eLS

Synthetic biology: molecular tools for engineering organisms

A20883

Author: Daniel C. Ducat[1,2]

Affiliations:

[1]MSU-DOE Plant Research Laboratory, Michigan State University, 612 Wilson Road, East Lansing, MI, USA; [2]Department of Biochemistry & Molecular Biology, Michigan State University, 603 Wilson Road, East Lansing, MI, USA.

***Advanced article**

<insert> **Abstract:**

Synthetic biology is a biological engineering discipline based on abstracting living systems through the lens of physical engineering concepts. In particular, synthetic biology places an emphasis on the characterization of simple parts that can be modularly assembled into configurations that give rise to complex, higher-order behaviours. Within the past 2 decades, this approach has enabled the development number of new molecular biology tools for modifying living systems in order to investigate fundamental processes or imbuing functions into cells that do not exist in nature. While specific synthetic biology applications span a huge range of seemingly unrelated disciplines (from biofuel producing microbes, to malaria-resistant mosquitos, to living medical therapies), these distinct examples derive from reuse and rearrangement of relatively limited set of cell engineering technologies. The continuing development of these core molecular biology tools for controlling gene expression, protein activity, and signalling networks can promote ever-more ambitious biological engineering projects.

<insert> **Key words**:

Synthetic biology, modularity, standardization, genetic circuit, gene expression, metabolic engineering, predictive engineering.

<insert> **Key Concepts:**

- *Synthetic biology is a discipline for biological engineering using principles of physical engineering*
- *Complex biological systems can be abstracted as a set of core "parts" with separable functions*
- *A biological part is modular when it confers a discrete function which can be repurposed and rearranged in many different contexts while maintaining fidelity*
- *Biological systems naturally display many modular features that are evident at many different scales*

- *Synthetic biology attempts to identify, characterize, and repurpose biological modules to build devices with novel functions*
- *The field has created several new molecular biology tools for controlling living systems through exploiting the recombination of modular parts*
- *The tools of synthetic biology are frequently used to imbue organisms with technologically useful traits*
- *As synthetic biology expands beyond model microbes, its potential to contribute to our fundamental understanding of biology continues to increase*

Introduction:

At its core, synthetic biology possesses distinct principles and philosophies that can be simplified as the attempt to bring electrical engineering concepts, such as standardization, modularity, and "bottom-up" design, to the biological sciences. While there are many differing opinions about what defines "synthetic biology" and how it contrasts with existing terms such as 'bioengineering' or 'metabolic engineering', a core conceit of the discipline is that biological systems can be abstracted as an network of modular biological "parts" (Cameron et al., 2014). Identifying biological modules that retain their function when repurposed into other systems, and learning design principles for effectively connecting parts to one another holds the promise to allow programing of cellular behaviours that are not found in nature. In this way, biological systems can be conceptualized as analogous to computer hardware that are composed of highly-defined modules (e.g. resistors, transistors), which can be rearranged into a myriad of complex circuits. Analogously, the capacity to predictively design increasingly ambitious biological systems relies upon the detailed characterization and standardization of these simplest components and understanding design principles for wiring inter-module connections.

Synthetic biology has developed rapidly within the last two decades through identification of an increasingly large number of separable modules that can be used to control the activity of genes and proteins within the cell. A biological part is robustly modular if it can be separated from the larger unit, then repurposed in a different context while fully retaining its characteristics. A classically-recognized example of a module is the promoter region of a gene (Figure 1A), which initiates transcription of downstream DNA under a defined set of environmental conditions, but which is routinely used to confer that expression profile on a new gene. Numerous other examples of natural modularity in biological systems can be found across many different scales (Figure 1B). At the same time that decreasing genome sequencing costs have provided a wealth of genome sequences and have highlighted the modular designs within natural systems (Bhattacharyya et al., 2006; Kashtan and Alon, 2005; Pawson and Nash, 2003; Ravasz, 2002; Wagner et al., 2007), decreases in DNA synthesis costs (Kosuri and Church, 2014) have enabled researchers to utilize a modular approach to the design of engineered pathways. Research within the field of synthetic biology often proceeds through the creation of libraries of biological parts, that can be efficiently recombined (Ellis et al., 2011) into a myriad of different configurations and screened for a desired output. This has allowed synthetic biologists to adopt strategies that are more analogous to physical engineering workflow of iterative rational design (i.e. cycles of "design-build-test-learn"; Figure 1C).

In this short review, we briefly discuss a number of molecular tools that have emerged from synthetic biology and are allowing cells to be controlled with increasing precision. We place special emphasis on examples that illustrate how modularity informs design of synthetic biology tools. While the range of applications for these tools is beyond the scope of this review, we highlight how modular assembly can generate complex systems from a limited set of relatively simple components.

Refining Control of Gene Expression

Refined methodologies for the control of gene expression underlie many of the more elaborate designs in synthetic biology. While control of heterologous gene expression has been a staple of all bioengineering efforts, recent efforts to more precisely standardize transcription and translation of gene targets have increased the precision, allowing increasing quantitative and predictive design of heterologous gene circuits. Therefore, some of the most fundamental "parts" of a synthetic biology toolkit are simply libraries of genetic elements that can drive expression of inserted genes at predictable levels and are standardized to one another across a broad tuneable range (Hammer et al., 2006). An illustrative example are the efforts to characterize the relationship between a given ribosome binding site (RBS) sequence and its kinetics of inducing ribosome binding and protein translation. This has enabled development of "RBS calculators" that both predict the affinity of a ribosome to a specific RBS sequence while also taking into account local secondary mRNA structure between the RBS and the neighbouring upstream 5' untranslated region (5'-UTR) and downstream coding sequence (Figure 2A; (Bonde et al., 2016; Espah Borujeni et al., 2014)). Such secondary structure effects can be difficult to completely predict, and illustrate biological limitations on modularity that may need to be overcome by suitable design principles for connecting component parts.

An excellent example of a library of defined promoter/RBS elements developed for *E. coli* utilized a design strategy to connect promoter-RBS elements that greatly reduced the impact of local genomic context (Figure 2B). As often noted for a given promoter or RBS element, the authors first observed that distinct genes placed under identical promoter-RBS combinations led to unpredictable gene expression levels (Mutalik et al., 2013). The majority of this variability could be attributed to sequence-dependent secondary mRNA structure, for example, the formation of hairpins that block ribosome access (Figure 2A). To circumvent this problem, the authors designed "bicistronic" operons with an invariant upstream coding sequence that could largely eliminate secondary RNA folding effects (Mutalik et al., 2013). Ribosome translocation through the upstream element unfolds the secondary structure of the RNA near the internal RBS, allowing for much greater prediction of downstream genes (Figure 2B). Similar efforts have been directed towards the characterization of the effectiveness of DNA terminator elements to quantify their capacity to insulate a genetic element from the genomic context in which is inserted (Cambray et al., 2013). Collectively, the goal of such studies is to allow *in silico* assembly of a regulatory region that will express a gene of interest (GOI) at a precise and predictable level.

Construction of synthetic transcription factors that recognize user-specified DNA sequences is an alternative approach that has helped to broaden the options available for controlling gene expression. Transcription activator-like effector (TALE) domains are

one promising class of protein sequences that can be used in the construction of designer transcription factors due to their ability to be constructed to recognize and bind to any target DNA sequence. TALE domains are naturally found in clusters containing multiple repeats of the 33-35 amino acid (aa) sequence, each of which can vary in two specific residues that dictate recognition of an A, T, C, or G nucleotide (Figure 3A). A string of these domains in series defines a particular DNA sequence to which the TALE repeat will bind. Cys2-His2 zinc-finger (ZF) domains represent a similar opportunity for encoding protein targeting to a custom DNA sequence, as each ZF domain possesses affinity to a particular 3 nucleotide triplet (reviewed in (Gaj et al., 2013)). Varying the amino acids that contact the base pairs within the major groove has allowed for the design of domains that have specificity for many of the 64 possible triplet combinations, albeit with different degrees of specificity (Kim et al., 2011).

In their natural context, TALE and ZF domains are found as the DNA-binding region within transcription factors, while other domains of the protein specify activities related to transcriptional activation or repression. For example, plant pathogens in the *Xanthomonas* genus inject transcription factors into host plant cells through type III secretion that are guided to bind to promoter regions in the plant nucleus through TALE repeat regions. A separate modular domain acts as a strong activator of transcription, ultimately leading to upregulation of genes that enhance the fitness of the pathogen (e.g., sugar transporter genes that increase the extracellular carbohydrate availability). The same construction strategy can be used to append different functional domains onto custom TALE or ZF proteins, creating proteins that bind to specific sequences and either repress or activate gene expression (Gaj et al., 2013). For instance, adding a VP16 domain will create a targeted transcription factor that activates gene expression in mammalian cells while a KRAB domain will conversely repress nearby genes. When synthetic TALE or ZF transcription repressor proteins are used in combination with post-transcriptional repressors (e.g. RNA interference; RNAi) the result can often be nearly complete gene repression.

An analogous engineering tool that has received a great deal of attention recently is the nuclease-containing CRISPR (clustered regularly interspaced short palindromic repeat) system. This refers to a natural anti-viral system evolved in bacteria to recognize foreign genetic material and to target a nuclease activity specifically to these sequences. The CRISPR system involves a nuclease protein, Cas9, which binds to specific small non-coding guide RNAs (ngRNA) that acts to direct Cas9's activity to specific genetic sequences by base pairing (Figure 3B). As the guide RNA is encoded separately from Cas9, engineers can readily encode custom guide RNAs to target Cas9 to bind to virtually any sequence. As Cas9 has endogenous nuclease activity, the capacity to direct it to bind to virtually any target sequence has greatly assisted genomic editing efforts by allowing DNA breaks to be introduced at precise locations. Both TALE and ZF-domain containing proteins have also been used widely for genome editing by appending modular domains with nuclease activity onto TALE or ZF proteins. Directed cleaving of DNA *in vivo* greatly increases the frequency of homologous recombination at that site, allowing for targeted gene disruption or insertions. Double-strand breaks (e.g. achieved by appending Fok1 nuclease domain) frequently lead to inaccurate repair, making this approach useful for creating indels and disrupting gene function, while single-stranded breaks can allow for seamless integration of alternative DNA sequence through homologous recombination

(Gaj et al., 2013). The core features of CRISPR, TALE-, and ZF-containing synthetic transcription factors are similar – they allow creation of custom proteins that bind user-specified nucleotide sequences and can then be modified to locally activate/repress transcription, block translation, or create DNA breaks to introduce targeted genetic modifications. For a detailed review of the comparative strengths and weaknesses of genome editing and control of gene expression via CRISPR, TALEs and ZFs, see (Gaj et al., 2013).

Post-translational Control of Protein Activities in Space and Time

In addition to methods to influence the activation and translation of target genes, tools have been developed that allow manipulation of the levels and activities of resultant proteins. Here again, much of this work relies upon the inherent modular organization of natural proteins and their regulatory sequences. For example, SH2 domains generally confer binding affinity to the proline-rich sequence PxxP, and is a domain that is repeated across 110 distinct proteins in humans alone (Liu et al., 2006). A particular combination of modular interaction domains that each provide an incremental binding affinity can cooperatively create a protein-protein interface of high specificity (Pawson and Nash, 2003).

The ability to control the binding of properties of proteins can be used to improve the performance of heterologous pathways by emulating the natural cellular strategy of segregating factors into distinct micro-domains or compartments in order to increase local concentration while reducing deleterious cross-talk (Figure 3C). For example, eukaryotic cells improve pathway efficiency by compartmentalizing reactions within organelles, effectively concentrating components of related pathways, increasing total flux, and insulating external pathways from unwanted cross-talk or toxic intermediates. Similar benefits are routinely achieved in metabolic and signalling pathways by co-localization to a common micro-domain within the cell, or upon a shared scaffolding surface. For example, elegant studies of yeast mitogen activated protein kinase (MAPK) pathways have demonstrated that scaffold proteins (e.g. Ste5) recruit and concentrate kinases from a related signalling cascade in order to increase fidelity, decrease crosstalk, increase reaction speed, and preserve spatial elements of a signal (Figure 1B; (Gordley et al., 2016)).

In an effort to capture similar benefits for heterologous metabolic pathways and reconstructed signalling systems, engineers have designed a variety of synthetic scaffolds with the ultimate goal of creating a programmable subcellular surface *in vivo*. One of the earliest synthetic scaffold designs was simply a string of protein-protein interaction domains all encoded on the same peptide (Figure 3C). By modifying enzymes of a heterologous mevalonate-production pathway so that they contained the corresponding ligand domains, the authors aimed to concentrate the pathway and metabolic intermediates to a subcellular domain within the cytosol. This strategy appeared to increase the output of a heterologous melvalonate production pathway by up to ~100 fold in one instance (Dueber et al., 2009). Yet, unpredictability in the actual structure formed by these simple synthetic scaffolds can limit their utility for other applications. Since then, a variety of biological scaffolds have been engineered, using

RNA, DNA, and polymerizing proteins, all of which use a similar premise of attaching modular binding domains that allow target proteins to be recruited to the scaffold structure (Siu et al., 2015). Yet, despite numerous improvements and examples of increasing efficacy of these designs, a truly predictable self-assembling scaffold that can be modified to form a specific desired architecture *in vivo* remains an unfulfilled goal of the field (Young et al., 2017).

Just as important as transcription and translation, degradation rates control the steady state level of any cellular protein – providing another potential layer to modulate the activity of engineered systems. One successful approach has been to utilize regulatory peptide sequences that are conditionally recognized by cellular machinery that direct targets to the proteasome. For example, the ssrA degradation pathway acts as a control system for damaged mRNA in bacteria by appending a 13 aa sequence onto the C-terminus of any polypeptide associated with a stalled ribosome. This 13 aa signal sequence is recognized by an adapter protein, SspB, which also interacts with the proteosomal complex ClpXP, thereby targeting these suspect proteins to be degraded (Figure 3D). Detailed mechanistic understanding of ClpXP-mediated protein degradation (Baker and Sauer, 2012) laid the foundation for engineers to co-opt it as a mechanism to inducibly downregulate protein levels. By encoding the ssrA tag directly within gene sequence of a heterologous construct, the protein can be degraded when expression of SspB is induced (Figure 3D). This effectively allows for inducible conditional knockouts for any protein that can be modified at the C-terminus, and has been used to downregulate essential genes for the purpose of metabolic engineering (Brockman and Prather, 2015) and to study the function of essential cell components for which genomic knockouts cannot be generated (Ricci et al., 2016). Similar strategies have been utilized with alternative proteosomal machinery, including the Lon protease (Cameron and Collins, 2014).

Another powerful class of molecular biology tools that can be used to precisely control the activity of proteins within a cell in both spatial and temporal dimensions are light-responsive proteins termed optogenetic switches. Some of the earliest examples of such devices were derived from the repurposing of microbial channelrhodopsins (Boyden et al., 2005), but more devices have relied upon modular LOV (light-oxygen-voltage) domains and portions of phytochromes and crytochromes. Key features of optagenetic devices are that light can both induce a rapid conformational change and also that the activated state can be reversed on short time scales by light of different wavelengths (e.g. phytochromes) or within seconds to minutes in the dark (e.g. cryptochromes). Domains that undergo such conformational state switching can be appended onto other functional proteins (e.g. enzymes, transcription factors) to control their function in a light-dependent fashion. This allows a researcher extreme precision to control activation of an optagenetically-controlled pathway both temporally and spatially. A variety of different optogenetic switches have been developed that allow control over gene expression, protein localization, control of ion transport, and activation of signalling cascades at defined subcellular locations (see (Pastrana, 2011) for a non-technical overview, and (Olson and Tabor, 2014; Shcherbakova et al., 2015) for more detailed reviews of mechanisms and applications).

"Bottom Up": Emergent Complexity from Interactive Components

Individual components, such as defined promoters or well-characterized modular protein-protein interaction domains, are useful tools on their own, but the promise of synthetic biology lies in the capacity to combine individual "parts" in a predictive manner to assemble complex systems with customizable functions. A commonly used synthetic biology component for controlling gene expression is a genetic logic gate, which is conceptualized in relation to simple computational operations. For example, an AND gate is only active if both inputs A and B are TRUE, while an OR gate is active if either A or B, or both, are TRUE (Figure 4A). Sixteen genetic logic gates are possible for a two-input function, and all of these been encoded genetically (Siuti et al., 2013), often via multiple independent approaches. Conceptualizing genetic circuits like logic gates inspired some of the earliest synthetic biology devices, such as the Repressilator and Toggle switch (Collins et al., 2000; Elowitz and Leibler, 2000). Yet the real utility of such genetic units is highlighted when multiple individual modules are connected in series with one another, enabling dramatically more complex behaviour to be encoded within the cell (Brophy and Voigt, 2014; Moon et al., 2012) (Figure 4B).

Increasingly, logic circuits are being used to drive designer expression of key genes towards theraputic or biotechnology goals. For example, a recent report created a probiotic *E. coli* strain with the capability to provide a readout for rapid detection of gut inflammation and demonstrate proof-of-concept within a mouse model. The design involved linking a novel histidine kinase receptor specific for the detection of thiosulfate (a marker of gut inflammation) to a synthetic AND gate, which drove the expression of a readout reporter only when the circuit was induced, and the cells experienced the environment of an inflamed mouse gut (Daeffler et al., 2017). A similar strategy utilized a circuit that responded to a chemical agent (anhydrotetracycline), and triggered a feedback loop, allowing bacteria within a gut to "remember" and report the chemical exposure long after the priming event (Kotula et al., 2014). Both of these studies illustrated how simple genetic circuits can be recombined to make bacterial strains that could reside within the digestive tract and act as "sentinels" to report that the system has been exposed to a specific insult and/or become imbalanced.

Akin to connecting logic gates within a single cell to construct more complicated devices, simple circuits that reside in separate cells of a larger population can be connected to one another to create complex higher-order population behaviours. Some of the best-known early examples of these intercellular circuits took advantage of a class diffusible signalling molecules called acyl-homoserine lactones (AHL), derived from bacterial quorum sensing pathways. AHLs are generated with a cell through the action of acyl-homoserine-synthases, and subsequently can diffuse through the plasma membrane and into neighbouring cells (Waters and Bassler, 2005). When an AHL binds into the hydrophobic core of a cognate AHL receptor, it stabilizes the protein, which can bind to promoter elements to regulate gene expression. When connected to one another across different cells by intercellular signals, even a simple circuits can produce startlingly complex behaviours at the population-level, including coordinated "blinking" across the community, or pulse-like waves of gene activation and repression (Danino et al., 2010). Connecting circuits across individual cells has been compared to distributed computation by parallel processors and one advantage of this approach is that it can enable a large design space with a relatively limited number of unique "parts," because

the same components can be reused in distinct cell types within the consortia (Regot et al., 2011).

AHLs are only one class of molecule that can be exchanged as a signal between cells, other efforts to program modular microbial communities take advantage of cross-feeding between distinct species to more effectively distribute metabolic labour across consortia partners. While microbes can be designed to exchange any number of metabolic intermediates (e.g. much research has been conducted on closely-related species engineered to have complimentary auxotrophies (Mee and Wang, 2012)), exchange of key metabolites can allow for the generation of communities composed of two or more "metabolic specialist species" that compartmentalize highly different reactions. In one recent example, two microbial partners were engineered to grow together on pretreated cellulosic material (corn stover) by compartmentalizing the reaction for breaking cellulose down to soluble oligosaccharides within *Trichoderma reesei*, while an engineered partner species, *E. coli*, consumed the released sugars to produce isobutanol (Minty et al., 2013).

As with many of the examples above, synthetic consortia also can serve as an illustration of the limitations of the abstraction of modularity in biological systems. In the previous example, it is notable that the genes encoding the cellulases contained of *T. reesei* are theoretically transferrable parts that, if expressed in *E. coli*, could be used to generate a single species with the metabolic capability of the consortia. In this instance, knowledge limitations - about cellulase maturation, export, and organization upon extracellular scaffolding complexes (i.e. cellulosomes) – can complicate our capacity to engineer *E. coli* for efficient cellulose degradation. In other instances, incompatibilities between metabolic processes may hinder integration of two desirable pathways into one host organism. For example, the complex metallocenters of nitrogenases that fix atmospheric nitrogen are notoriously oxygen sensitive and require numerous maturation factors. Aerobic species must often go to extreme lengths to maintain active nitrogenases; confining nitrogenase activity to specialized compartments, time periods or differentiated cell types (Compaoré and Stal, 2010; Flores and Herrero, 2010). Although ambitious efforts exist to express nitrogenases in heterologous hosts, many of these have been met with limited success in part because of the complexity of the pathways and metabolic incompatibilities.

Alternatively, there are examples that utilize modular consortia to compartmentalize metabolic abilities across a consortium that would be difficult to program within a single host. For example, *Azotobacter vinelandii* is a nitrogen-fixing bacterium that has been modified to secrete ammonium by knocking out nifL, a transcription factor that represses nitrogenase activity under nitrogen-replete conditions. Ammonium-secreting *A. vinelandii* has been shown to grow in co-culture with select alga species and plants (Ambrosio et al., 2017; Ortiz-Marquez et al., 2013), enhancing available nitrogen resources for the autotroph by effectively bootstrapping nitrogenase activity onto photosynthetic organisms. By contrast, retaining an active heterologous nitrogenase within the photosynthetic species would be complicated by the high oxygen partial pressure caused by the oxygen evolving activity of photosystem II. In a conceptually similar vein, the cyanobacterium *Synechococcus elongatus PCC 7942* has been modified to efficiently fix carbon and secrete it in the form of sucrose, and can therefore be used as a photosynthetic module to construct a variety of light-driven

autotroph/heterotroph consortia (Figure 4C; (Hays et al., 2017)). In this case, the heterotrophic species within the consortia can act as a conversion module to transform the sugar to higher-value compounds (Figure 4C). Eight distinct co-cultures have been published with this modular autotroph/heterotroph design, including; utilizing *Pseudomonas putida* to create the bioplastic poly-hydroxybutyrate (Löwe et al., 2017), *Rhodotorula glutinis* for the generation of long-chain fatty acids (Li et al., 2017), and *A. vinelandii* to create an artificial carbon-for-nitrogen symbiotic exchange (Smith and Francis, 2016). In some instances, productivity of the synthetic consortia exceeded that of attempts to rewire the metabolism of cyanobacteria for direct photoproduction of the same target compound (Weiss and Ducat, 2017), illustrating the potential benefit of metabolic specialization and compartmentalization.

Expanding the Synthetic Biology Toolkit Beyond Model Organisms

In order for synthetic biology to realize its broadest potential it must continue to escape the confines of the best-studied model microbes and apply the foundational principles and philosophy of the discipline towards building reliable parts lists within alternative species. Many of the core themes of synthetic biology have permeated into other disciplines, and specific tools have been adapted for use in a number of organisms. Yet, it could be argued that most model systems outside of *E. coli* and *S. cerevisiae* lack a core set of well-defined biological parts that have been characterized in a standardized manner or been rigorously analysed for their degree of modularity. Because complex synthetic pathways and circuits are built in a manner that is dependent upon robust and predictable functioning of simple elements (Figure 4), the lack of core components (e.g. even small libraries of promoter elements with standardized activities) can hinder the translation of the most ambitious synthetic biology applications into other species.

It is frequently said that deep knowledge of a topic is further enhanced when one must teach that knowledge to another; synthetic biology offers a similar promise of "learning-by-building". Although the emphasis of many discussions on the potential of synthetic biology relate to therapeutic, energy, environmental, or other biotechnological applications, perhaps one of the most impactful aspects of the use of synthetic biology will be in furthering our fundamental knowledge of the organization and evolution of biological systems (Bashor et al., 2010). The tools of synthetic biology allow a researcher to manipulate organisms with increasing precision, while the framework enables the systematic investigation of core assumptions about the network organization of living systems by building analogous, simplified genetic devices. Put differently, much of our present day knowledge of biology comes from a scientific tradition of "learning by breaking" – for example the creation and study of mutants. Synthetic biology offers an alternative approach to test and expand our knowledge through the construction of new systems that will only function as intended if our underlying assumptions are relatively accurate. Recognition of the enabling potential of establishing foundational modular "parts lists" across a wide swath of organisms will greatly assist in accelerating the impact of synthetic biology, both for futuristic technological applications, and deepening our fundamental understanding of the organization of life.

Acknowledgements:

This work was supported by the Office of Science of the U.S. Department of Energy DE-FG02-91ER2002, as well as NSF Grant CBET #1437657; which also supported some of the primary research discussed in this review.

References:

Ambrosio, R., Ortiz-Marquez, J.C.F., Curatti, L., 2017. Metabolic engineering of a diazotrophic bacterium improves ammonium release and biofertilization of plants and microalgae. Metab. Eng. 40, 59–68. doi:10.1016/j.ymben.2017.01.002

Baker, T.A., Sauer, R.T., 2012. ClpXP, an ATP-powered unfolding and protein-degradation machine. Biochim. Biophys. Acta - Mol. Cell Res. doi:10.1016/j.bbamcr.2011.06.007

Bashor, C.J., Horwitz, A.A., Peisajovich, S.G., Lim, W.A., 2010. Rewiring Cells: Synthetic Biology as a Tool to Interrogate the Organizational Principles of Living Systems. Annu. Rev. Biophys. 39, 515–537. doi:10.1146/annurev.biophys.050708.133652

Bhattacharyya, R.P., Reményi, A., Yeh, B.J., Lim, W.A., 2006. Domains, Motifs, and Scaffolds: The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits. Annu. Rev. Biochem. 75, 655–680. doi:10.1146/annurev.biochem.75.103004.142710

Bonde, M.T., Pedersen, M., Klausen, M.S., Jensen, S.I., Wulff, T., Harrison, S., Nielsen, A.T., Herrgård, M.J., Sommer, M.O.A., 2016. Predictable tuning of protein expression in bacteria. Nat. Methods 13, 233–236. doi:10.1038/nmeth.3727

Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., Deisseroth, K., 2005. Millisecond-timescale, genetically targeted optical control of neural activity. Nat. Neurosci. 8, 1263–1268. doi:10.1038/nn1525

Brockman, I.M., Prather, K.L.J., 2015. Dynamic metabolic engineering: New strategies for developing responsive cell factories. Biotechnol. J. 10, 1360–1369. doi:10.1002/biot.201400422

Brophy, J.A.N., Voigt, C.A., 2014. Principles of genetic circuit design. Nat. Methods 11, 508–520. doi:10.1038/nmeth.2926

Cambray, G., Guimaraes, J.C., Mutalik, V.K., Lam, C., Mai, Q.A., Thimmaiah, T., Carothers, J.M., Arkin, A.P., Endy, D., 2013. Measurement and modeling of intrinsic transcription terminators. Nucleic Acids Res. 41, 5139–5148. doi:10.1093/nar/gkt163

Cameron, D.E., Bashor, C.J., Collins, J.J., 2014. A brief history of synthetic biology. Nat. Rev. Microbiol. 12, 381–390. doi:10.1038/nrmicro3239

Cameron, D.E., Collins, J.J., 2014. Tunable protein degradation in bacteria. Nat. Biotechnol. 32, 1276–1281. doi:10.1038/nbt.3053

Collins, J.J., Gardner, T.S., Cantor, C.R., 2000. Construction of a genetic toggle switch in Escherichia coli. Nature 403, 339–342. doi:10.1038/35002131

Compaoré, J., Stal, L.J., 2010. Oxygen and the light-dark cycle of nitrogenase activity in two unicellular cyanobacteria. Environ. Microbiol. 12, 54–62. doi:10.1111/j.1462-2920.2009.02034.x

Daeffler, K.N., Galley, J.D., Sheth, R.U., Ortiz-Velez, L.C., Bibb, C.O., Shroyer, N.F., Britton, R.A., Tabor, J.J., 2017. Engineering bacterial thiosulfate and tetrathionate sensors for detecting gut inflammation. Mol. Syst. Biol. 13, 923. doi:10.15252/msb.20167416

Danino, T., Mondragón-Palomino, O., Tsimring, L., Hasty, J., 2010. A synchronized

quorum of genetic clocks. Nature 463, 326–330. doi:10.1038/nature08753

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A. V, Prather, K.L.J., Keasling, J.D., 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotechnol. 27, 753–759. doi:10.1038/nbt.1557

Ellis, T., Adie, T., Baldwin, G.S., 2011. DNA assembly for synthetic biology: from parts to pathways and beyond. Integr. Biol. 3, 109. doi:10.1039/c0ib00070a

Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403, 335–338. doi:10.1038/35002125

Espah Borujeni, A., Channarasappa, A.S., Salis, H.M., 2014. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res. 42, 2646–2659. doi:10.1093/nar/gkt1139

Flores, E., Herrero, A., 2010. Compartmentalized function through cell differentiation in filamentous cyanobacteria. Nat. Rev. Microbiol. 8, 39–50. doi:10.1038/nrmicro2242

Gaj, T., Gersbach, C.A., Barbas, C.F., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. doi:10.1016/j.tibtech.2013.04.004

Gordley, R.M., Bugaj, L.J., Lim, W.A., 2016. Modular engineering of cellular signaling proteins and networks. Curr. Opin. Struct. Biol. doi:10.1016/j.sbi.2016.06.012

Hammer, K., Mijakovic, I., Jensen, P.R., 2006. Synthetic promoter libraries--tuning of gene expression. Trends Biotechnol. 24, 53–5. doi:10.1016/j.tibtech.2005.12.003

Hays, S.G., Yan, L.L.W., Silver, P.A., Ducat, D.C., 2017. Synthetic photosynthetic consortia define interactions leading to robustness and photoproduction. J. Biol. Eng. 11, 4. doi:10.1186/s13036-017-0048-5

Kashtan, N., Alon, U., 2005. Spontaneous evolution of modularity and network motifs. Proc. Natl. Acad. Sci. 102, 13773–13778. doi:10.1073/pnas.0503610102

Kim, S., Lee, M.J., Kim, H., Kang, M., Kim, J.-S., 2011. Preassembled zinc-finger arrays for rapid construction of ZFNs. Nat. Methods 8, 7–7. doi:10.1038/nmeth0111-7a

Kosuri, S., Church, G.M., 2014. Large-scale de novo DNA synthesis: technologies and applications. Nat. Methods 11, 499–507. doi:10.1038/nmeth.2918

Kotula, J.W., Kerns, S.J., Shaket, L.A., Siraj, L., Collins, J.J., Way, J.C., Silver, P.A., 2014. Programmable bacteria detect and record an environmental signal in the mammalian gut. Proc. Natl. Acad. Sci. 111, 4838–4843. doi:10.1073/pnas.1321321111

Li, T., Li, C.-T., Butler, K., Hays, S.G., Guarnieri, M.T., Oyler, G.A., Betenbaugh, M.J., 2017. Mimicking lichens: incorporation of yeast strains together with sucrose-secreting cyanobacteria improves survival, growth, ROS removal, and lipid production in a stable mutualistic co-culture production platform. Biotechnol. Biofuels 10, 55. doi:10.1186/s13068-017-0736-x

Liu, B.A., Jablonowski, K., Raina, M., Arcé, M., Pawson, T., Nash, P.D., 2006. The Human and Mouse Complement of SH2 Domain Proteins-Establishing the Boundaries of Phosphotyrosine Signaling. Mol. Cell 22, 851–868. doi:10.1016/j.molcel.2006.06.001

Löwe, H., Hobmeier, K., Moos, M., Kremling, A., Pflüger-Grau, K., 2017. Photoautotrophic production of polyhydroxyalkanoates in a synthetic mixed culture of Synechococcus elongatus cscB and Pseudomonas putida cscAB. Biotechnol. Biofuels 10, 190. doi:10.1186/s13068-017-0875-0

Mee, M.T., Wang, H.H., 2012. Engineering ecosystems and synthetic ecologies. Mol. Biosyst. 8, 2470. doi:10.1039/c2mb25133g

Minty, J.J., Singer, M.E., Scholz, S.A., Bae, C.-H., Ahn, J.-H., Foster, C.E., Liao, J.C., Lin, X.N., 2013. Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proc. Natl. Acad. Sci. 110, 14592–14597. doi:10.1073/pnas.1218447110

Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C., Voigt, C.A., 2012. Genetic programs constructed from layered logic gates in single cells. Nature 491, 249–253. doi:10.1038/nature11516

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.-A., Tran, A.B., Paull, M., Keasling, J.D., Arkin, A.P., Endy, D., 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. Nat. Methods 10, 354–360. doi:10.1038/nmeth.2404

Olson, E.J., Tabor, J.J., 2014. Optogenetic characterization methods overcome key challenges in synthetic and systems biology. Nat. Chem. Biol. 10, 502–511. doi:10.1038/nchembio.1559

Ortiz-Marquez, J.C.F., Do Nascimento, M., Zehr, J.P., Curatti, L., 2013. Genetic engineering of multispecies microbial cell factories as an alternative for bioenergy production. Trends Biotechnol. 31, 521–529. doi:10.1016/j.tibtech.2013.05.006

Pastrana, E., 2011. Optogenetics: controlling cell function with light. Nat. Methods 8, 24–25. doi:10.1038/nmeth.f.323

Pawson, T., Nash, P., 2003. Assembly of cell regulatory systems through protein interaction domains. Science 300, 445–452. doi:10.1126/science.1083653

Ravasz, E., 2002. Hierarchical Organization of Modularity in Metabolic Networks. Science (80-.). 297, 1551–1555. doi:10.1126/science.1073374

Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellén, J., Peeters, T., Hohmann, S., de Nadal, E., Posas, F., Solé, R., 2011. Distributed biological computation with multicellular engineered networks. Nature 469, 207–211. doi:10.1038/nature09679

Ricci, D.P., Melfi, M.D., Lasker, K., Dill, D.L., McAdams, H.H., Shapiro, L., 2016. Cell cycle progression in Caulobacter requires a nucleoid-associated protein with high AT sequence recognition. Proc. Natl. Acad. Sci. U. S. A. 113, E5952–E5961. doi:10.1073/pnas.1612579113

Shcherbakova, D.M., Shemetov, A.A., Kaberniuk, A.A., Verkhusha, V. V., 2015. Natural Photoreceptors as a Source of Fluorescent Proteins, Biosensors, and Optogenetic Tools. Annu. Rev. Biochem. 84, 519–550. doi:10.1146/annurev-biochem-060614-034411

Siu, K.H., Chen, R.P., Sun, Q., Chen, L., Tsai, S.L., Chen, W., 2015. Synthetic scaffolds for pathway enhancement. Curr. Opin. Biotechnol. doi:10.1016/j.copbio.2015.08.009

Siuti, P., Yazbek, J., Lu, T.K., 2013. Synthetic circuits integrating logic and memory in

living cells. Nat. Biotechnol. 31, 448–452. doi:10.1038/nbt.2510

Smith, M.J., Francis, M.B., 2016. A Designed A. vinelandii-S. elongatus Coculture for Chemical Photoproduction from Air, Water, Phosphate, and Trace Metals. ACS Synth. Biol. 5, 955–961. doi:10.1021/acssynbio.6b00107

Wagner, G.P., Pavlicev, M., Cheverud, J.M., 2007. The road to modularity. Nat. Rev. Genet. 8, 921–931. doi:10.1038/nrg2267

Waters, C.M., Bassler, B.L., 2005. Quorum Sensing : Communication in Bacteria. Annu. Rev. Cell Dev. Biol. 21, 319–346. doi:10.1146/annurev.cellbio.21.012704.131001

Weiss, T.L., Ducat, D.C., 2017. Designing stable, synthetic, light-driven cyanobacteria-heterotroph consortia for bioproduction. Metabolic Engineering. doi: 10.1016/j.ymben.2017.10.009

Young, E.J., Burton, R., Mahalik, J.P., Sumpter, B.G., Fuentes-Cabrera, M., Kerfeld, C.A., Ducat, D.C., 2017. Engineering the bacterial microcompartment domain for molecular scaffolding applications. Front. Microbiol. 8. doi:10.3389/fmicb.2017.01441

<insert> **Further Reading**:

Kirschner, M. W., & Gerhart, J. C. (2006). *The plausibility of life: Resolving Darwin's dilemma*. Yale University Press.

Church, G. M., & Regis, E. (2014). *Regenesis: how synthetic biology will reinvent nature and ourselves*. Basic Books.

Lim, W., Mayer, B., & Pawson, T. (2014). *Cell signaling: principles and mechanisms*. Taylor & Francis.

Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metabolic engineering*, *14*(3), 189-195.

Purnick, P. E., & Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nature reviews Molecular cell biology*, *10*(6), 410-422.

Lazebnik, Y. (2002). Can a biologist fix a radio?—Or, what I learned while studying apoptosis. *Cancer cell*, *2*(3), 179-182.

<insert> **Glossary:**

Part: a unit of biology that is a smaller component of a larger system and has separable functions (e.g. protein domain, gene promoter).

Module: a functional unit (part) of biological systems that retains its intrinsic properties irrespective of the context it is placed and what other units it is connected to.

Circuit: a collection of modules connected to one another such that the network is capable of a programmed higher-order function. For example, a biological "logic circuit" that can detect 2 input stimuli and activate (or repress) an output gene according to a defined standard (see Figure 4A).

<insert> **Figures and Tables:**
All figures within this manuscript are original and do not require permission for use.

Figure 1:

Abstraction of biology as composed of modular subunits. A core tenant of synthetic biology is the conceptualization of biological systems as composed of many, relatively simple interconnected parts that can be recombined in a modular fashion. **A)** A gene coding region is readily recognized as containing modular features, including the promoter, ribosome binding site (RBS), coding sequence, and terminator. If these "parts" are to be used in a truly modular fashion, it should be possible to repurpose them in a different context, yet retain their core function. Promoter "i" (dark green – bottom left) represents a highly modular promoter element because it drives transcription of the downstream sequence in a highly predictable fashion, no matter what the sequence is. Promoter "ii" (light green) displays variable properties depending on the context, and therefore possesses poor modularity. **B)** Modularity in the design of living systems can be found across many scales. This includes protein domains, which often contain homologous sequence to domains in other proteins. The SH2 (Src-homology domain 2) is a domain with the self-contained property of binding amino acid sequences P-X-X-P, and is a domain naturally found widely across many proteins in eukaryotes. Whole proteins are often modular and can be exported from one organism to another while retaining their function, or even repurposed in different contexts in the same cell for different functions. Here, the MAPKKK Ste11 is at the top of the kinase cascade for signalling responses to both mating factor and osmotic shock in yeast. The function of Ste11 (grey box) remains the same while the context (i.e. which scaffold protein it is associated with; either Ste5 or Pbs2) has important implications for its output. At larger scales, examples of modularity of tissue types or whole organs can be found (e.g. the capacity to transplant hearts across a relatively large evolutionary space). Yet, at each biological scale, examples of poor modularity also exist (e.g. cannot transplant brain tissue from even closely related species). This highlights the necessity of utilizing a process **C)** to characterize biological parts in a standardisable way for their functionality, thereby identifying valuable parts and design principles that facilitate biological engineering using a modular approach.

Figure 2:

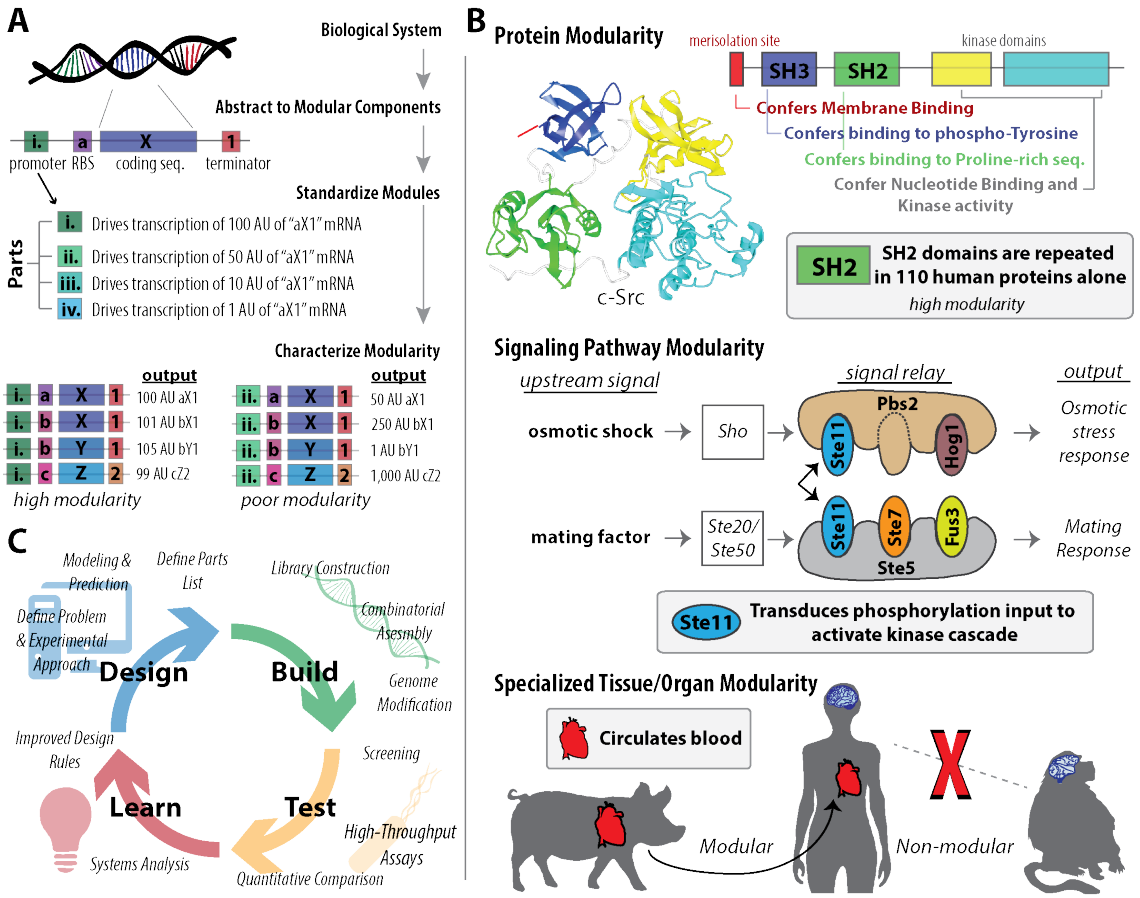
Appropriate design principles can improve the modularity of component parts. An example of a lack of modularity within biology can be found in the activity of RBS elements, which **A)** often exhibit variability in the degree of translation they promote, depending on context. For example, the same RBS element may drive a high expression of GFP (green fluorescent protein; top) but low expression of RFP (red fluorescent protein) due to unexpected nucleotide interactions within the mRNA that cause formation of secondary structure that inhibits ribosome binding (bottom). More predictive expression of a broad range of genes can be achieved through design principles that mitigate these problems. **B)** Creating bi-cistronic elements that consist of a leading RBS (RBS1) and a standardized leader sequence allows for upstream binding of ribosomes. The helicase activity of ribosomes translocating through the leader sequence disrupts secondary structure, revealing the internal RBS (RBS2). This design has been successfully implemented to greatly reduce the variability in gene expression that is achieved when using a given promoter/RBS combination, regardless of the target gene to be expressed (Mutalik et al., 2013).

Figure 3:

Examples of modular molecular biology tools utilized in synthetic biology. A) TALE effectors are characterized by a modular DNA-targeting region that is composed of multiple repeats of a protein domain. Each domain is nearly identical to the others, except that they can vary in two key amino acid residues (see blow-up insert) and these two residues confer specificity for binding to a target nucleotide. When multiple domains are connected in series, they can bind to a target DNA sequence by arraying next to one another within the major groove of the DNA (right). **B)** Cas9 protein recognizes small non-coding RNAs that have a characteristic hairpin sequence. When bound to a guide RNA (red) the Cas9 protein is able to use standard base pairing interactions to bind to the complimentary sequence within a target genome. Both Cas9 and TALE proteins can be readily modified with a functional domain (FD) to confer a desired function that will preferentially affect the target sequence: capacity to induce double-stranded or single-stranded DNA breaks, or domains that enhance/repress recruitment of transcriptional machinery. **C)** Concentration of proteins to a subcellular location (top) is a recurring theme to improve fidelity and efficiency within signalling and metabolic pathways. Artificial scaffolds have been constructed by encoding a string of binding domains (e.g. SH2 domains) on a single polypeptide that correspond to ligand domains that are appended to target proteins. When the artificial scaffold is expressed, it recruits the target proteins through receptor-ligand interactions, effectively concentrating the enzymes relative to one another. Early designs consisted only of single, isolated scaffolds, while more recent examples have favoured scaffolding proteins that can self-assemble into defined, marcromolecular arrays (depicted as tiled hexagons). **D)** Protein degradation can be experimentally controlled by modifying target proteins so they encode C-terminal "degron" tags (ssrA tag). These peptides are typically recognized by endogenous proteasome machinery (ClpX) and targeted for degradation. By introducing point mutations to the ssrA tag sequence, the marked protein can only be recognized when the adaptor protein (SspB) is present to recruit it to the proteasome, allowing for inducible downregulation of the target by controlling the expression of sspB.

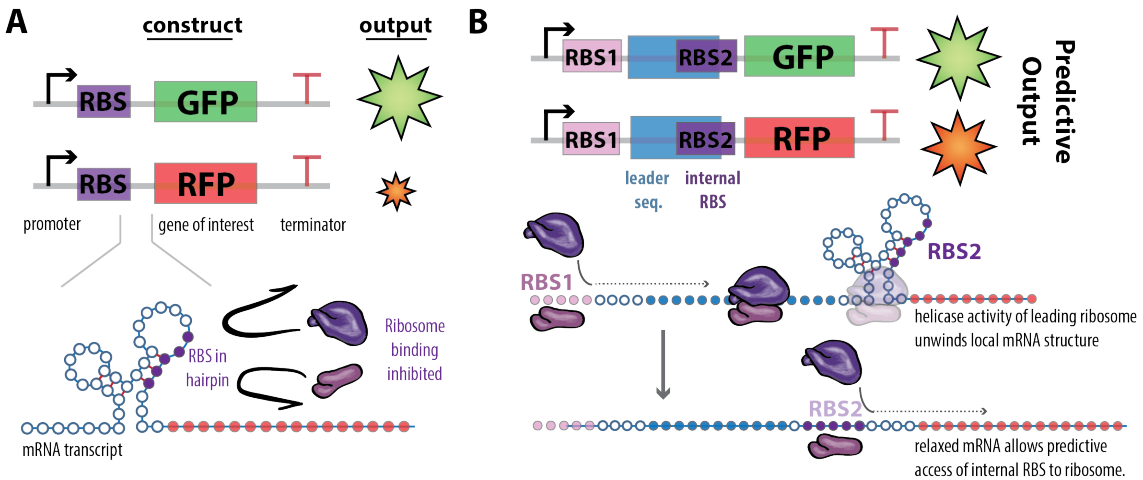
Figure 4:

Assembly of simple modules to create increasingly complex circuits and systems. A) Two-component logic gates process two input signals and activate an output response (1) when the appropriate conditions are met. A lookup table (grey inset) illustrates the output of 4 basic logic gates under each condition of for commonly-used gates. **B)** Connection of component logic gates together in series can allow for higher-order complexity in genetic circuit design. For instance, connecting 3 AND gates will create a coincidence detector that activates a target output (e.g. gene expression; red line) only when all 4 input criteria are met. Other complex output patterns, such as "memory" or oscillatory outputs can be generated by connecting simple circuits. Complex behaviours can arise when feedback loops are present in otherwise simple networks – e.g. a device that activates only when exposed to two signals, and retains memory of this activation (bottom). **C)** Individual cells, or species can also be abstracted as modules within a larger community. Here, an autotrophic module (the cyanobacterium *S. elongatus*) has been engineered to utilize photosynthesis to fix carbon and export a simple sugar (sucrose). In synthetic communities, this can be regarded as an "autotrophic" module that provides organic carbon to power other desirable metabolic reactions in heterotrophic modules. Combining species modules can confer desired properties into synthetic consortia without having to engineer complex processes (e.g. light-harvesting or nitrogen-fixation) into a single chassis.


A
Biological System
Abstract to Modular Components
promoter RBS coding seq. terminator
Standardize Modules
Parts
Drives transcription of 100 AU of "aX1" mRNA
Drives transcription of 50 AU of "aX1" mRNA
Drives transcription of 10 AU of "aX1" mRNA
Drives transcription of 1 AU of "aX1" mRNA
Characterize Modularity
output
100 AU aX1
101 AU bX1
105 AU bY1
99 AU cZ2
high modularity
50 AU aX1
250 AU bX1
1 AU bY1
1,000 AU cZ2
poor modularity
B
Protein Modularity
merisolation site
kinase domains
SH3
SH2
Confers Membrane Binding
Confers binding to phospho-Tyrosine
Confers binding to Proline-rich seq.
Confer Nucleotide Binding and Kinase activity
c-Src
SH2 domains are repeated in 110 human proteins alone
high modularity
Signaling Pathway Modularity
upstream signal
signal relay
output
osmotic shock
Sho
Pbs2
Ste11
Hog1
Osmotic stress response
mating factor
Ste20/ Ste50
Ste11
Ste7
Fus3
Ste5
Mating Response
Ste11
Transduces phosphorylation input to activate kinase cascade
Specialized Tissue/Organ Modularity
Circulates blood
Modular
Non-modular
C
Modeling & Prediction
Define Parts List
Library Construction
Define Problem & Experimental Approach
Combinatorial Assembly
Design
Build
Genome Modification
Improved Design Rules
Screening
Learn
Test
High-Throughput Assays
Systems Analysis
Quantitative Comparison

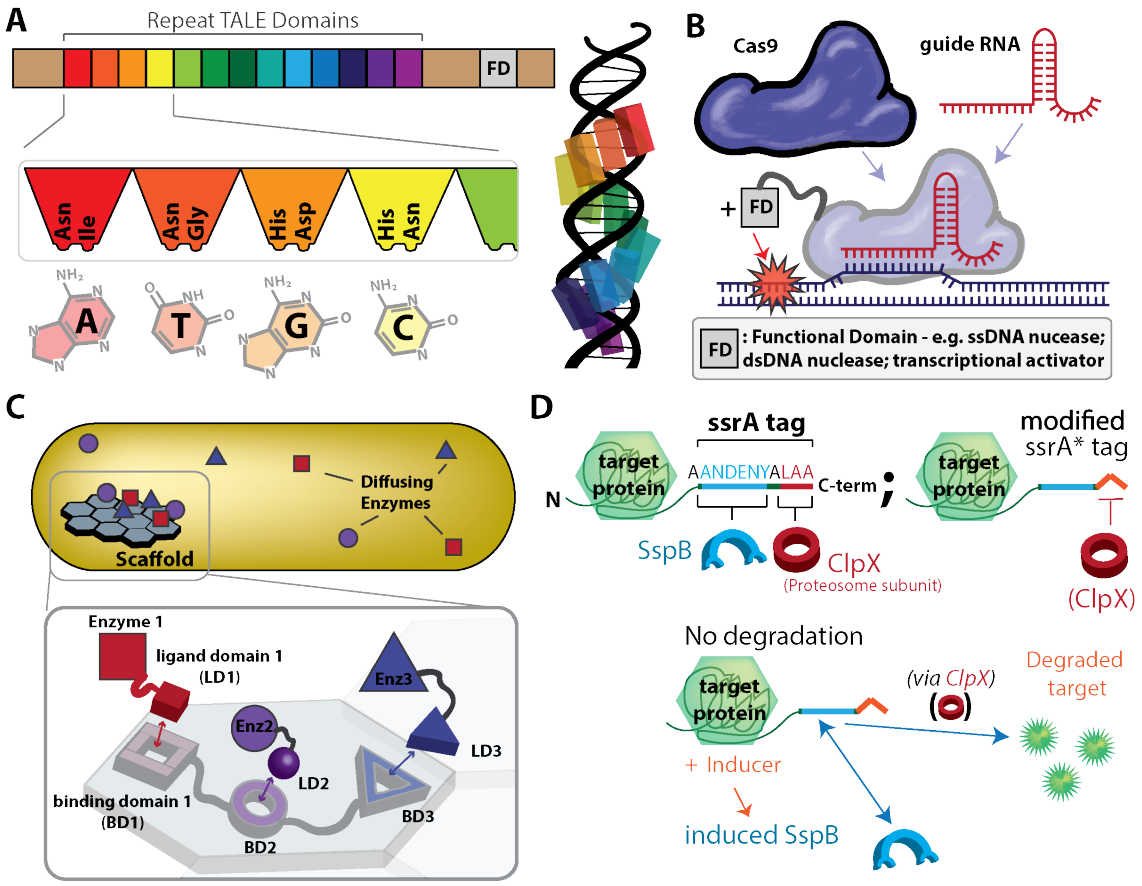
1

2


A
construct
output
RBS
GFP
RBS
RFP
promoter
gene of interest
terminator
RBS in hairpin
Ribosome binding inhibited
mRNA transcript
B
RBS1
RBS2
GFP
RBS1
RBS2
RFP
Predictive Output
leader seq.
internal RBS
RBS1
RBS2
helicase activity of leading ribosome unwinds local mRNA structure
RBS2
relaxed mRNA allows predictive access of internal RBS to ribosome.

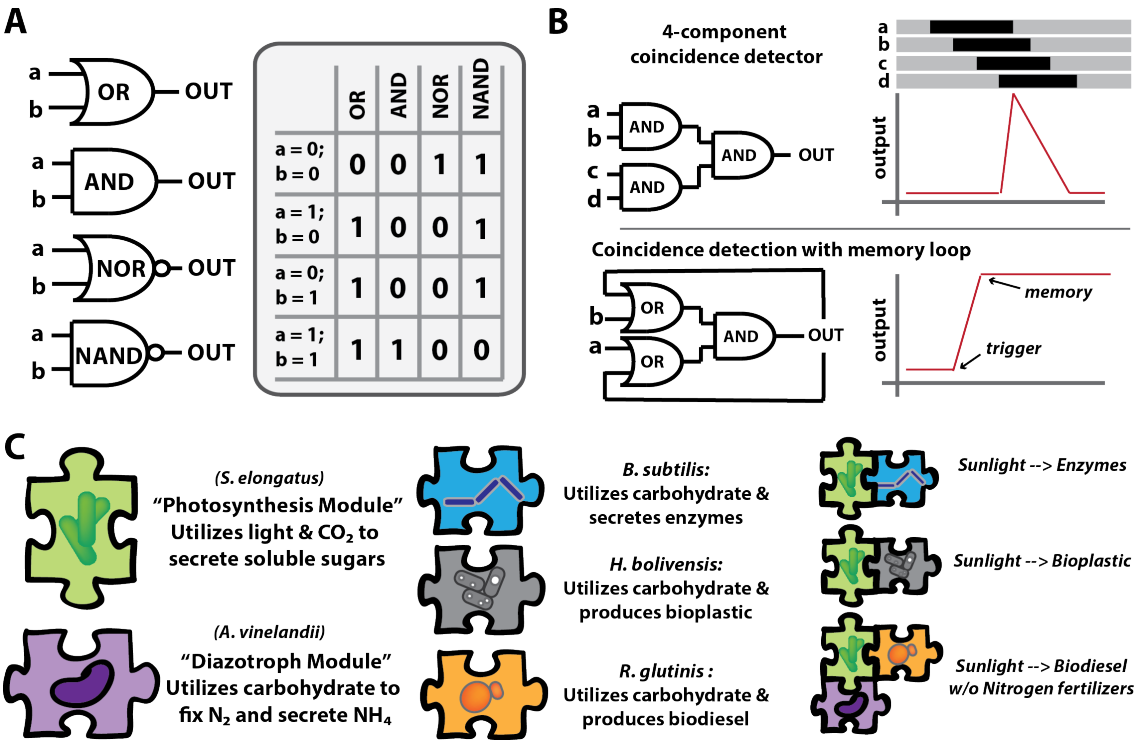
1

2


A
Repeat TALE Domains
FD
Asn Ile
Asn Gly
His Asp
His Asn
A
T
G
C
B
Cas9
guide RNA
+ FD
FD : Functional Domain - e.g. ssDNA nucease; dsDNA nuclease; transcriptional activator
C
Diffusing Enzymes
Scaffold
Enzyme 1
ligand domain 1 (LD1)
Enz3
Enz2
LD3
LD2
binding domain 1 (BD1)
BD3
BD2
D
ssrA tag
target protein
N
AANDENYALAA
C-term
SspB
ClpX
(Proteosome subunit)
modified ssrA* tag
(ClpX)
No degradation
(via ClpX)
Degraded target
+ Inducer
induced SspB

1

2


A
a
b
OR
OUT
AND
NOR
NAND
a = 0; b = 0
a = 1; b = 0
a = 0; b = 1
a = 1; b = 1
B
4-component coincidence detector
c
d
output
Coincidence detection with memory loop
memory
trigger
C
(S. elongatus)
"Photosynthesis Module"
Utilizes light & CO_2 to secrete soluble sugars
(A. vinelandii)
"Diazotroph Module"
Utilizes carbohydrate to fix N_2 and secrete NH_4
B. subtilis:
Utilizes carbohydrate & secretes enzymes
H. bolivensis:
Utilizes carbohydrate & produces bioplastic
R. glutinis :
Utilizes carbohydrate & produces biodiesel
Sunlight --> Enzymes
Sunlight --> Bioplastic
Sunlight --> Biodiesel w/o Nitrogen fertilizers

1