Chapter 1.

Synthetic Approaches to Understanding Biology

The construction of synthetic circuits has illuminated how cells process and respond to environmental signals, and how cells can be engineered to perform useful functions. Recent progress in the regulatory functions of noncoding RNAs suggests that nucleic acids can be employed as a design substrate for programming cellular function. However, as discussed below, the construction of increasingly complex systems will require an appreciation and understanding of trade-offs between biological functions. An understanding of how organisms tolerate and manage trade-offs in function between different environments could provide design principles for building robust systems. In addition, such work could lead to an understanding of whether the architecture of genetic circuits can shape the organization of larger-scale ecological networks.

1. Synthetic strategies for understanding genetic regulation

1.1 Reconstruction of genetic circuits to understand regulation

The complex functions of biological systems such as organisms, metabolic and developmental pathways, and proteins are ordinarily studied by analysis of genetic and biochemical perturbations. The modularity of biological components like genes and proteins enables a complementary approach: one can construct and analyze synthetic systems such as genetic circuits, organisms, and proteins with unnatural monomers inspired by their natural counterparts. Synthetic biology is the construction of existing or novel biological function from constituent components. Researchers may desire to do this for several reasons – one is to manipulate or measure existing biological systems in more sophisticated ways, leading to a greater understanding of biology. The study of how the structure of synthetic circuits relates to their behavior can potentially illuminate Nature's "design principles," or rules for how evolution has solved the adaptive challenges faced by an organism, although very few studies have yet linked network architecture to adaptive strategy for a given organism and environment. However, several examples of building genetic circuits to understand the biological significance of these architectures are noted below.

An examination of the adaptive and functional significance of network architecture is a daunting task given the complexity and diversity of examples from the natural world. One strategy to tackle this problem is not to break down the complexity of natural

networks, but to construct such systems from the bottom-up. The construction of biological regulatory circuits is at the heart of synthetic biology, and can be envisioned as a physical model of circuit architecture for testing hypotheses about the information processing, adaptive, and functional significance of such architectures.

Observations of the structure of genetic regulatory networks in yeast, bacteria, and other organisms have shown that these networks are comprised of repeated patterns of interaction between genes, knows as motifs.[1] Negative feedback, multi-input motifs, and feedforward loops are among the motifs that are observed far more often than would be expected for a random network. Several groups have tested the functional and adaptive significance of these motifs using a combination of theory, modeling, and construction of synthetic motifs. One example comes from the work on transcriptional feed-forward loops from the Alon lab. Feed-forward loops (FFLs) are three-gene motifs, where a first gene regulates a second, and the two in turn co-regulate a third gene. Two widely observed types are the *coherent* and *incoherent* FFL. Coherent FFLs consist of two activators that regulate a third component, while incoherent FFLs consist of one activator and one repressor.[2] Modeling these motifs with simple differential equations suggested that they have different information processing capabilities: signals that activate the expression of the first gene (the "input") are propagated with different temporal dynamics and strengths. For example, coherent FFLs act as signal delays and are able to filter transient inputs from affecting output gene expression. Incoherent FFLs can act as signal accelerators, showing faster induction of output gene expression than simple regulation by a single transcription factor. To test these models, Mangan and Alon examined the

arabinose utilization system of *Escherichia coli*, which shows FFL topology.[3] They found that the motif displayed "sign-sensitive" kinetics: the induction of motif output was slower than the signal decay after input was removed. This function could be employed by cells to filter noisy inputs (to avoid spurious gene expression, which is metabolically costly), while still being able to turn genes off rapidly when the input signal is removed. Thus, experimental evidence confirms the functional and potentially adaptive significance of a highly abundant feature of network organization.

Another example of the biological insight gained by constructing synthetic circuits comes from post-translational signaling pathways. Eukaryotes utilize mitogen-activated protein kinases (MAPKs) for a variety of signaling functions. In yeast, environmental signals initiate a cascade of phosphorylation events between kinases, ultimately resulting in changes in gene expression of a number of pathways, such as osmolarity responses and mating pathways. The MAPKs are associated with "scaffold" proteins of uncertain function: does the scaffold merely bring kinases in close physical proximity, or does it regulate, activate, and otherwise provide an additional point of control on the cascade? To test this, the Lim group replaced native protein-protein interactions between the kinase and scaffold with heterologous interaction domains from other proteins and found that the cascade was able to function properly.[4] The researchers extended these findings by creating synthetic scaffolds that brought together kinases that are not naturally associated. The synthetic scaffolds possessed novel input-output properties, demonstrating that scaffold proteins could be useful in evolving or engineering new cellular behaviors. In addition, these results convincingly showed that simple tethering is

sufficient to explain MAPK cascade behavior. This work demonstrates how synthetic approaches can complement and extend analytical approaches for biological discovery.

One underlying question in each of these cases is whether the observed network features represent functional units of information processing in the cell, or are merely evolutionary "artifacts" without functional significance. As discussed in section 4.3, general and global network topology and architecture can often be mistakenly ascribed to adaptive origins. A challenge for researchers is to test the functional and adaptive significance of observed patterns in genetic networks.

1.2 Construction of genetic circuits as an engineering discipline

Synthetic genetic circuits are valuable tools for understanding the natural world, and also have shown great utility in the engineering of biology. One advantage to a theoretical and practical understanding of biological circuit behavior is the ability to both design and evolve applications to many pressing technological problems such as therapeutics, energy, bioremediation, and material synthesis. Several examples of synthetic biology as an engineering discipline are described below.

Therapeutics and Human Disease

The diversity of chemical compounds made by biological systems found in Nature is staggering. Many natural compounds have historically been utilized as antibiotics and

other therapeutics.[5] Several groups have engineered microbes to produce fine chemicals for therapeutic purposes.[6] The Keasling group at UC Berkeley has engineered *Saccharomyces cerevisiae* to produce arteminisic acid, a precursor to artemisinin, a potent anti-malarial naturally found in the plant *Artemisia annua*.[7] Currently, the chemical synthesis of artemisinin is cost prohibitive to the population of the Third World, where it is needed most. The researchers were able to use an engineered mevalonate pathway, an amorphadiene synthase, and a cytochrome P450 monooxygenase from *A. annua* to produce artemisinic acid. The engineered yeast produced higher artemisin yields that *A. annua*, although the authors note that industrial scale-up and optimization will be required to make this route to production cost-effective.

In addition to engineering microbes to produce therapeutics, several groups are exploring the use of live cells *as* therapeutics.[6, 8, 9] Microbes in their natural state are endowed with many functions that could be utilized to discriminate between healthy and disease states (such as receptors and environmental sensing components) and act in a therapeutic manner (such as synthesizing therapeutic proteins, invading disease cells, or synthesizing chemicals, as above). Towards these aims, the Voigt group at UC San Francisco engineered *E. coli* to sense and destroy cancer cells by environment-dependent control of invasion.[10] The authors used invasin from the pathogen *Yersinia pseudotuberculosis* as an output that allowed *E. coli* to invade mammalian cells. To render the bacteria cancer cell-specific, invasin expression was controlled by several heterologous sensors: the *Vibrio fisheri* quorum sensing circuit, the hypoxia responsive fdhF promoter, or the arabinose-responsive araBAD promoter. Each of these is designed to induce invasin expression and

bacterial invasion only in the presence of tumor cell environments (for example, tumors are highly hypoxic) or via external, researcher-inducible control. The authors were able to demonstrate invasion of several cancer-derived cell lines with these engineered bacteria, demonstrating that cells can be programmed with sensors and outputs to achieve therapeutic functionality. Taken together, this work shows that natural functions of bacteria can be re-engineered and augmented to construct useful functions.

Biological Pattern Formation

The coordinated organization of cells in specific patterns is a classic example of complex function in many organisms and is central to the development of multicellular organisms from a single fertilized oocyte. Pattern formation typically involves signaling and communication between cells, processing of these signals, and modulation of the expression of a variety of genes. The ability to design pattern formation could hold great utility in applications such as tissue engineering and biomaterials. Towards these aims, several groups have explored how collections of cells can be programmed to form user specified patterns. In one example, researchers enabled *E. coli* cells to communicate with each other using the quorum sensing system from *Vibrio cholerae*.[11] Engineered "receiver cells" were designed to express fluorescent proteins based on the concentration of the signaling molecule synthesized by a "sender" cell. The receiver cells were designed such that they were responsive only to a defined range of signaling molecules, analogous to a bandpass filter. The range of signaling molecule the receiver cells was responsive to could be tuned by changing the kinetic parameters of the underlying

information processing circuit. The authors were able to use combinations of sender and receiver cells to create two-dimensional patterns on a lawn of cells such as a bulls-eye, ellipses, and clovers. Thus, the engineering of underlying functionality (i.e., quorum sensing) and gene circuits allowed spontaneous pattern formation in a population of cells.

Other approaches have been used to program pattern formation that are inspired more from lithographic and printing techniques rather than developmental pattern formation. In a stunning example of engineering synthetic functions into organisms, the Voigt and Ellington groups constructed a strain of bacteria that could sense red light and control gene expression accordingly.[12] To accomplish this, the researchers constructed a chimeric two-component photorhodopsin system from the cyanobacteria *Synechocystis* in *E. coli*. 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 at resolutions up to 100 megapixels per square inch. The control of pattern formation in living cells could have important applications in constructing complex patterned biomaterials, tissue engineering, and parallel biological computation. In an extension of this work, Tabor and Voigt (personal communication) have enabled massively parallel 'edge detection' of a projected image such that cells communicate to discriminate and delineate boundaries between cells sensing light and dark regions of the image. These efforts hold great potential to explore how biology uses large numbers of computational elements (in this case, cells) to compute complex problems and to combine "top-down" lithographic-style patterning with "bottom-up" parallel computation to specify patterns.

1.3 Nucleic acids as a substrate for designing circuits

The success of rational design of desired function in an engineering discipline is largely determined by the design substrate(s) available. One strategy to design component-level function in biology is to use nucleic acids as a substrate. The recognition that RNA plays a central role as not only an information carrier but as a catalyst of biochemical reactions and of genetic regulation has driven the development of diverse strategies for using RNA and DNA in multiple applications.

Roles of RNA in contemporary biology and in the evolution of life

RNA is pervasive in fundamental biological processes. As an information-rich molecule, it is responsible for carrying information between DNA and the translational machinery and guiding the processing and editing of ribosomal RNA (rRNA). In addition, RNA primes the process of DNA replication and individual nucleotides are used as cofactors in enzymatic reactions. Although RNA naturally is composed of only 4 nucleotides (in contrast to the 20 amino acids that make up proteins), it is able to fold into diverse tertiary structures that can display binding and catalytic activity. The crystal structure of the ribosome revealed that the catalyst of protein synthesis is in fact RNA – a function that likely doomed the RNA world to extinction, replaced by the protein universe. In 1980, Cech and Inoue discovered that the splicing reaction of a rRNA intermediate in *Tetrahymena* was able to proceed in the absence of a protein enzyme via a specific

tertiary structure in the RNA able to perform the catalysis, which they termed a ribozyme.[13] Several other naturally occurring cleavase ribozymes have been discovered in central roles in biological functions, including RNA processing and viral genome replication. The biochemical functionality of RNA is highlighted by the recent discovery of RNA regulatory elements in prokaryotic metabolic genes.[14] Breaker and co-workers discovered conserved RNA structures in the 5' untranslated regions (UTR) of genes involved in cofactor metabolism in *Bacillus subtilis*[15]. These structures were able to directly bind the small molecule cofactor, and by virtue of a conformational change upon binding, occlude the ribosome-binding site (RBS) of the mRNA and control translation. These riboswitches have now been observed in prokaryotes, archaea, and eukaryotes and have been found to regulate translation, transcript stability, and alternative splicing in response to metabolite effectors.

In the past 10 years, the discovery and characterization of a highly conserved RNA-based regulatory mechanism has fundamentally altered the conception of the biological function of RNA in cells. RNA interference (RNAi) was first described in *Caenohabditis elegans* and later shown to be present in many eukaryotes such as fission yeast and mammalian cells[16]. In RNAi, a small (21 nt) double-stranded RNA effector guides the sequence-specific silencing of genes. RNAi silencing is guided either by exogenously delivered small interfering RNAs (siRNAs) or by endogenously produced microRNAs (miRNAs). These RNAs silence the expression of target genes in several ways: one is by base-pairing complementarity to a target transcript, mediated by a multi-protein complex known as RISC (RNA-induced silencing complex). Catalysis is performed by the

endonuclease Argonaute 2 in the complex. RNAi has been used as a tool for targeted gene knockdown in mammalian cells, but has also been shown to be a significant mechanism for cellular regulation. miRNAs are able to fine-tune gene expression during differentiation and development by recognizing the 3' UTR of target genes. The loss of miRNA-mediated regulation in mutants has been linked to oncogenesis as well as developmental defects.[17]

Taken together, these discoveries clearly demonstrate the importance of RNA in biological systems far beyond the original conception that RNA was merely an information carrier. One way that the powerful biochemical, regulatory, and informational properties of RNA can be expanded upon is by directed evolution in the laboratory - recapitulating Darwinian selection on populations of functional RNAs.

Evolving function from populations of RNA

Contemporary theories on biological origins suggest that life arose via a self-replicating molecule or assembly of self-replicating molecules. Because of the almost unique complementarity inherent in nucleobases, it has been suggested that nucleic acids or nucleic acid-like molecules were the first self-replicators, although some researchers hold that self-replicating peptides or lipid systems may have preceded or emerged in parallel with nucleic acids.[18] An early nucleic acid replicator is also tempting because of the prevalence of molecular fossils in modern cells that are related to nucleic acids, such as ATP and other cofactors. Given the fact that the ribosome is at its core a ribozyme, it

seems highly likely that there was a complex RNA world that preceded the modern world of protein catalysts.[19] This RNA world may have in turn descended from an early nucleic acid replicator by duplication, parasitism, and diversification.

Researchers have recapitulated the Darwinian evolution of RNA self-replicators as well as other functional RNAs.[20] One of the earliest examples of extracellular evolution of molecules is the work done by the Spiegelman group.[21] In this work, the researchers discovered that the bacteriophage Qβ utilized a RNA-dependent RNA polymerase to replicate its RNA genome. In this simple system (containing replicase, template RNA, and nucleotides) the researchers found that the system showed autocatalytic kinetics, indicating that self-propagation of the viral genome was occurring and could do so *in vitro* with the necessary components. Furthermore, the Qβ system was used to set-up an extracellular evolution experiment where mutant templates could "compete" with one another for the limited pool of replicases and nucleotides, eventually selecting for template sequences that were able to replicate faster than the original (parent) template[22].

The strategy of competing populations of diverse RNA molecules (analogous to Darwinian evolution) was used by Ellington and Szostak to evolve ligand-binding species. The authors termed the evolved ligand-binding RNAs "aptamers" (from the Greek *aptus*, to fit). This work[23] demonstrated how randomized pools of molecules could be selected to bind organic dyes. Further work demonstrated that aptamers could be evolved to bind other small molecules, proteins, oligosaccharides, as well as cells, tissues, and organisms (by virtue of molecular recognition on the cellular surface).[24]

Aptamers rival protein antibodies in binding specificity and affinity, and have been shown to elicit minimal immunological responses[25]. These properties have led to the development of aptamer therapeutics[25]. The first such therapy to gain FDA approval is an aptamer that binds the vascular endothelial growth factor (VEGF), used in the treatment of age-related macular degeneration.[26]

The success of aptamer selections has enabled several groups to explore the unique structural biochemistry of binding between aptamer and ligand. Because aptamers are evolved solely for binding functionality, the resulting structures are not optimized for other functionality and can provide insight into the universe of possible chemical solutions for RNA-ligand binding. Three-dimensional structural analyses have provided insights into the nature of recognition by nucleic acid-aptamer complexes.[24] The enclosure of large regions of the ligand by the nucleic acid is the basis for specific recognition of the ligand in aptamer complexes. Multiple intermolecular contacts between the nucleic acid and ligand provide specificity. For example, steric occlusion of a methyl group prevents binding of caffeine to the theophylline aptamer.[27] Specific hydrogen bonding creates an interaction with the ligand similar to a base-pair in the AMP aptamer.[28, 29] Aminoglycoside antibiotics are bound by their aptamers by a combination of both electrostatic and hydrogen bonds that create complementarity between ligand and aptamer.[30] Aptamers selected to bind peptides and proteins from HIV-Rev and other viruses often involve stacking, electrostatic contacts, and induced fit of both the peptide and the aptamer.[31, 32] In nucleic acid - ligand binding, the structurally similar nucleotides are limited in the number of ways that they can be structured around a given ligand. The

interaction of ligands into aptamer binding sites often displays imperfect complementarity, which can be compensated for by adaptive recognition – structural rearrangement of the ligand and/or the aptamer.[33, 34] Furthermore, many aptamers contain disordered loop regions that acquire an ordered structure upon ligand binding. Adaptive recognition and the formation of specific contacts between aptamer and ligand are responsible for a general trade-off between binding affinity and an inability of several aptamers to bind variant forms of their specified ligand. For example, aptamers selected to bind the HIV Rev peptide with high affinity are unable to bind even single mutant forms of the peptide, limiting their use as therapeutics for the rapidly mutating HIV virus.[35]

In parallel with the evolution of binding functionality in aptamers, several groups developed schemes for the evolution of catalytic ability from randomized nucleic acid pools. Catalytic RNAs are found throughout biology, as in the protein translation machinery (the ribosome), as well as controlling gene expression[36] and carrying out splicing.[37] Using similar methods to the *in vitro* selection of aptamers (creating pool diversity, partitioning higher fitness molecules, and amplification) researchers have been able to discover RNAs that catalyze a wide range of reactions, including ligation of other nucleic acids,[38] cleavage of nucleic acids,[39] and even a ribozyme that catalyzes an alcohol dehydrogenase reaction,[40] showing that ribozymes could perform redox chemistry. Catalytic nucleic acids have allowed structural insights to nucleic acid function and have played roles in engineering new functions, as described below.

The crystal structures of several natural and *in vitro* selected ribozymes have been solved, giving insights on structural design principles for RNA catalysis.[41] One of the more general themes that emerges from this body of work is that ribozymes catalyze reactions in the same ways that proteins do: by forming substrate-binding sites to decrease the entropic cost of attaining the transition state, having more favorable interactions with the transition state structure versus the ground state, enabling chemistries that involve the movement of protons, and raising the energy of the bound substrate relative to the transition state. Ribozymes are not limited to using only metal ions as functional groups in catalysis (as was proposed in the infancy of ribozyme research), but can use nucleotide bases, sugar hydroxyls and phosphate backbone to accomplish chemical functions[41]. Due to limitations in the diversity of side chains, ribozymes are not as adept at catalyzing the wide variety of reactions that protein enzymes can, which may be one reason that the protein-based biology we observe today triumphed over its RNA-based ancestors.

A recently published ribozyme structure highlights the ability of directed evolution to find novel biochemical "solutions" to a particular catalytic "problem". The L1 ligase was isolated by *in vitro* selection to catalyze the ligation of two RNA molecules.[42] The ribozyme catalyzes a nucleophilic attack by a 3'-hydroxyl group on the phosphorus of the ribozyme's 5'-triphosphate, creating a new phosphodiester linkage and releasing the pyrophosphate. The ligase creates a catalytic pocket by a unique triple base interaction, with contributing residues from invariant nucleotides on each of three helices[43]. This catalytic core juxtaposes the ends of the ligation reactants with a Mg^{2}+ ion. Although this structure and mechanism share some features with that of the hammerhead ribozyme, the

nucleotide geometry involved in catalysis in this ligase has not been observed before, but was "discovered" by directed evolution of ligase function from a random pool. This suggests that even with an incomplete understanding of biophysical principles, directed evolution remains a powerful approach to creating functional nucleic acids. The directed evolution of nucleic acid functionality has enabled the construction of useful biotechnological tools.

Harnessing evolution and design of RNA to engineer function

RNA has two properties that enable researchers to engineer and evolve novel and useful functions for biotechnological applications. One is the above noted biochemical functionality of RNA, such as binding and catalysis. Another is the informational functionality of RNA, the ability of the primary sequence of nucleotides to be "read off", amplified, and manipulated with available enzymes and techniques in a test tube. While all biological molecules can be said to possess information content in the sequences of monomers or chemical structures composing the molecule, the ability to transform or amplify this chemical information (such as protein sequence or metabolite functional groups) has not been demonstrated. In contrast, the information for a particular ribozyme-catalyzed reaction is embodied in the sequence of the RNA molecule, which can be reverse transcribed and PCR amplified to create millions of copies of the ribozyme information. Researchers have exploited this dual informational/biochemical property to develop strategies for biosensing, genetic regulation, cell-specific targeting, and other applications.

An example of combining directed evolution with rational design to create functional nucleic acids is work done by the Sullenger lab on reversible aptamer-based anticoagulants. Sullenger and colleagues first selected an aptamer to bind the clotting factor Ixa.[44] Blood plasma based screening revealed that the aptamer was a potent anticoagulant that acts by blocking further activation of the clotting factor zymogen cascade. The authors then introduced "antidote control" to the aptamer through addition of a complementary oligonucleotide that hybridizes to the aptamer. Thus, the design of an aptamer-antidote pair is intuitive: each aptamer sequence inherently carries the prescription for its antidote. The authors further demonstrated the power of nucleic acid engineering by demonstrating the use of this anticoagulant – antidote pair in a mouse model.[45] To overcome the inherent instability of nucleic acids in the bloodstream, chemically modified nucleotides were introduced together with a pendant cholesterol group to improve bloodstream retention time.

In another example of the dual roles of selection and design in creating functional nucleic acids, several groups have created cell-targeting RNA molecules towards the goal of disease-cell specific drug delivery. The Levy group at Albert Einstein College of Medicine used aptamers selected to the known cancer biomarker prostate-specific membrane antigen (PSMA) that displayed high affinity and specificity.[46] They then conjugated these aptamers to the drug gelonin – due to receptor cycling to and from the membrane, the aptamer-drug conjugate bound PSMA on the cancer cell surface and was internalized, killing the cell through the action of gelonin.[47] The aptamer-drug conjugates

showed 600-fold increased potency compared to cells that did not express PSMA. These results demonstrate how selection can be used to create "magic bullet" therapeutics that target disease cells while leaving healthy cells unharmed, lowering the dose of drug needed and mitigating side-effects to the patient. In separate studies the same PMSA aptamer was used to deliver gene-targeting siRNAs to prostate cancer cells,[48, 49] extending the therapeutic possibilities of that aptamer delivery method. One promising possibility for the creation of cell targeting conjugates is the advent of whole-cell selection procedures. In the above example, PSMA, a known cancer marker, was purified to select aptamers against. However, in many cases extracellular markers of disease are not known. In these cases, selection of aptamers to living cells can be used to discriminate differences on the cell surface. In a recent study,[50] aptamers were generated by selecting against entire leukemia cells. Specificity was achieved by performing a "negative selection" against healthy cells. A negative selection partitions and discards aptamers that bind the negative target (in this case healthy cells) after the selection step (in this case, partitioning of aptamers that bind leukemia cells). Cell-based aptamer selections can be used to evolve aptamers to discriminate between any cell types, and can show utility as cell-specific therapeutics when coupled with drugs or toxins as above.

Information processing with nucleic acids

The dual functions of nucleic acids observed in biology of information carrier and biochemical actuator have been coupled to demonstrate how sets of biomolecules can perform elementary computation with molecular inputs and outputs. A seminal paper by

Adelman demonstrated that nucleic acids can be used not only to encode information but could perform elementary computational operations that could theoretically scaled to yield massive computational power.[51] These efforts can aid in creating complex circuits for nanoscale computation and in understanding how biological systems process information given the constraints of physics at the molecular scale. Several examples are described below.

Stojanovic and Stefanovic used deoxyribozymes to demonstrate the computational abilities of nucleic acids could interpret multiple simultaneous inputs by designing a DNA automaton that could play a game of "tic-tac-toe" against a human player.[52] The outputs of the device were deoxyribozyme cleavage events that yielded single-stranded oligonucleotides. The user-supplied inputs were also single-stranded DNA. Boolean logic linking inputs to outputs was accomplished by engineering allostery into the deoxyribozymes – upon addition of input oligos, the catalysts cleave, yielding an output.[53] The authors were able to incorporate allosteric domains in such a way that the higher order logic functions required to play the game against a human were accomplished. The automaton strategy is thus "hard-wired" in the structures and arrangements of the deoxyribozymes. While this work required a large amount of empirical tuning (to achieve reasonable signal-to-noise ratios, for example), it remains as one of the first examples of building a complex, predictable network from programmable enzymes.

In recent work, Winfree and colleagues demonstrated how cascades of logic functions can be implemented using strand displacement of DNA, without catalysis.[54] The authors were able to show that DNA-based logic gates using their design (based completely on base-pair hybridization) showed the hallmarks of digital abstraction employed in silicon-based electronics such as logic, cascading, signal restoration, fan-out, and modularity. Because of these features, the authors were able to build a complex device using 11 logic gates and performing complex computational tasks such as signal amplification, restoration, and threshold behavior. This work establishes design principles for information processing by nucleic acids, and could be used to control nanoscale devices *in vitro*, to analyze complex chemical samples, or to interface with existing biological circuits inside living cells.

Genetic regulation: information processing and biochemical functionality inside cells

Several groups have begun to use the ease of RNA directed evolution and rational RNA design to construct components for the regulation of gene expression programs inside cells. Researchers have used the ability to rationally design interaction energetics between RNA molecules (via base-pairing) to construct a post-translational regulation system in *E. coli*.[55] Translation in prokaryotes requires the interaction of the ribosome with a (RBS), and as such, translation can be inhibited by occluding the RBS. The researchers used a *cis* stem-loop structure that binds the RBS to silence gene expression. In an elegant demonstration of the power of base-pairing specificity and binding energetics between RNA molecules, the authors then used a *trans* RNA molecule to

activate gene expression. The *trans* RNA targets the *cis* repressive stem-loop via base-pairing, causing a conformational change that activates translation. The interactions of *cis* and *trans* elements are highly specific due to the combinatorial nature of base-pairing, and as such the same design schemes could be used to construct a large number of orthoganal *cis-trans* riboregulator pairs to build increasingly complex networks.

Several groups have utilized the ligand-binding properties of RNA as well as sequence specificity to engineer novel behaviors in living cells. The theophylline-binding aptamer was inserted in the 5' UTR of a reporter gene in *S. cerevisiae* to achieve ligand-dependent regulation of the target gene.[56] Another group then used this scheme to control gene expression in *E. coli* and evolved the dynamic range of this system, from 12-fold difference in gene expression with and without theophylline to a 36-fold difference.[57] The evolved riboswitch was then cloned upstream of the CheZ gene which led to control of bacterial motility in the presence of theophylline[58]. Cells were able to "follow" a trail of theophylline arranged specifically on a plate. Although the engineered chemotactic system differs from the normal control mechanism in several ways (for example, regulation is based on slow changes in gene expression rather than the wildtype protein phosphorylation cascade), the "phenotype level" function is the same – cells sense and move towards a user-specified chemo-attractant. These studies show how RNA can be designed, evolved, and engineered to meet the functional demands for engineering complex behaviors and functions in living cells.

Finally, several qualitative and quantitative differences between regulation of gene expression with noncoding RNA and regulation with proteins have recently been elucidated. Hwa and colleagues combined theoretical and experimental approaches to study regulation by a class of *trans*-acting noncoding RNAs in *E. coli*, small RNA (sRNA). There are over 70 identified sRNAs in *E. coli* that have been implicated in regulating diverse functions such as osmotic response, quorum sensing oxidative stress, DNA damage, and others[59]. This work found that regulation by sRNAs involved a "threshold-linear" response where repression is tight under a threshold of sRNA synthesis rates and is linear depending on target gene and sRNA synthesis above the threshold. In addition, compared with protein regulation, sRNA regulation displayed characteristic noise resistance and a capability for "hierarchical cross-talk" - many sRNAs bind several targets, and because target mRNAs can titrate sRNA by binding them, two target genes can thus have an indirect affect on each other. These results demonstrate that in addition to "programmability" and ease of design, engineering genetic regulation with noncoding RNAs may allow engineers to build novel quantitative and qualitative modes of regulation in living cells.

2. Constraints and trade-offs with increasing complexity

As we begin to assemble components to create functional devices, a new understanding of potential trade-offs and design criteria for optimization will be necessary. Trade-offs and constraints can bias the design criteria of natural and engineered systems. The

engineering of more complex systems will require an understanding and appreciation of the trade-offs between biological functions.

2.1 Trade-offs shape circuit design and function

An understanding of constraints and trade-offs in system function is critical for understanding how to construct more complex biological devices. Design principles will hold the greatest utility if underlying complexity can be effectively "hidden" from users at higher levels. An understanding of the foundations of trade-offs and constraints will enable biological engineers to have a set of "rules" for composing modules and a guideline for expected behavior. Recent work in circuit analysis suggests that trade-offs between system functions may require a greater appreciation of trade-offs at the cellular scale.

Recent work has described how the architecture of the *E. coli* heat-shock response is able to balance a trade-off between temporal response and efficiency. Heat shock causes the rapid unfolding and aggregation of many proteins in the cell, compromising normal function. Cells respond to heat shock by inducing the expression of heat-shock proteins (Hsp), chaperones that enable unfolded proteins to fold to their native structure, as well as proteases that degrade misfolded protein that are beyond repair.[60] Because the effects of heat shock occur on the order of seconds, the heat shock response must be activated quickly. However, the spurious induction of this massive cellular response is

metabolically costly for the cell, requiring tight regulation of its induction. Thus, the system must be optimized for several functionalities such as robustness, efficiency, response speed, and rejection of noise. Given limited cellular resources and energy, these demands are often contradictory – for example, high turnover rates in heat shock regulator production are necessary for a fast response, but come at a cost to metabolic efficiency.[61] Recent work has demonstrated how the architecture of the heat shock system balances such trade-offs.[62] For example, numerous feedback and feed-forward loops were identified that couple the induction dynamics, steady-state levels, and decay of multiple regulators of the heat shock response. Interestingly, the researchers note that the evolution of control architectures to balance these trade-offs results in new fragilities and constraints in the system, a recurring theme in the analysis of biological and man-made systems. The inherent trade-offs and constraints between biological functions and traits are also a significant determinant in shaping how organisms adapt and evolve.

2.2 Trade-offs in life history evolution

The recognition that trade-offs and intrinsic constraints can play a large role in determining the composition of systems has a lengthy precedent in evolutionary biology and ecology. A central assumption in many theories of trait evolution is the existence of trade-offs between functions. Trade-offs are based in fundamental mechanical, physiological, or thermodynamic constraints in metabolism, behavior, morphology, and other functions. One clear example of such constraints is the widely observed trade-off is the trade-off between fertility and survival. To take one case, the survival versus fertility

trade-off in the guppy *Trinidata sp.* is determined by fluid mechanics of swimming.[63] Because guppies are viviparous (give birth to live offspring rather than laying eggs) individuals with large litter/clutch sizes tend to be larger and more rotund than individuals with small litter sizes. The difference in size creates more drag in the water, making it more difficult for the individual to avoid predators. This trade-off can shape adaptive strategies available to a population and can reveal "signatures" of adaptation in the field due to selection determining the optimal balance of traits along the trade-off (from high fertility to high survival). Thus, populations that experience low predation would tend to optimize fertility over survivability, while populations experiencing high predation would tend to optimize survivability. The knowledge of mechanistic determinants of a trade-off can thus illuminate the ecological and evolutionary history of populations in the wild.

A related and prominent trade-off theory in biology is that of *r* versus *K* selection.[64] Organisms under *K* selection are predicted to optimize utilization of resources, such as when the population is near its carrying capacity and resources are scarce. *r* selection occurs when resources are abundant and organisms maximize per capita growth rate. For example, organisms generally face a trade-off between two modes of ATP production, respiration and fermentation – fermentation is faster but less efficient than respiration. Environmental conditions can select for organisms displaying a rate or yield strategy in this case, as is detailed below.[65] Although a given organism may employ more than one ecological strategy, the hypothesis of a general trade-off between growth rate and yield

(the efficiency of resource utilization) remains central to theories regarding the evolution of cooperation and evolution of generalists versus specialists.

Similar, but importantly distinct, trade-offs are well established in metabolic function. The laws of thermodynamics imply a trade-off between the rate of ATP production (moles ATP / time) and the yield (moles ATP / moles substrate) for any catabolic reaction.[66] This trade-off has been experimentally demonstrated in *E. coli* and *S. cerevisiae.*[65] One example is the rate-yield trade-off in two ubiquitous modes of sugar degradation, respiration and fermentation. Respiration is rapidly saturated at high levels of substrate or limiting oxygen, such that organisms can increase the rate of ATP synthesis by fermentation in addition to respiration. However, the yield for fermentation is much lower than respiration (2 moles ATP / mol glucose for fermentation, 32 moles ATP / mol glucose for respiration). Thus, cells can either "choose" to produce ATP rapidly (fermentation) or to extract a higher yield of ATP from a given amount of substrate (respiration). Pfeiffer and Bonhoeffer[66] showed that high yield ATP production can be viewed as cooperation between cells and can evolve in spatially structured environments.

An understanding of potential design constraints comes from an intimate knowledge of the (often) conflicting strategies of coping with environmental challenges. One recent example that has been elucidated is the trade-off between multiplication inside a host and survival outside a host in *E. coli* infecting bacteriophage.[67] Here, the authors found that phage mortality rate outside the bacterial host is inversely correlated with multiplication

rate for the viruses inside the host. Survival time outside of a host is a critical phase of the life cycle for parasites transmitted via the environment. Coliphages can have significant variability in the period between hosts – for example, a phage may be able to infect other *E. coli* cells in the same animal gut (short time between infection), or may be excreted into the external environment (long time between infection). The authors were able to propose and support a mechanism for the trade-off between reproduction and survival. The major determinant of viral survival is the stability of the capsid shell. This study found that the surfactic mass of the capsid as well as the density of the packaged phage genome was positively correlated with survival[67]. However, the energetic and temporal demands for stronger and more densely packed capsids resulted in slower virion production rates inside the host. In this case, evolving population are forced to optimize and balance a trade-off between fitness benefits in two different environments. Human engineers designing a system such as a bacteriophage would need a mechanistic understanding of capsid thickness and genome packaging to determine which function to optimize (reproduction or survival) or how to balance functionality in their engineered system.

3. Genome-scale organization

As we move towards an understanding of how constraints, trade-offs, and optimization drives the evolution of biological function, an open question is whether there are overarching principles or themes that accurately depict the organization of entire biological systems. Such “systems design principles” would be valuable in genome-scale

engineering. The capability to chemically synthesize large DNA sequences is rapidly becoming faster, cheaper, and more reliable such that the synthesis of entire genomes could be feasible for academic or industrial labs.[68] At this point, the rate-limiting step of engineering biology will not be fabrication (i.e., cloning), but will be design. In addition, the existence of fundamental genome-scale design themes could enable a deep understanding of adaptation at the systems level.

3.1 Organizing biology: engineering perspectives

One paradigm for designing biological systems is the use of an abstraction hierarchy for managing complexity.[68] Abstraction hierarchies are widely used in software engineering. For example, high-level programming languages (such as C++) enable programmers to read and write code in a form that is understandable to humans. The high level language is translated into machine-level instructions, which are in turn translated into bit strings for interpretation by the machine. Thus, software engineers can work on specific parts of the hierarchy such as high-level languages or machine-level instructions.

The effect is two-fold, and has many potential similarities in biology and evolution. The first is that the abstraction allows ease of design. Specialists in designing instruction sets (machine-level) do not have to create novel machine-level methods for each new high-level application, whereas high-level programmers do not need to concern themselves with the detail of machine-level operations when writing new code. Another effect of an abstraction hierarchy is the promotion of diversity in high-level applications. Software

engineers are able to use, re-use, and combine high-level functions (specified at the machine-level) to create a myriad of applications. The concept of the recombination of functional modules to create novel systems has parallels in protein evolution. Recent work with eukaryotic signaling proteins has shown how the recombination of a common set of domains can create novel input-output functionality such as allosteric gating, signal integration, and ultrasensitivity.[69]

Several abstraction hierarchies for synthetic biology have been proposed. One is the "parts-devices-systems" hierarchy.[70] In this concept, at the lowest level of complexity, are parts – discrete sequences of DNA that encode for functional units such as (but not limited to) genes, RBS, transcriptional terminators, and promoters. These parts can be characterized and assembled into devices. Devices include multiple parts with a specified function, such as a transcriptional inverter, a signal amplifier, or a toggle switch. At a higher level, devices can be assembled into functional systems. Some examples include oscillators, pattern-forming cells, and tumor-invading bacteria described previously. One goal of synthetic biology is to explore the utility of such abstraction hierarchies in the pursuit of engineering biological function.

3.2 Optimality principles in network organization

Optimal function is a fundamental and shared idea between engineering and evolutionary biology. One open question is toward what functions evolution has optimized metabolic networks. Organisms face multiple challenges from their environments, including

fluctuating nutrient sources, a battery of stress conditions, and potentially deleterious (and beneficial) interactions with other organisms. The optimal organization of metabolic networks would shed insight on the evolutionary forces that have shaped networks as well as provide "design principles" for the forward engineering of organisms for metabolic engineering. The Palsson group has shown that the "object" of metabolic network organization (the "objective function" in their parlance) is to maximize growth, and as such, potential fluxes through a metabolic network can be derived.[71] For example, microbes have evolved to efficiently convert carbon and energy into biomass, the creation of more cells. In an elegant demonstration of this objective function, Fong and Palsson grew *E. coli* on glycerol (a non-preferred carbon source for this organism) and observed that metabolic fluxes did not operate according to the optimality principle.[72] However, after applying selective pressure by repeated growth on glycerol, the network evolved to maximize its growth rate on this substrate according to the predictions described by an *in silico* model. The utility of this simple evolutionary principle has been validated by its use to predict essential genes in metabolic networks, as well as its use to optimize ethanol production in *S. cerevisiae*.

However, alternative selective pressures may result in metabolic networks that are optimized for other functions besides growth. The Sauer group analyzed metabolic gene knockouts in *Bacillus subtilis* and found that several mutants grew faster than wildtype, showing that bacteria does not operate solely according to maximized biomass production.[73] The authors found that *B. subtilis* has suboptimal metabolism because it invests a significant amount of resources in anticipation of changing environmental

conditions, which the authors term a "stand-by mode". For example, most identified knockout 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, network organization 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.

3.3 Game theoretic strategies for maximizing competitive ability

Optimality principles such as those described above often make the implicit assumption that populations gradually ascend peaks in a fitness landscape, and as such, gradually increase in fitness over time. However, fitness is often determined and dependent on the fitness of other members of an organism's population or on other interacting species. For example, the selective advantage of a particular tree height depends on the height of surrounding trees. The evolution of a successful immune response to a pathogen will elicit selection pressure leading to enhanced variants of the pathogen. Thus, while optimization theory is useful in describing gradual adaptation, game theoretic explanations have been utilized to describe evolution among interdependent agents. Game theory originated to describe interactions among independent economic agents, and was quickly applied to evolutionary biology. The success of a given strategy in a

game theoretic setting is dependent on the strategies of other players. The biological analog is that the success of a given genome is dependent on the make-up of the population.

The classic example of game theory is the Prisoner's Dilemma.[74] In this game, players receive a benefit if they cooperate (W), but a larger benefit if they act selfishly (X, the so-called "temptation to defect") when the other player cooperates (with Y being the penalty for cooperating when the other player defects). When both players defect, each receives a penalty Z. The Prisoner's Dilemma arises when the payoffs for each action are ranked

$$X > W > Z > Y$$

Thus, strategies that always cooperate can be exploited by strategies that always defect. However, populations of defectors are less fit than populations of cooperators. The Prisoner's Dilemma highlights the difficulty of understanding how cooperation arises from populations of organisms and is captured in the "tragedy of the commons" – how can groups of cooperating individuals resist invasion by selfish individuals? In recent work, MacLean and Gudelj demonstrate mechanisms that can enable cooperators to coexist with selfish individuals: spatial and temporal heterogeneity can allow either coexistence or competitive exclusion, depending on trade-offs in metabolism.[65]

Interestingly, observations of classical Prisoner's Dilemma situations in Nature are lacking. One clear example comes from the work of Turner and Chao on the bacteriophage ϕ-6.[75] The authors were able to measure the payoff matrix associated with cooperation (the manufacture of diffusible products inside the bacterium during infection)

and defection (the sequestration of these products). In this study, selfishness (defection) evolved in viral populations with high multiplicity of infection (MOI, the number of independent viruses infecting a single bacterium) and actually lowered the fitness of the evolved viruses relative to the ancestral virus. Spatial segregation of viral genomes via infection at low MOI evolved clones that showed high fitness and cooperative behavior. Thus, the outcome of simple evolution experiments was heavily dependent on the frequency of other viral genomes in the environment, showing how game theory can complement optimization theory for describing adapting populations.

The fitness of a given network feature may in fact be frequency-dependent: like the success of rare alleles, specific regulatory features may be selected for depending on their relative abundance in a population. An intriguing hypothesis is that regulatory and metabolic architectures are organized to promote cooperation among individuals. Investigations along these lines are only now beginning to be explored.

3.4 Robustness as an organizing principle

One potential recurring design theme in biology may be the organization of cellular networks to be robust to perturbation (both genetic and non-genetic perturbation). Robustness has a long precedent in both engineering disciplines as well as biology. As far back as 50 years ago, Waddington described the robustness of morphological features of organisms to perturbations during development, and ascribed this robustness to natural selection operating to produce organisms that were optimized to some intermediate form

(i.e., stabilizing selection).[76] The idea that robustness is a dominant organizing principle in biological networks is intriguing and continues to be investigated today.

Robustness to non-genetic change

Evolving populations undoubtedly face perturbation and uncertainty from their environments. Organisms have evolved countless mechanisms to maintain function during changes in external (and internal) conditions. For example, the maintenance of osmotic balance, metabolite concentrations, and other homeostatic mechanisms can be considered mechanisms for robustness to non-genetic change, as can thermoregulation in endothermic organisms, flight stabilization in birds, and predator avoidance behavior in higher organisms. The robustness of cellular networks has been explored as well. Barkai and Leibler used a simple two-state model to show how the connectivity of the components of the bacterial chemotaxis network confers robust adaptation.[77] Chemotaxis is a process by which bacteria are able to move toward or away from certain chemicals in the environment by a series of "smooth runs" of motility in one direction punctuated by events of "tumbling" where the direction for a next run is chosen randomly. By adjusting tumbling frequency, the cell is able to bias this random walk behavior and direct its motion toward or away from chemical gradients. Adaptation in this system refers to the invariance of tumbling frequency in environments with homogenous chemical gradients. This adaptation is robust to variations in the environment as well as the levels of chemotaxis regulators and the biochemical parameters of interactions between those

components.[78] This model and further experimental observations show that robustness in adaptation is a direct outcome of the organization of the network.

Recent work has highlighted the importance of robustness to changes in the internal environment of the cell. The presence of noise and the fundamental limits of deterministic behavior at the molecular level suggest that biological systems have evolved to cope with and exploit stochastic behavior in gene expression. Noise in gene expression is a ubiquitous feature of the natural world at the molecular scale and has been demonstrated to arise from the small numbers of molecules involved in the process.[79] Noise intrinsic to gene expression is thought to be dictated by fluctuations in mRNA levels, which may arise from fluctuations in promoter states or the random births and deaths of mRNAs themselves, and has also been shown to result from fluctuations in factors extrinsic to the genes themselves (including pathway specific and global factors of gene expression such as the levels of transcription factors, nucleic acid polymerases, and ribosomes[80]). Noise has been shown to be critical in several biological processes, including determination of competence in *B. subtilis*[81], eye color-vision development in *Drosophila melanogaster*[82], and in viral latency.[83] One future challenge is to demonstrate whether the control of noise is used by biological systems to adapt to specific environments.

Several organizational features of biological networks have been suggested to enable cells to cope with or exploit noise in their internal environment. Negative feedback loops have been demonstrated theoretically and experimentally to reduce noise in gene

expression by damping large fluctuations above the mean.[84] Recently, the Serrano group demonstrated that self-repression reduces noise compared to an unregulated gene, and was effective in reducing both intrinsic and extrinsic noise.[85] The organization of the genome could play a role in reducing gene expression noise as well. Swain showed that genes organized in operons strongly reduced variation between the genes via simulations and analytical derivations.[86] Noise attenuation by operons was effective even in the presence of multiple RBS. Later, experimental work by Tabor and Ellington at the University of Texas at Austin (personal communication) was able to confirm that the organization of genes in operons was an effective means to reduce variation between two genes. Nature, as well as engineers, may be able to utilize genomic organization of this kind to design systems with lower variability between two genes. The importance of these types of noise control with respect to biological function is a central question for future study.

Robustness to mutation

The robustness of the phenotype to underlying mutational change can be central to shaping the emergence of traits and phenotypes in evolution as well as disease, especially in the context of a "genetic capacitor". Mechanisms that confer robustness can act as capacitors for phenotypic change by masking the expression of genetic variation and suddenly revealing this variation when the robustness mechanism is impaired.[87, 88] Recent work in *Candida albicans* has demonstrated how the chaperone Hsp90 acts as a molecular buffering mechanism and can affect the evolution of drug resistance by

enabling novel mutations to have immediate physiological consequences. In addition, the effects of mutant Src kinases can be masked by Hsp90 in mammalian cells, leading to oncogenic phenotypes when Hsp90 is impaired. Thus, Hsp90 can act as a capacitor for phenotypic diversity from genetic variance.

Recent studies have suggested that buffering and genetic capacitance may be a general feature of complex regulatory networks[87]. Several patterns have emerged from studies of gene regulatory networks such as scale-free architecture, small-world structure, and the abundance of regulatory motifs – small "building-blocks" of larger networks. One central question for the engineering of biological systems is if these concepts of robustness can be extended into a set of design principles for specifying the degree of robustness to genetic and non-genetic change. Further work is needed to explore whether such systems-level features in biology are emergent properties of the organization of simpler units (devices and components). In addition, any systems-level property or design themes of regulatory architecture will undoubtedly be shaped by trade-offs between different functions. One example is the trade-off between robustness and fragility in biological and technological systems: the observation that a system robust to one class or type of perturbation is often sensitive or fragile with respect to another. The general occurrence of trade-offs between robustness and fragility [89] has been suggested in diverse complex systems such as the Internet,[90] the immune system,[91] and diseases such as metabolic syndrome [92] and cancer. [93] The "robust yet fragile" nature of biological systems may be a unifying force shaping the evolution of network architecture across the molecular, cellular, organismal, and ecological scales.

Robustness and power-law distributions: a cautionary tale

As discussed above, many biological networks display power-law distributions of connections per node. Albert et al.[94] found that networks with power-law distributions are robust to random perturbations: upon removal of random nodes, the average path length (the distance between any two nodes) changes very little. Power-law networks thus tend to retain overall topology by virtue of alternate paths between nodes, and act to minimize the cascading effects of node removal.[95] These and other theoretical examinations of power-law networks have led to the hypothesis that networks have evolved power-law distributions precisely because of this robustness to node removal and perturbation, and the ubiquity of power-law distributions in biological networks represents a deep and shared evolutionary pressure towards robust function.[95] However, several bodies of evidence contradict this intriguing hypothesis. An examination of networks in biological and physical systems and across spatial scales found that power-law structures exist in a wide spectrum of networks, including networks that have never been under any sort of natural or functional selection[96]. For example, the chemical reaction networks of planetary atmospheres (including Earth, Venus, and Jupiter) found power-law distributions in these systems. The power-law distribution may thus be a general feature of networks (both natural and human-engineered), much like the Gaussian distribution is a general feature of distributions in Nature. While this observation is fascinating, it explains little about the evolution of robustness or the adaptive significance of such global features of cellular network architecture. Thus, the construction and analysis of the

ecological and evolutionary consequences of regulatory mechanisms and architectures is a central area of study for systems and synthetic biology.

4. Open questions: from cellular to ecological networks

Understanding how regulatory architecture affects fitness and how network structure has adapted to cope with noise highlights central themes in systems biology. One current challenge is to understand population genetic parameters (such as fitness and diversity) in terms of cellular network architecture and dynamics, or how metabolic, regulatory, and molecular interaction networks combine to produce the phenotypes observed in Nature. A complementary goal is to understand the adaptive significance of network architecture and how genetic regulation in single cells may shape higher-order ecological interactions.

Open questions and outline of work

Several lines of inquiry at increasing levels of organization (molecular to ecological) will enable researchers to better engineer biology.

(1) How can we use the biochemical and informational functions of nucleic acids to regulate cellular circuits and networks?

(2) How do regulatory circuits regulate trade-offs between cellular functions? Can circuits be engineered to produce new optimal phenotypes, (for example, in terms of fitness in new environments)?

(3) How can ecology-scale interactions between organisms and their environment be understood in terms of regulatory networks? Are there systems-level design principles that suggest adaptation to given ecological strategies?

Progress towards answering these questions should provide insight into how adaptation has shaped the organization of genetic regulatory systems, and illuminate strategies for designing and constructing useful biological technologies as well. Ideally, this and other work will contribute to emerging themes for understanding the emergence of biological function in complex systems, from molecules to organisms.

5. References

1. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet 31, 64-8 (2002).
2. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci U S A 100, 11980-5 (2003).
3. Mangan, S., Zaslaver, A. & Alon, U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. J Mol Biol 334, 197-204 (2003).
4. Dueber, J. E., Yeh, B. J., Chak, K. & Lim, W. A. Reprogramming control of an allosteric signaling switch through modular recombination. Science 301, 1904-8 (2003).
5. Chin, Y. W., Balunas, M. J., Chai, H. B. & Kinghorn, A. D. Drug discovery from natural sources. Aaps J 8, E239-53 (2006).
6. Hughes, E. H. & Shanks, J. V. Metabolic engineering of plants for alkaloid production. Metab Eng 4, 41-8 (2002).
7. Ro, D. K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940-3 (2006).
8. Castagliuolo, I. et al. Engineered E. coli delivers therapeutic genes to the colonic mucosa. Gene Ther 12, 1070-8 (2005).
9. Critchley, R. J. et al. Potential therapeutic applications of recombinant, invasive E. coli. Gene Ther 11, 1224-33 (2004).
10. Anderson, J. C., Clarke, E. J., Arkin, A. P. & Voigt, C. A. Environmentally controlled invasion of cancer cells by engineered bacteria. J Mol Biol 355, 619-27 (2006).
11. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. Nature 434, 1130-4 (2005).
12. Levskaya, A. et al. Synthetic biology: engineering Escherichia coli to see light. Nature 438, 441-2 (2005).
13. Inoue, T., Sullivan, F. X. & Cech, T. R. Intermolecular exon ligation of the rRNA precursor of Tetrahymena: oligonucleotides can function as 5' exons. Cell 43, 431-7 (1985).
14. Tucker, B. J. & Breaker, R. R. Riboswitches as versatile gene control elements. Curr Opin Struct Biol 15, 342-8 (2005).
15. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419, 952-6 (2002).
16. Dykxhoorn, D. M., Novina, C. D. & Sharp, P. A. Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol 4, 457-67 (2003).
17. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet 39, 673-7 (2007).
18. Saghatelian, A., Yokobayashi, Y., Soltani, K. & Ghadiri, M. R. A chiroselective peptide replicator. Nature 409, 797-801 (2001).

19. Gesteland, R. F., Cech, T. & Atkins, J. F. The RNA world : the nature of modern RNA suggests a prebiotic RNA world (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2006).
20. Joyce, G. F. Forty years of in vitro evolution. Angew Chem Int Ed Engl 46, 6420-36 (2007).
21. Haruna, I. & Spiegelman, S. Autocatalytic synthesis of a viral RNA in vitro. Science 150, 884-6 (1965).
22. Kacian, D. L., Mills, D. R., Kramer, F. R. & Spiegelman, S. A replicating RNA molecule suitable for a detailed analysis of extracellular evolution and replication. Proc Natl Acad Sci U S A 69, 3038-42 (1972).
23. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818-22 (1990).
24. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. Science 287, 820-5 (2000).
25. Yan, A. C., Bell, K. M., Breeden, M. M. & Ellington, A. D. Aptamers: prospects in therapeutics and biomedicine. Front Biosci 10, 1802-27 (2005).
26. Ng, E. W. et al. Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. Nat Rev Drug Discov 5, 123-32 (2006).
27. Zimmermann, G. R., Wick, C. L., Shields, T. P., Jenison, R. D. & Pardi, A. Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. Rna 6, 659-67 (2000).
28. Lin, C. H. & Patel, D. J. Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP. Chem Biol 4, 817-32 (1997).
29. Nonin-Lecomte, S., Lin, C. H. & Patel, D. J. Additional hydrogen bonds and base-pair kinetics in the symmetrical AMP-DNA aptamer complex. Biophys J 81, 3422-31 (2001).
30. Jiang, L. & Patel, D. J. Solution structure of the tobramycin-RNA aptamer complex. Nat Struct Biol 5, 769-74 (1998).
31. Tao, J. & Frankel, A. D. Electrostatic interactions modulate the RNA-binding and transactivation specificities of the human immunodeficiency virus and simian immunodeficiency virus Tat proteins. Proc Natl Acad Sci U S A 90, 1571-5 (1993).
32. Puglisi, J. D., Chen, L., Frankel, A. D. & Williamson, J. R. Role of RNA structure in arginine recognition of TAR RNA. Proc Natl Acad Sci U S A 90, 3680-4 (1993).
33. Tan, R. & Frankel, A. D. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. Biochemistry 33, 14579-85 (1994).
34. Das, C., Edgcomb, S. P., Peteranderl, R., Chen, L. & Frankel, A. D. Evidence for conformational flexibility in the Tat-TAR recognition motif of cyclin T1. Virology 318, 306-17 (2004).
35. Bayer, T. S., Booth, L. N., Knudsen, S. M. & Ellington, A. D. Arginine-rich motifs present multiple interfaces for specific binding by RNA. Rna 11, 1848-57 (2005).

36. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428, 281-6 (2004).
37. Fedor, M. J. & Williamson, J. R. The catalytic diversity of RNAs. Nat Rev Mol Cell Biol 6, 399-412 (2005).
38. Robertson, M. P. & Ellington, A. D. Design and optimization of effector-activated ribozyme ligases. Nucleic Acids Res 28, 1751-9 (2000).
39. Breaker, R. R. & Joyce, G. F. A DNA enzyme that cleaves RNA. Chem Biol 1, 223-9 (1994).
40. Tsukiji, S., Pattnaik, S. B. & Suga, H. An alcohol dehydrogenase ribozyme. Nat Struct Biol 10, 713-7 (2003).
41. Doherty, E. A. & Doudna, J. A. Ribozyme structures and mechanisms. Annu Rev Biophys Biomol Struct 30, 457-75 (2001).
42. Robertson, M. P. & Ellington, A. D. In vitro selection of nucleoprotein enzymes. Nat Biotechnol 19, 650-5 (2001).
43. Robertson, M. P. & Scott, W. G. The structural basis of ribozyme-catalyzed RNA assembly. Science 315, 1549-53 (2007).
44. Rusconi, C. P. et al. RNA aptamers as reversible antagonists of coagulation factor IXa. Nature 419, 90-4 (2002).
45. Rusconi, C. P. et al. Antidote-mediated control of an anticoagulant aptamer in vivo. Nat Biotechnol 22, 1423-8 (2004).
46. Lupold, S. E., Hicke, B. J., Lin, Y. & Coffey, D. S. Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. Cancer Res 62, 4029-33 (2002).
47. Chu, T. C. et al. Aptamer:toxin conjugates that specifically target prostate tumor cells. Cancer Res 66, 5989-92 (2006).
48. Chu, T. C., Twu, K. Y., Ellington, A. D. & Levy, M. Aptamer mediated siRNA delivery. Nucleic Acids Res 34, e73 (2006).
49. McNamara, J. O., 2nd et al. Cell type-specific delivery of siRNAs with aptamer-siRNA chimeras. Nat Biotechnol 24, 1005-15 (2006).
50. Shangguan, D. et al. Aptamers evolved from live cells as effective molecular probes for cancer study. Proc Natl Acad Sci U S A 103, 11838-43 (2006).
51. Adleman, L. M. Molecular computation of solutions to combinatorial problems. Science 266, 1021-4 (1994).
52. Stojanovic, M. N. & Stefanovic, D. A deoxyribozyme-based molecular automaton. Nat Biotechnol 21, 1069-74 (2003).
53. Stojanovic, M. N., Mitchell, T. E. & Stefanovic, D. Deoxyribozyme-based logic gates. J Am Chem Soc 124, 3555-61 (2002).
54. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. Science 314, 1585-8 (2006).
55. Isaacs, F. J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22, 841-7 (2004).
56. Werstuck, G. & Green, M. R. Controlling gene expression in living cells through small molecule-RNA interactions. Science 282, 296-8 (1998).

57. Desai, S. K. & Gallivan, J. P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. J Am Chem Soc 126, 13247-54 (2004).
58. Topp, S. & Gallivan, J. P. Guiding bacteria with small molecules and RNA. J Am Chem Soc 129, 6807-11 (2007).
59. Levine, E., Zhang, Z., Kuhlman, T. & Hwa, T. Quantitative characteristics of gene regulation by small RNA. PLoS Biol 5, e229 (2007).
60. Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. Science 295, 1664-9 (2002).
61. El-Samad, H., Kurata, H., Doyle, J. C., Gross, C. A. & Khammash, M. Surviving heat shock: control strategies for robustness and performance. Proc Natl Acad Sci U S A 102, 2736-41 (2005).
62. Kurata, H. et al. Module-based analysis of robustness tradeoffs in the heat shock response system. PLoS Comput Biol 2, e59 (2006).
63. Ghalambor, C. K., Reznick, D. N. & Walker, J. A. Constraints on adaptive evolution: the functional trade-off between reproduction and fast-start swimming performance in the Trinidadian guppy (Poecilia reticulata). Am Nat 164, 38-50 (2004).
64. MacArthur, R. H. & Wilson, E. O. The theory of island biogeography (Princeton University Press, Princeton, N.J.,, 1967).
65. MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. Nature 441, 498-501 (2006).
66. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. Science 292, 504-7 (2001).
67. De Paepe, M. & Taddei, F. Viruses' Life History: Towards a Mechanistic Basis of a Trade-Off between Survival and Reproduction among Phages. PLoS Biol 4, e193 (2006).
68. Endy, D. Foundations for engineering biology. Nature 438, 449-53 (2005).
69. Dueber, J. E., Mirsky, E. A. & Lim, W. A. Engineering synthetic signaling proteins with ultrasensitive input/output control. Nat Biotechnol 25, 660-2 (2007).
70. Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2, 2006 0028 (2006).
71. Edwards, J. S., Ibarra, R. U. & Palsson, B. O. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 19, 125-30 (2001).
72. Fong, S. S. & Palsson, B. O. Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. Nat Genet 36, 1056-8 (2004).
73. Fischer, E. & Sauer, U. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. Nat Genet 37, 636-40 (2005).
74. Nowak, M. A. & Sigmund, K. Evolutionary dynamics of biological games. Science 303, 793-9 (2004).
75. Turner, P. E. & Chao, L. Prisoner's dilemma in an RNA virus. Nature 398, 441-3 (1999).

76. Waddington, C. H. & Robertson, E. Selection for developmental canalisation. Genet Res 7, 303-12 (1966).
77. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. Nature 387, 913-7 (1997).
78. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. Nature 397, 168-71 (1999).
79. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. Science 309, 2010-3 (2005).
80. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. Nat Genet 31, 69-73 (2002).
81. Suel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440, 545-50 (2006).
82. Samoilov, M. S., Price, G. & Arkin, A. P. From fluctuations to phenotypes: the physiology of noise. Sci STKE 2006, re17 (2006).
83. Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. & Schaffer, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122, 169-82 (2005).
84. Hooshangi, S. & Weiss, R. The effect of negative feedback on noise propagation in transcriptional gene networks. Chaos 16, 026108 (2006).
85. Dublanche, Y., Michalodimitrakis, K., Kummerer, N., Foglierini, M. & Serrano, L. Noise in transcription negative feedback loops: simulation and experimental analysis. Mol Syst Biol 2, 41 (2006).
86. Swain, P. S. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. J Mol Biol 344, 965-76 (2004).
87. Bergman, A. & Siegal, M. L. Evolutionary capacitance as a general feature of complex gene networks. Nature 424, 549-52 (2003).
88. Rutherford, S., Hirate, Y. & Swalla, B. J. The Hsp90 capacitor, developmental remodeling, and evolution: the robustness of gene networks and the curious evolvability of metamorphosis. Crit Rev Biochem Mol Biol 42, 355-72 (2007).
89. Doyle, J. & Csete, M. Rules of engagement. Nature 446, 860 (2007).
90. Doyle, J. C. et al. The "robust yet fragile" nature of the Internet. Proc Natl Acad Sci U S A 102, 14497-502 (2005).
91. Kitano, H. & Oda, K. Robustness trade-offs and host-microbial symbiosis in the immune system. Mol Syst Biol 2, 2006 0022 (2006).
92. Kitano, H. et al. Metabolic syndrome and robustness tradeoffs. Diabetes 53 Suppl 3, S6-S15 (2004).
93. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J., 3rd & Doyle, J. Robustness of cellular functions. Cell 118, 675-85 (2004).
94. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. Nature 406, 378-82 (2000).
95. Kim, B. J., Yoon, C. N., Han, S. K. & Jeong, H. Path finding strategies in scale-free networks. Phys Rev E Stat Nonlin Soft Matter Phys 65, 027103 (2002).
96. Wagner, A. Robustness and evolvability in living systems (Princeton University Press, Princeton, N.J., 2005).