Current Biology 20, R772–R779, September 14, 2010 ©2010 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2010.06.049

Using Synthetic Biology to Understand the Evolution of Gene Expression

Review

Travis S. Bayer[1,2]

The evolution of phenotype is often based on changes in gene expression rather than changes in protein-coding sequence. Gene expression is controlled by complex networks of interacting regulators that act through a variety of biochemical mechanisms. Perturbation of these networks can have profound effects on the fitness of organisms. This highlights an important challenge: the investigation of whether the mechanisms and network architectures we observe in Nature evolved in response to selective pressure — and, if so, what that pressure might have been — or whether the architectures are a result of non-adaptive forces. Synthetic biologists aim to construct artificial genetic and biological systems to increase our understanding of Nature as well as for a number of biotechnological applications. In this review, I will highlight how engineering 'synthetic' control of gene expression provides a way to test evolutionary hypotheses. Synthetic biology might allow us to investigate experimentally the evolutionary paths not taken by extant organisms.

Introduction

It has long been hypothesized that the evolution of phenotype is often based on changes in gene expression rather than changes in encoded proteins [1,2]. Many animal proteins show remarkable sequence conservation over one billion years of divergent evolution and are able to substitute functionally for each other *in vivo*. Some of the first examples of this conservation include the ability of the mouse *Hox* genes to specify segment identity in *Drosophila* [3,4] and the ability of mouse *Pax-6* to trigger eye development in flies [5]. Given this striking conservation of protein sequence and function, a large number of studies have correlated the differences in when, where, and at what level genes are expressed with morphological and functional differences between organisms. For example, experimental analysis of the *HoxD* complex has shown that novel regulatory control emerged in the evolution of tetrapod limbs from paired fins of fish [6]. Changes in the regulation and expression of conserved genes are also involved in the evolution of complex plant leaves [7], avian feathers [8], butterfly spots [9,10], and tissue-fate specification in echinoderms [11].

The study of the evolution of gene expression and genetic control networks is confounded by the fact that we are often only able to study the network in a single model organism. For example, it is difficult to ascertain whether the hierarchical organization of transcriptional control [12] (where a handful of master regulators control a large number of genes) is due to a functional advantage or is the result of non-adaptive processes, such as genome duplication. One could investigate the functional value of hierarchical control by comparison to closely related networks that have an alternative organization of genetic control. Synthetic biology, which entails the design, construction, and re-engineering of living systems [13,14], can provide a way to do precisely this. In this review, I will highlight cases where synthetic biology has been used to rewire regulatory networks to probe the adaptive and functional value of the biological systems we observe in Nature.

[1]Centre for Synthetic Biology and Innovation and [2]Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, UK.
E-mail: t.bayer@imperial.ac.uk

Synthetic Biology

Synthetic biology is the design and construction of biological systems guided by engineering principles, with the aim of understanding biology or producing useful biological technologies. While a precise definition and scope has been discussed elsewhere [14,15], in broad terms synthetic biology encompasses a wide range of focus areas, including alternative chemistries, artificial cells, self-replicating macromolecules, and *in silico* life forms. The construction of genetic circuits is one of the defining research areas of synthetic biology and is (currently) the most relevant focus area for understanding the evolution of gene expression.

Genetic Circuits

The ability to engineer genetic regulatory elements to function analogously to electronic circuits was first described in several seminal papers a decade ago that described a genetic toggle switch (a synthetic, bistable gene-regulatory network) [16] and a genetic oscillator [17] constructed from well-characterized promoters and transcription factors (such as the *tet* repressor (TetR), the *lac* repressor (LacI), and lambda phage regulators). Subsequent papers have described circuits that are able to count cellular events [18], form patterns [19], control intracellular noise [20,21], have memory [22], and replicate predator–prey behavior [23], amongst other functions. The construction and analysis of genetic circuits has been instrumental in demonstrating the role of architecture in determining network stability [24]. What distinguishes synthetic biology from earlier genetic engineering is the (somewhat obvious) realization that the use of well-characterized genetic components and predictive mathematical models of genetic circuit behavior is a powerful approach for designing complex biological functions. In addition, a push towards community-wide standards of construction, design, and measurement aims to move synthetic biology from tinkering with biological systems to a true engineering discipline [25]. As the capacity to synthesize long sections of DNA increases [26–28] and the price of synthesis decreases, the limiting factor will be the ability to design and validate complex biological circuits rather than the ability to assemble them. The use of well-characterized parts and models are complemented by approaches that take advantage of diversity on a gene-circuit and genome scale.

Synthetic Chemistries and Other Approaches

A complementary focus of synthetic biology has been the study of how biological processes and functions (particularly Darwinian evolution) could arise from purely physical and

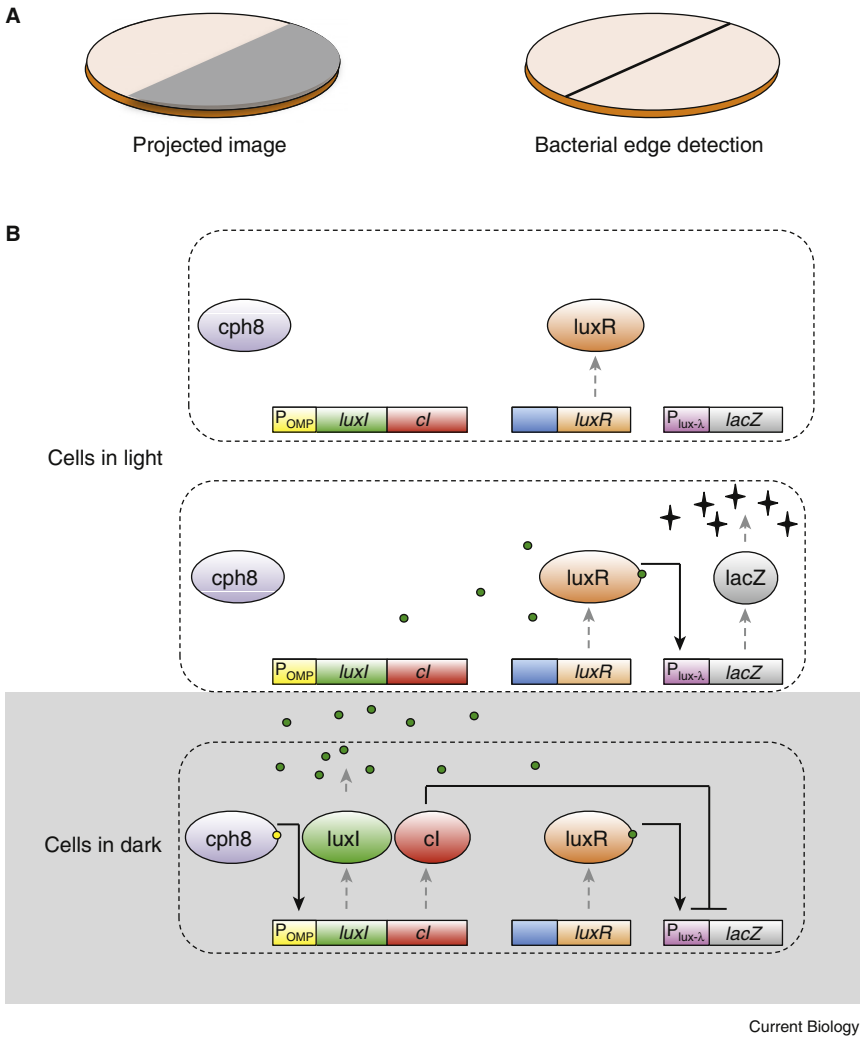

Figure 1. Construction of a complex genetic circuit to detect edges in a projected image.

A simplified schematic of the bacterial edge detection circuit described in the main text. (A) The intended functionality of the edge detection circuit. An image is projected onto a lawn of cells. After the biological 'computation' is performed, the bacterial population delineates the edge between light and dark regions on the plate by producing a black pigment. (B) The genetic circuit encoding edge detection. Cells in the dark (bottom) induce expression of luxI and cI due to the engineered cph8 protein, which is phosphorylated (yellow circle) and activates gene expression from the OMP promoter in the dark. The diffusible acylhomoserine lactone (AHL, green circles), generated by the AHL synthase luxI, binds to luxR and activates expression from the lux-λ promoter driving *lacZ* expression; however, *lacZ* is dominantly repressed by cI. Cells at the edge of the image (in the light but in close proximity to the dark regions) receive the AHL signal and induce *lacZ* expression via the effects of luxR, producing black pigment (represented by black crosses) as an output. cI is not induced because cph8 is inactive in the presence of light. Cells in the light but far away from the light/dark boundary do not receive the AHL signal or induce luxI/cI expression, resulting in no expression of lacZ.

chemical starting points. There have been significant efforts to construct a protocell that compartmentalizes a self-replicating molecule or metabolic process. Hanczyc *et al.* [29] demonstrated that mineral particles were able to nucleate the formation of lipid vesicles from a solution of micelles and that these vesicles could divide and absorb RNA. Further work showed that protocells were able to take up synthetic DNA that had been designed to provide a template for the extension of a complementary strand when activated nucleotides were supplied exogenously [30].

Several groups have explored how the synthesis of unnatural nucleobases can be used to expand the genetic system to six (rather than four) nucleotides [31,32]. Yang *et al.* [33] have shown that this third basepair can be enzymatically incorporated into DNA strands, which offers the possibility of a novel form of information storage in biology. DNA can also be used as a material for the assembly of complex nanoscale structures (often termed 'DNA origami'), where groups have described the construction of two-dimensional crystals and lattices [34], tubes [35], three-dimensional scaffolds [36], and layers of helices that can be designed to assemble into virtually any three-dimensional shape [37]. The design of nanostructures that could self-assemble and template the assembly of more nanostructures would allow Darwinian evolution to proceed and could be used as a tool to evolve new functions and structures.

Designing Biological Systems: Rational and Evolutionary Approaches

A recent example of the use of well-characterized genetic components and mathematical modeling to engineer complex behavior in cells is the construction of bacteria to detect edges in a projected image. To accomplish this, Tabor *et al.* [38] built upon previous work in which bacteria were engineered to sense light [39] and respond with a pigment output following the expression in *Escherichia coli* of a chimeric two-component phytochrome system from a cyanobacterium. When coupled to the expression of a chemical output, this function allowed a lawn of bacteria to act as a photographic film: projection of an image onto the lawn results in the recording of a high-definition, two-dimensional chemical image. To enable bacterial edge detection, the authors connected additional genetic circuitry to the light-sensing component (Figure 1), each part of which was constructed independently, characterized, and modeled to predict the behavior of the full system. Tabor *et al.* [38] engineered a cell–cell communication component and a simple genetic logic such that cells in the dark areas of the image produce a diffusible chemical signal. A genetic logic gate acts analogously to an electronic logic gate, by integrating two signals and performing simple logic such as AND or OR. By implementation of genetic logic, only cells that sense the light areas of the image and the diffusible signal are able to generate a pigment as a read-out of the image edge. These efforts demonstrate how design and construction of complex biological

behavior proceeds from the drawing board to a functional piece of DNA in living cells.

Construction of a gene network from genetic components to behave in a predictable way remains a task that requires iterative rounds of *post hoc* tweaking, redesign, and testing. Often, this tinkering is constrained by the limited number of available biological parts, such as promoters, or by incomplete knowledge or prediction of circuit behavior. Several groups have used directed evolution as a tool to achieve desired circuit performance. Yokobayashi *et al.* [40] combined rational engineering with directed evolution in which a non-functional genetic logic gate evolved into a functioning genetic device. Recent work by Ellis *et al.* [41] showed that the synthesis and screening of libraries of standard parts can be a powerful approach to design gene networks. In this work the authors randomized regions of the *lac* and *tet* promoters, which resulted in promoters with different levels of transcription. They were then able to use computational models to select which promoter variants to use in the design of their synthetic networks. The *lac* and *tet* promoters were used to create a toggle switch that acts as a biological timer for controlling the onset of yeast sedimentation. The timing of the sedimentation event could be predictably tuned by changing the relative strength of the promoters. This work demonstrates that biological engineers can take advantage of diversity in designing systems rather than attempting to limit diversity. The two necessary ingredients for Darwinian change — genetic variation and selection (in this case provided by the model) — provide unique opportunities for engineering biology.

The generation and screening of diversity to acheive desired function has been used to optimize rationally designed genetic circuits and can also be applied at the genome scale. Wang *et al.* [42] recently described a method for introducing DNA mismatches, insertions, and deletions across the *E. coli* chromosome in a single cell or in a population of cells. The method relies on homology between transformed oligonucleotides and chromosomal DNA. Wang *et al.* [42] were able to automate this process and termed it MAGE (multiplex automated genome engineering). The MAGE method was used to 'evolve' overproduction of lycopene in an engineered strain of *E. coli*. In this selection, over 15 billion mutants were generated in 25 cycles of mutagenesis. Clones were isolated that showed fivefold increases in lycopene production over the ancestral strain. The generation of diversity and selection for desired function is an important (and somewhat unique) tool available to the synthetic biologist to engineer genetic regulatory circuits and networks.

Testing Hypotheses in the Evolution of Gene Expression

One of the major future contributions of synthetic biology to the study of evolution may be to discriminate whether regulatory network features arose through natural selection or through other, non-adaptive processes. For example, theoretical studies of genome duplication have found that network motifs (such as feed-forward loops) arise frequently, even in the absence of selection [43,44]. In addition, Lynch [45] has discussed how genetic drift, mutation, and recombination are sufficient to explain transcription control network topology, especially in populations of limited size.

These investigations suggest that the features of networks we observe may be what Gould and Lewontin termed 'spandrels' of genome evolution [46]. In architecture, spandrels correspond to the space between a curved arch and rectangular enclosure, and these regions are usually heavily decorated in cathedrals. Although spandrels were later co-opted for aesthetic purposes, they are the by-products of the building rules and, as such, were not intentionally designed. Some features of regulatory networks, such as the abundance of small motifs, modular organization, and complexity, may well be by-products of the mechanistic processes of genome evolution. However, the emergence of features by non-adaptive mechanisms does not preclude their later adaptive value to an organism, particularly if environmental conditions change. The construction and analysis of synthetic regulatory networks could be a powerful way to test hypotheses surrounding the emergence and adaptive value of genetic regulatory networks.

Optimality in Biology

One of the fundamental predictions of evolutionary theory is that protein expression levels are optimized during evolution to maximize fitness in a given environment. However, fitness as a function of expression level is rarely measured. Synthetic biology allows researchers to design simple genetic circuits where the expression level of a protein (or proteins) of interest can be precisely tuned to measure organism fitness across a range of expression levels. In an example of this, Dekel and Alon [47] characterized the costs and benefits of the expression of the *lac* operon in *E. coli*. The *lac* operon encodes LacZ, which cleaves lactose as the first step in its catabolism, and LacY, which imports lactose into the cell. The benefit of LacZ and LacY expression is the energy and carbon provided by lactose metabolism, while the cost is the metabolic burden of LacZ and LacY protein expression. Costs of *lac* expression were measured using the non-metabolizable inducer of *lac*, IPTG, and assessing the decrease in growth rate. The authors used these data to parameterize a fitness function for cells as a function of lactose. As would be expected, for a given lactose concentration in the environment there is an optimal expression level of the *lac* operon genes. Dekel and Alon then performed 'laboratory evolution' in different lactose environments and found that *E. coli* evolved *lac* operon expression to the predicted level. This work is a simple yet powerful demonstration that cells evolve expression levels to an optimal solution of a cost–benefit function. Construction of precise genetic control of expression of a metabolic pathway could be used to ask whether cost–benefit analysis [48] applies to more complex networks in the cell.

Optimal function is a shared idea between engineering and evolutionary biology. One open question is towards which functions evolution has optimized metabolic networks. Organisms face multiple challenges from their environments, including fluctuating nutrient sources, a variety of stress conditions, and potentially deleterious (or beneficial) interactions with other organisms. The Palsson group [49] has shown that in some cases the 'objective' of genetic regulation of metabolism is to maximize growth and, in this way, gene expression levels and metabolic flux through a given pathway can be predicted. In an elegant demonstration of this objective function, Palsson and colleagues [49] grew *E. coli* on glycerol (a non-preferred carbon source for this organism) and observed that gene expression evolved to maximize its growth rate according to the predictions described by an *in silico* model. The utility of this simple evolutionary principle has been further validated by its use to predict essential genes [50].

Alternative selective pressures may, however, result in regulatory and metabolic networks that are optimized for other functions besides growth. An analysis of gene deletions in *Bacillus subtilis* revealed that several mutants grew faster than wild type, showing that gene expresion levels have not evolved solely to maximize biomass production [51]. The authors found that *B. subtilis* invests a significant amount of resources in anticipation of changing environmental conditions, which the authors term a 'stand-by mode'. For example, most identified deletion mutants displaying increased biomass production were regulators of alternative phenotypic states of *B. subtilis*, such as sporulation and competence. These developmental pathways are activated in starvation conditions and are repressed in nutrient-rich environments. Thus, gene expression is a compromise between rapid growth in resource-abundant environments and the anticipation of environmental change. These results demonstrate how multiple selective pressures can shape metabolic networks and suggest that human engineers will similarly have to balance multiple functions in designing useful biological systems.

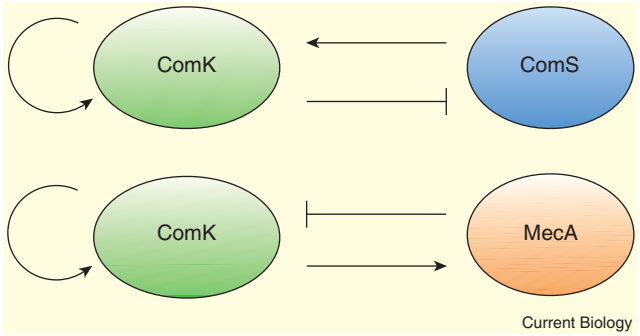

Figure 2. Rewiring the *B. subtilis* competence circuit.

(Top) The natural ComK genetic circuit. ComK induces is own expression and inhibits the expression of ComS; in turn, ComS is able to induce ComK expression. (Bottom) The synthetic ComK circuit, in which ComK, as well as inducing its own expression, induces the expression of MecA, an inhibitor of ComK expression. In the synthetic circuit, ComK induces the expression of its inhibitor, whereas in the natural circuit ComK represses the expression of its inducer.

Synthetic Circuits to Study Functional and Fitness Contributions

The above examples highlight an important challenge in evolutionary biology: explaining why the genetic network architectures we observe in Nature have evolved to solve a particular problem an organism faces in its environment. This challenge is often complemented by the question of which selective forces (i.e. environmental or cellular conditions) have shaped the biological systems we observe in modern organisms. The null hypothesis is simply that a particular architecture has arisen by non-selective forces and that multiple architectures would be sufficient to achieve the biological functionality observed. For example, the circuit that regulates mating in ascomycete yeasts uses an activator in *Candida albicans* and a repressor in *Saccharomyces cerevisiae*, but appears to encode identical regulatory logic [52]. As stated above, asking the 'why' question in biology is often confounded by the availability of only one system in a limited number of model organisms.

The construction of synthetic versions of natural circuits is a powerful way to interrogate questions of 'why' in biology. Along these lines, Cagatay *et al.* [53] studied the circuit that regulates competence in *B. subtilis*. Competence refers to the developmental state of *B. subtilis* in which a fraction of cells in a population will take up and incorporate extracellular DNA under periods of environmental stress. In wild-type *B. subtilis*, competence is controlled by a master regulator, ComK, which induces its own expression to form a positive feedback loop. ComK represses ComS expression, and ComS in turn induces ComK. This has the effect of forming a net negative-feedback loop, which generates a transient pulse of competence upon environmental induction. In the synthetic version of this circuit, ComK induces the expression of the regulator MecA, which in turn represses ComK. This circuit topology again forms a net negative-feedback loop but differs in the specific topology: in the native circuit, ComK represses its inducer, whereas in the synthetic circuit ComK induces its repressor (Figure 2). The authors found that the synthetic circuit was able to generate wild-type competence dynamics and mimic wild-type competence frequency. However, analysis of competence at the single-cell level was able to reveal the differences between the wild-type and synthetic circuits and shed light on why the particular wild-type circuit topology has evolved. Cagatay *et al.* [53] found that there was significantly more variation in the duration of the competence pulses between cells expressing the wild-type circuit. The 'noisy' operation of the wild-type circuit may be beneficial in unpredictable environments.

Synthetic Rewiring Studies

Synthetic approaches have also been used to study how the fitness and phenotype of the cell respond to 'rewiring' of the transcriptional regulatory network (Figure 3). Isalan *et al.* [54] added new 'links' in the regulatory network of *E. coli* by constructing synthetic constructs that consisted of a regulatory region with a different transcription factor or σ-factor protein-coding region. When this synthetic construct was introduced into cells, the swapped promoter controlled expression of the transcription factor or σ-factor, which in turn regulated the expression of genes downstream in the network and created new network motifs: for example, if regulatory gene A induced expression of gene B, then a synthetic construct in which the protein-coding region of A is controlled by B had the effect of introducing a positive-feedback loop. Isalan *et al.* [54] constructed and screened approximately 600 such rewirings and found that 95 percent resulted in viable cells. This included rewiring the expression of global regulators of gene expression such as σ^{70}, which controls the expression of approximately 1,000 genes. The authors also applied selection pressure to their set of synthetic construct strains to assess whether rewiring networks could provide a route to evolvability. After serial passaging, they found that specific rewirings were fitter than the wild-type strain with respect to growth in liquid media, longevity in stationary phase, and survival after heat shock. For growth in liquid culture, the fittest clones were those that rewired the control of flagellar biosynthesis genes. These genes are non-essential and are costly for the cell in terms of metabolic resources. In the longevity and heat-shock challenges the surviving clones all harbored rpoS–ompR constructs, where the promoter for rpoS (a σ-factor) drives the expression of ompR (a DNA-binding response regulator of two-component systems). These results

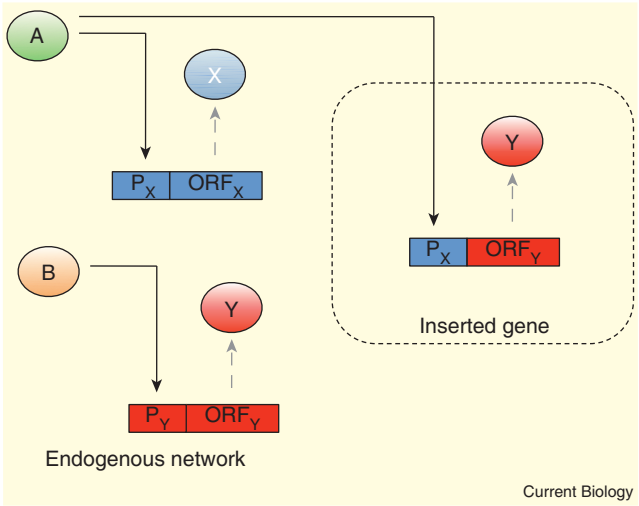

Figure 3. Rewiring transcriptional control in *E. coli*.

The approach taken to rewire the transcriptional control network of *E. coli*. In the endogenous network, regulator A induces the expression of gene X and regulator B induces the expression of gene Y. By adding a construct that fuses the promoter of X and the open reading frame (ORF) of Y, a new regulatory link is added to the cell, whereby regulator A can now induce the expression of gene Y.

suggest two important evolutionary points: firstly, organisms can evolve by small changes in their regulatory network; and secondly, variation in the network is largely tolerated. The accumulation of variation in a population provides a basis on which natural selection can act at a later time or in novel environments.

Re-engineering approaches can also be used to parse the potential functional advantages of the observed complexity of many genetic circuits found in Nature. Recent work by Smith and Davidson [55] has shown that a gene regulatory network that specifies skeletogenic mesenchyme fate in sea urchins operates in parallel with a back-up circuit to ensure correct skeletogenic specification if the first circuit is disabled. Two effectors (*pmar1* and *blimp1*) are activated at the same time and by the same inputs. The authors constructed a re-engineered gene circuit that removed a downstream binding site for the back-up effector (*blimp1*). The wild-type and re-engineered systems could then be perturbed using morpholino antisense oligonucleotides. Under perturbation of *pmar1*, the wild-type circuit was able to correctly specify cell fate, whereas the re-engineered circuit was not. This 'fail-safe' circuitry ensures that development will proceed reliably and reproducibly, even under perturbation, and is an example of a case in which complexity is favored over simpler circuit topologies.

Future Directions

Progress in synthetic biology will allow more far-reaching questions to be investigated in the evolution of gene expression and regulation. Several technological driving forces are enabling researchers to design and construct larger and more complex biological systems. For example, the process of physically synthesizing and assembling pieces of DNA is rapidly becoming cheaper, faster, and more accessible to a wider range of laboratories. Efforts by Gibson *et al.* [27,28] have demonstrated how a combination of *in vitro* ligation and *in vivo* recombination can be used to assemble genes, pathways, and genomes. These technological advances will make it possible for a standard molecular biology laboratory to redesign and build entire chromosomes or genomes with reasonable labor and time investments. However, the ability to synthesize and construct large fragments of DNA has far outpaced the ability to design novel systems. The development of automated tools for designing 'well-behaved' biological parts [56,57] has shown promise in enabling researchers to forward engineer genetic circuits and will be necessary for the management of the complexity of large circuit design.

Linking Synthetic and Systems Approaches: Hypothesis Generation and Testing

The synthetic approaches described above complement more classical and systems approaches to determine fitness effects of evolutionary rewiring of gene expression control, and can provide ways to test hypotheses generated by these methods. Babu *et al.* [58] used the experimentally determined transcriptional network of *E. coli* to computationally predict the targets of transcriptional regulators across 175 prokaryotic genomes, which were then used to predict the regulatory network topology for each organism. The authors found that networks have evolved by placing orthologous genes in different transcriptional regulatory motifs. Organisms with similar lifestyles had similar preservation of interactions and motifs, suggesting functional advantages of a given motif for a given environmental niche. Investigating motif to niche fitness is an ideal problem for a synthetic approach.

Similarly, the dissection of gene expression level at quantitative trait loci has shed light on how genomes evolve in response to selection. Tirosh *et al.* [59] compared allele-specific expression of *S. cerevisiae*, *S. paradoxus*, and their hybrid to determine whether differences in expression were linked to *cis* (in the gene) or *trans* (in the regulatory factor) effects. They found that the majority of expression differences were due to changes in *cis*, although changes in *trans* were associated with genes that responded differently to the environment in each species. Similarly, sequencing methods developed by Brem and co-workers [60] were able to identify polygenic regulatory evolution in yeast species, or the accumulation of *cis* mutants that act the same way on gene expression in a pathway (increase or decrease activity). Recoded genomes that manipulate many *cis* or *trans* regions at once will allow us to understand their effects on organism fitness, which has been shown by the above studies to be central to genomic evolution.

Redesigning Pathways and Genomic Features

The ability to design and build biological systems at the genomic scale will open new avenues for studying how selective forces shape the patterns of gene expression, regulatory control, and genome architecture. For example, the adaptive value of features, such as operon organization, presence of introns, antisense transcription, and overlapping genes (amongst others), could be assessed by synthesizing entire genomes or subsets of genomes that lack or reorganize these features. To take one example, overlapping genes have been discovered in viral [61] and mitochondrial genomes [62], as well as in free-living microbes and eukaryotes [63,64]. Despite the enormous variety of microbial metabolisms and environments, overlapping genes appear to be a consistent feature of microbial genomes. Overlaps

are important in minimizing genome size (such as in viruses) and in the regulation of gene expression by introducing transcriptional or translational coupling [65,66]. For example, in the *E. coli trp* operon the termination codon of *trpE* overlaps with the start codon of *trpD* such that premature termination of *trpE* reduces the expression of *trpD* [67]. Parsing of the contributions of gene overlap on gene expression and fitness will require genome-scale redesign of protein-coding regions, given that approximately one-third of genes across sequenced microbial genomes are overlapping. In addition, removal of overlapping genes and addition of synthetic regulatory control could shed light on why overlapping genes are selected as a regulatory mechanism rather than (or in addition to) transcriptional, translational, and other mechanisms.

One area in which synthesis and manipulation of entire genomes may have a profound impact is the study of the genomic basis of lifestyle strategies in organisms. Lifestyle strategy refers to complex phenotypes, such as specialists or generalists in nutrient acquisition, or cooperative or selfish behavior in social organisms [68,69]. Recent work has begun to define genomic 'signatures' of particular lifestyle strategies in marine bacteria. Lauro *et al.* [70] identified genomic features of marine bacteria that have evolved to grow optimally at high nutrient concentrations (copiotrophic species) or low nutrient concentrations (oligotrophic species) by comparing the genome sequences of a model copiotroph (*Photobacterium angustum*) and oligotroph (*Sphingopyxis alaskensis*). Besides differences in the type and diversity of transporters, cell motility, and defense mechanisms, copiotrophs possess an abundance of genes related to signal transduction and transcriptional regulation. This is consistent with the hypothesis that copiotrophs are able to respond to a number of environmental stimuli and utilize multiple substrates for rapid growth. These results bring up interesting hypotheses that could be tested using synthetic approaches: can oligotrophic microbes evolve towards copiotrophic lifestyles (or vice versa) by manipulating the expression of existing protein-encoding genes, acquisition of novel protein-coding genes, or some combination thereof? This question could be addressed by a stepwise rewiring of gene-expression networks with the intention of switching an oligotroph to a copiotrophic lifestyle (or vice versa), which could delineate the minimum number of genomic changes necessary to occupy a new environmental niche. The ability to re-engineer lifestyle strategy and nutrient acquisition may also be useful in engineering industrial microbes for fermentation and biotechnological applications.

Resurrecting Ancient Proteins and Pathways

In the same way that synthetic biology allows us to build and evaluate alternative schemes of genetic regulatory control, it may allow us to construct and experimentally interrogate ancestral gene expression schemes and regulatory circuits. Recent advances in DNA synthesis and computational phylogenetics have made it possible to experimentally 'resurrect' ancient proteins in the laboratory [71–73]. Essentially, the sequence of an ancestral protein is inferred by using phylogenetics and statistical inferences on a family of modern day proteins and this sequence (or multiple candidate sequences) is chemically synthesized and expressed in a host organism or studied *in vitro* [74]. While the structure of ancestral gene regulatory networks and expression levels is undoubtedly difficult to infer given existing techniques, it may be possible to recreate some ancestral expression schemes using synthetic approaches. For example, while gene duplication is believed to be a major source of genetic novelty [75], the evolution of gene expression and regulatory control between duplicate genes is less understood. Computational analysis of microarrays, gene sequences, and regulatory networks suggests that, in duplicated genes, the expression profile of one gene evolves rapidly while the other gene maintains the ancestral (pre-duplication) expression profile [76,77]. Replacing natural regulation of duplicated genes with synthetic control would allow the effects of coupled (and divergent) gene regulation to be explored. Construction of multiple mutants with a range of combinations of coupled expression and different ancestral proteins could provide a recapitulation of the evolutionary route taken after gene duplication and could allow for the experimental investigation of the fitness effects of duplication.

Conclusions

The expanding ability to design and manipulate genetic regulatory schemes in living cells will allow biologists to tackle important questions in the evolution and emergence of gene expression control. There are several emergent features of biological systems that have been hypothesized to be favored in evolution, such as the modularity of regulatory and protein interactions [78], the robustness to environmental and mutational perturbations [79], and the ability of a regulatory network to foster evolvability [80]. Construction of biological systems will also allow for the parsing of the relative contributions of other mechanisms of diversity generation, such as protein-domain duplication and non-coding RNA control. The design, synthesis, and analysis of synthetic genomes with alternative genetic regulatory organization could become a powerful tool in asking precisely why the systems-level features we observe in biology have appeared throughout evolution.

Acknowledgements

The author thanks Jeff Tabor for useful discussions and anonymous reviewers for helpful comments. Synthetic biology in the Bayer lab is supported by EPSRC grant EP/G036004/1.

References

1. Dickinson, W.J. (1988). On the architecture of regulatory systems: evolutionary insights and implications. Bioessays *8*, 204–208.
2. King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107–116.
3. Malicki, J., Cianetti, L.C., Peschle, C., and McGinnis, W. (1992). A human HOX4B regulatory element provides head-specific expression in Drosophila embryos. Nature *358*, 345–347.
4. Malicki, J., Schughart, K., and McGinnis, W. (1990). Mouse Hox-2.2 specifies thoracic segmental identity in Drosophila embryos and larvae. Cell *63*, 961–967.
5. Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science *267*, 1788–1792.
6. Spitz, F., Gonzalez, F., Peichel, C., Vogt, T.F., Duboule, D., and Zakany, J. (2001). Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. Genes Dev. *15*, 2209–2214.
7. Hofer, J., Gourlay, C., Michael, A., and Ellis, T.H. (2001). Expression of a class 1 knotted1-like homeobox gene is down-regulated in pea compound leaf primordia. Plant Mol. Biol. *45*, 387–398.
8. Prum, R.O. (2005). Evolution of the morphological innovations of feathers. J. Exp. Zool. B Mol. Dev. Evol. *304*, 570–579.
9. Keys, D.N., Lewis, D.L., Selegue, J.E., Pearson, B.J., Goodrich, L.V., Johnson, R.L., Gates, J., Scott, M.P., and Carroll, S.B. (1999). Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. Science *283*, 532–534.

10. Brunetti, C.R., Selegue, J.E., Monteiro, A., French, V., Brakefield, P.M., and Carroll, S.B. (2001). The generation and diversification of butterfly eyespot color patterns. Curr. Biol. *11*, 1578–1585.
11. Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science *311*, 796–800.
12. Ihmels, J., Levy, R., and Barkai, N. (2004). Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae. Nat. Biotechnol. *22*, 86–92.
13. Andrianantoandro, E., Basu, S., Karig, D.K., and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. Mol. Syst. Biol. *2*, 2006.0028.
14. Endy, D. (2005). Foundations for engineering biology. Nature *438*, 449–453.
15. Benner, S.A., and Sismour, A.M. (2005). Synthetic biology. Nat. Rev. Genet. *6*, 533–543.
16. Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in Escherichia coli. Nature *403*, 339–342.
17. Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335–338.
18. Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G., and Collins, J.J. (2009). Synthetic gene networks that count. Science *324*, 1199–1202.
19. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H., and Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. Nature *434*, 1130–1134.
20. Austin, D.W., Allen, M.S., McCollum, J.M., Dar, R.D., Wilgus, J.R., Sayler, G.S., Samatova, N.F., Cox, C.D., and Simpson, M.L. (2006). Gene network shaping of inherent noise spectra. Nature *439*, 608–611.
21. Tabor, J.J., Bayer, T.S., Simpson, Z.B., Levy, M., and Ellington, A.D. (2008). Engineering stochasticity in gene expression. Mol. Biosyst. *4*, 754–761.
22. Ajo-Franklin, C.M., Drubin, D.A., Eskin, J.A., Gee, E.P., Landgraf, D., Phillips, I., and Silver, P.A. (2007). Rational design of memory in eukaryotic cells. Genes Dev. *21*, 2271–2276.
23. Balagadde, F.K., Song, H., Ozaki, J., Collins, C.H., Barnet, M., Arnold, F.H., Quake, S.R., and You, L. (2008). A synthetic Escherichia coli predator-prey ecosystem. Mol. Syst. Biol. *4*, 187.
24. Becskei, A., and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. Nature *405*, 590–593.
25. Purnick, P.E., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. Nat. Rev. Mol. Cell Biol. *10*, 410–422.
26. Cello, J., Paul, A.V., and Wimmer, E. (2002). Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science *297*, 1016–1018.
27. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., *et al.* (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science *319*, 1215–1220.
28. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.
29. Hanczyc, M.M., Fujikawa, S.M., and Szostak, J.W. (2003). Experimental models of primitive cellular compartments: encapsulation, growth, and division. Science *302*, 618–622.
30. Mansy, S.S., Schrum, J.P., Krishnamurthy, M., Tobe, S., Treco, D.A., and Szostak, J.W. (2008). Template-directed synthesis of a genetic polymer in a model protocell. Nature *454*, 122–125.
31. Yang, Z., Hutter, D., Sheng, P., Sismour, A.M., and Benner, S.A. (2006). Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. Nucleic Acids Res. *34*, 6095–6101.
32. Hirao, I., Harada, Y., Kimoto, M., Mitsui, T., Fujiwara, T., and Yokoyama, S. (2004). A two-unnatural-base-pair system toward the expansion of the genetic code. J. Am. Chem. Soc. *126*, 13298–13305.
33. Yang, Z., Sismour, A.M., Sheng, P., Puskar, N.L., and Benner, S.A. (2007). Enzymatic incorporation of a third nucleobase pair. Nucleic Acids Res. *35*, 4238–4249.
34. Winfree, E., Liu, F., Wenzler, L.A., and Seeman, N.C. (1998). Design and self-assembly of two-dimensional DNA crystals. Nature *394*, 539–544.
35. Rothemund, P.W., Ekani-Nkodo, A., Papadakis, N., Kumar, A., Fygenson, D.K., and Winfree, E. (2004). Design and characterization of programmable DNA nanotubes. J. Am. Chem. Soc. *126*, 16344–16352.
36. Shih, W.M., Quispe, J.D., and Joyce, G.F. (2004). A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. Nature *427*, 618–621.
37. Douglas, S.M., Dietz, H., Liedl, T., Hogberg, B., Graf, F., and Shih, W.M. (2009). Self-assembly of DNA into nanoscale three-dimensional shapes. Nature *459*, 414–418.
38. Tabor, J.J., Salis, H.M., Simpson, Z.B., Chevalier, A.A., Levskaya, A., Marcotte, E.M., Voigt, C.A., and Ellington, A.D. (2009). A synthetic genetic edge detection program. Cell *137*, 1272–1281.
39. Levskaya, A., Chevalier, A.A., Tabor, J.J., Simpson, Z.B., Lavery, L.A., Levy, M., Davidson, E.A., Scouras, A., Ellington, A.D., Marcotte, E.M., *et al.* (2005). Synthetic biology: engineering Escherichia coli to see light. Nature *438*, 441–442.
40. Yokobayashi, Y., Weiss, R., and Arnold, F.H. (2002). Directed evolution of a genetic circuit. Proc. Natl. Acad. Sci. USA *99*, 16587–16591.
41. Ellis, T., Wang, X., and Collins, J.J. (2009). Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat. Biotechnol. *27*, 465–471.
42. Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894–898.
43. Cordero, O.X., and Hogeweg, P. (2006). Feed-forward loop circuits as a side effect of genome evolution. Mol. Biol. Evol. *23*, 1931–1936.
44. Dwight Kuo, P., Banzhaf, W., and Leier, A. (2006). Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence. Biosystems *85*, 177–200.
45. Lynch, M. (2007). The evolution of genetic networks by non-adaptive processes. Nat. Rev. Genet. *8*, 803–813.
46. Gould, S.J., and Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc. R Soc. Lond. B Biol. Sci. *205*, 581–598.
47. Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature *436*, 588–592.
48. Kalisky, T., Dekel, E., and Alon, U. (2007). Cost-benefit theory and optimal design of gene regulation functions. Phys. Biol. *4*, 229–245.
49. Fong, S.S., Joyce, A.R., and Palsson, B.O. (2005). Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. Genome Res. *15*, 1365–1372.
50. Joyce, A.R., and Palsson, B.O. (2008). Predicting gene essentiality using genome-scale in silico models. Methods Mol. Biol. *416*, 433–457.
51. Fischer, E., and Sauer, U. (2005). Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. Nat. Genet. *37*, 636–640.
52. Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. Nature *443*, 415–420.
53. Cagatay, T., Turcotte, M., Elowitz, M.B., Garcia-Ojalvo, J., and Suel, G.M. (2009). Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell *139*, 512–522.
54. Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., Garriga-Canut, M., and Serrano, L. (2008). Evolvability and hierarchy in rewired bacterial gene networks. Nature *452*, 840–845.
55. Smith, J., and Davidson, E.H. (2009). Regulative recovery in the sea urchin embryo and the stabilizing role of fail-safe gene network wiring. Proc. Natl. Acad. Sci. USA *106*, 18291–18296.
56. Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. *27*, 946–950.
57. Villalobos, A., Ness, J.E., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006). Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics *7*, 285.
58. Babu, M.M., Teichmann, S.A., and Aravind, L. (2006). Evolutionary dynamics of prokaryotic transcriptional regulatory networks. J. Mol. Biol. *358*, 614–633.
59. Tirosh, I., Reikhav, S., Levy, A.A., and Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. Science *324*, 659–661.
60. Bullard, J.H., Mostovoy, Y., Dudoit, S., and Brem, R.B. (2010). Polygenic and directional regulatory evolution across pathways in Saccharomyces. Proc. Natl. Acad. Sci. USA *107*, 5058–5063.
61. Barrell, B.G., Air, G.M., and Hutchison, C.A., 3rd (1976). Overlapping genes in bacteriophage phiX174. Nature *264*, 34–41.
62. Fearnley, I.M., and Walker, J.E. (1986). Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. EMBO J. *5*, 2003–2008.
63. Fukuda, Y., Washio, T., and Tomita, M. (1999). Comparative study of overlapping genes in the genomes of Mycoplasma genitalium and Mycoplasma pneumoniae. Nucleic Acids Res. *27*, 1847–1853.
64. Spencer, C.A., Gietz, R.D., and Hodgetts, R.B. (1986). Overlapping transcription units in the dopa decarboxylase region of Drosophila. Nature *322*, 279–281.
65. Keese, P.K., and Gibbs, A. (1992). Origins of genes: "big bang" or continuous creation? Proc. Natl. Acad. Sci. USA *89*, 9489–9493.
66. Krakauer, D.C. (2000). Stability and evolution of overlapping genes. Evolution *54*, 731–739.
67. Oppenheim, D.S., and Yanofsky, C. (1980). Translational coupling during expression of the tryptophan operon of Escherichia coli. Genetics *95*, 785–795.
68. MacArthur, R.H., and Wilson, E.O. (1967). The Theory of Island Biogeography (Princeton, N.J.: Princeton University Press).
69. MacLean, R.C., and Gudelj, I. (2006). Resource competition and social conflict in experimental populations of yeast. Nature *441*, 498–501.
70. Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., DeMaere, M.Z., Ting, L., Ertan, H., Johnson, J., *et al.* (2009). The genomic basis of trophic strategy in marine bacteria. Proc. Natl. Acad. Sci. USA *106*, 15527–15533.

71. Benner, S.A., Sassi, S.O., and Gaucher, E.A. (2007). Molecular paleoscience: systems biology from the past. Adv. Enzymol. Relat. Areas Mol. Biol. *75*, 1–132, xi.
72. Thomson, J.M., Gaucher, E.A., Burgan, M.F., De Kee, D.W., Li, T., Aris, J.P., and Benner, S.A. (2005). Resurrecting ancestral alcohol dehydrogenases from yeast. Nat. Genet. *37*, 630–635.
73. Zhu, G., Golding, G.B., and Dean, A.M. (2005). The selective cause of an ancient adaptation. Science *307*, 1279–1282.
74. Thornton, J.W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. Nat. Rev. Genet. *5*, 366–375.
75. Wagner, A. (2008). Gene duplications, robustness and evolutionary innovations. Bioessays *30*, 367–373.
76. Conant, G.C., and Wolfe, K.H. (2006). Functional partitioning of yeast co-expression networks after genome duplication. PLoS Biol. *4*, e109.
77. Gu, X., Zhang, Z., and Huang, W. (2005). Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proc. Natl. Acad. Sci. USA *102*, 707–712.
78. Sauro, H.M. (2008). Modularity defined. Mol. Syst. Biol. *4*, 166.
79. Kitano, H. (2004). Biological robustness. Nat. Rev. Genet. *5*, 826–837.
80. Wagner, A. (2005). Robustness and Evolvability in Living Systems (Princeton, N.J.: Princeton University Press).