu/.

27. •• Bayer TS, Smolke CD: **Programmable ligand-controlled riboregulators of eukaryotic gene expression**. *Nat Biotechnol* 2005, **23**:337-343.

Riboregulators are constructed that turn expression on or off in the presence of a small molecule (theophylline). The system is demonstrated in yeast, but should also work in prokaryotes.

28. •• Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR: **Control of gene expression by a natural metabolite-responsive ribozyme**. *Nature* 2004, **428**:281-286.

A ribozyme is reported that is activated by the metabolite glucosamine-6-phosphate. Glucosamine-6-phosphate is produced by the GlmS enzyme, the mRNA of which is cleaved by the ribozyme. This arrangement forms a post-transcriptional negative feedback loop.

29. Werstuck G, Green MR: **Controlling gene expression in living cells through small molecule-RNA interactions**. *Science* 1998, **282**:296-298.

30. • Ham TS, Lee SK, Keasling JD, Arkin AP: **A tightly regulated inducible expression system utilizing the *fim* inversion recombination switch**. *Biotechnol Bioeng* 2006, **94**:1-4.

An input induces an invertase, which flips DNA to change the orientation of the promoter. This produces an irreversible switch with minimal leakiness.

31. •• Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JR: **Engineered riboregulators enable post-transcriptional control of gene expression**. *Nat Biotechnol* 2004, **22**:841-847.

Hairpins are designed to sequester ribosome-binding site and, thus, to inhibit translation. This inhibition is overcome by the transcription of a small regulatory RNA from a second promoter.

32. Isaacs FJ, Hasty J, Cantor CR, Collins JR: **Prediction and measurement of an autoregulatory genetic module**. *Proc Natl Acad Sci USA* 2003, **100**:7714-7719.

33. • Michalowski CB, Short MD, Little JW: **Sequence tolerance of the phage λ PRM promoter: implications for evolution of gene regulatory circuitry**. *J Bacteriol* 2004, **186**:7899-7999.

The authors mutate the circuit controlling the lysis–lysogeny cycle of bacteriophage λ. Variants of the biphasic P_{RM} promoter are found that have altered transfer functions.

34. Gardner TS, Cantor CR, Collins JR: **Construction of a genetic toggle switch in *Escherichia coli***. *Nature* 2000, **403**:339-342.

35. •• Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, Cantor CR, Collins JJ: **Programmable cells: interfacing natural and engineered gene networks**. *Proc Natl Acad Sci USA* 2004, **101**:8414-8419.
An excellent demonstration of the modularity and extensibility of the genetic toggle switch. The authors link it to several different inputs (quorum sensing, SOS response) and use it to control the formation of a biofilm.

36. Mandal M, Breaker RR: **Gene regulation by riboswitches**. *Nat Rev Mol Cell Biol* 2004, **5**:451-463.

37. Setty Y, Mayo AE, Surette MG, Alon U: **Detailed map of cis-regulatory input function**. *Proc Natl Acad Sci USA* 2003, **100**:7702-7707.

38. Arkin A, Ross J: **Computational functions in biochemical reaction networks**. *Biophys J* 1994, **67**:560-578.

39. •• Rackham O, Chin JW: **A network of orthogonal ribosome*rna pairs**. *Nat Chem Biol* 2005, **1**:159-166.
Selections are performed to create synthetic mRNA and rRNA interaction pairs in *E. coli*. The pairs do not cross-react and can be used simultaneously as control circuits.

40. Rackham O, Chin JW: **Cellular logic with orthogonal ribosomes**. *JACS* 2005, **127**:17584-17585.

41. •• Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB: **Gene regulation at the single-cell level**. *Science* 2005, **307**:1962-1965.
The biochemical parameters underlying a two-stage repressor cascade are inferred *in vivo*. Transfer functions describing this network are obtained in individual cells.

42. •• Hooshangi S, Thiberge S, Weiss R: **Ultrasensitivity and noise propagation in a synthetic transcriptional cascade**. *Proc Natl Acad Sci USA* 2005, **10**:3581-3586.
A three-stage transcriptional cascade is constructed. The transfer function and cell-to-cell variation is measured at each stage.

43. •• Pedraza JM, van Oudenaarden A: **Noise propagation in gene networks**. *Science* 2005, **307**:1965-1969.
The authors study how a noisy upstream part of a gene network will influence the downstream parts to which it is connected.

44. Mangan S, Alon U: **Structure and function of the feed-forward loop network motif**. *Proc Natl Acad Sci USA* 2003, **100**:11980-11985.

45. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of *Escherichia coli***. *Nat Genet* 2002, **31**:64-68.

46. Mangan S, Zaslaver A, Alon U: **The coherent feedforward loop serves as a sign-sensitive delay element in transcription network**. *J Mol Biol* 2003, **334**:197-204.

47. Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic gene expression in a single cell**. *Science* 2002, **297**:1183-1186.

48. Atkinson MR, Savigeau MA, Myers JT, Ninfa AJ: **Development of a genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli***. *Cell* 2003, **113**:597-607.

49. •• Fung E, Wong WW, Suen JK, Bulter T, Lee SG, Liao JC: **A synthetic gene-metabolic oscillator**. *Nature* 2005, **435**:118-122.
An oscillating genetic circuit is constructed from transcriptional and metabolic components. Some of the parts are from the glucose sensor (see Supplementary material).

50. Camilli A, and Bassler, BL: **Bacterial small-molecule signaling pathways**. *Science* **311**: 1113–1116.

51. • Collins CH, Arnold FH, Leadbetter JR: **Directed evolution of *Vibrio fischeri* LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones**. *Mol Microbiol* 2005, **55**:712-723.
The authors use random mutagenesis and DNA shuffling techniques to identify a LuxR variant that has increased specificity towards a non-natural acyl-homoserine lactone.

52. •• Collins CH, Leadbetter JR, Arnold FH: **Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR**. *Nat Biotechnol* 2006, **24**:708-712.
The researchers use directed evolution to alter the specificity of a quorum sensor to respond to different small molecules. This opens up multiple channels by which cells can be programmed to communicate.

53. Bulter T, Lee S-G, Wong WW, Fung E, Connor MR, Liao JC: **Design of artificial cell-cell communication using gene and metabolic networks**. *Proc Natl Acad Sci USA* 2004, **101**:2299-2304.

54. Blair DF, Berg HC: **Restoration of torque in defective flagellar motors**. *Science* 1988, **242**:1678-1681.

55. •• Chan LY, Kosuri S, Endy D: **Refactoring bacteriophage T7**. *Mol Systems Biol* 2005: doi: 10.1038/msb4100025.
The authors rationally reduce the regulatory complexity of the T7 genome and build it using DNA synthesis. The refactored genome yields a viable T7 phage.

56. • Feng X, Hooshangi S, Chen D, Li G, Weiss R, Rabitz H: **Optimizing genetic circuits by global sensitivity analysis**. *Biophys J* 2004, **87**:2195-2202.
An algorithm is developed to determine the parameters in a circuit model that should be targeted for mutagenesis to achieve a desired property.

57. • Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Phillips R: **Transcriptional regulation by the numbers: models**. *Curr Opin Genet Dev* 2005, **15**:116-124.
The physiochemical basis for the transfer function of a promoter is described.

58. Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, Church G: **Accurate multiplex gene synthesis from programmable DNA microchips**. *Nature* 2004, **432**:1050-1054.

59. •• 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* 2005, **102**:15971-15976.
The entire 3.5-megabase genome of a cyanobacterium is cloned into the *B. subtilis* genome.

60. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM: **Accurate multiplex polony sequencing of an evolved bacterial genome**. *Science* 2005, **309**:1728-1732.

61. •• Balagadde FK, You L, Hansen CL, Arnold FH, Quake SR: **Long-term monitoring of bacteria undergoing programmed population control in a microchemostat**. *Science* 2005, **309**:137-140.
A microchemostat with a 16 nL growth chamber is built and used to study a population control genetic circuit in *E. coli*.

62. •• Dane KY, Chan LA, Rice JJ, Daugherty PS: **Isolation of cell specific peptide ligands using fluorescent bacteria display libraries**. *J Immunol Methods* 2006, **309**:120-129.
A outer membrane display system is developed and used to select for peptides that bind specifically to breast cancer cells. Bacteria displaying these peptides adhere tightly to their target.

63. Yu YA, Shabahang S, Timiryasova TM, Zhang Q, Beltz R, Gentschev I, Goebel R, Szalay AA: **Visualization of tumors and metastases in live animals with bacteria and vaccinia virus encoding light-emitting proteins**. *Nat Biotechnol* 2004, **22**:313-320.

64. Lutz R, Bujard H: **Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements**. *Nucleic Acids Res* 1997, **25**:1203-1210.

65. Siegele DA, Hu JC: **Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations**. *Proc Natl Acad Sci USA* 1997, **94**:8168-8172.

66. Lutz R, Lozinski T, Ellinger T, Bujard H: **Dissecting the functional program of *Escherichia coli* promoters: the combined mode of action of Lac repressor and AraC activator**. *Nucleic Acids Res* 2001, **29**:3873-3881.

67. Little JW, Shepley DP, Wert DW: **Robustness of a gene regulatory circuit**. *EMBO J* 1999, **15**:4299-4307.

68. • Karig D, Weiss R: **Signal-amplifying genetic circuit enables *in vivo* observation of weak promoter activation in the Rhl quorum sensing system**. *Biotechnol Bioeng* 2004, **89**:709-718.
The *Pseudomonas* Rhl quorum sensing system produces a signal that is too weak to be detected. The authors use an inverter to amplify the signal.

69. Egland KA, Greenberg EP: **Conversion of the *Vibrio fischeri* transcriptional activator, LuxR, to a repressor**. *J Bacteriol* 2000, **182**:805-811.

70. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A: **Regulation of noise in the expression of a single gene**. *Nat Genet* 2002, **31**:13-14.

71. • Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, van Oudenaarden A: **Predicting stochastic gene expression dynamics in single cells**. *Proc Natl Acad Sci USA* 2006, **103**:7304-7309.
The authors use fluorescence data to estimate several simplified, but predictive parameters describing the stochastic dynamics of the genetic circuit controlling the lactose operon.