Modular Analysis and Design of Biological Circuits

Theodore W. Grunberg[a] and Domitilla Del Vecchio[b]

Abstract: Modularity has been subject of intense investigation in systems biology for more than two decades. Whether modularity holds in biological networks is a question that has attracted renewed attention in recent years with the advent of synthetic biology, mostly as a convenient property to perform bottom-up design of complex systems. A number of studies have appeared, which fundamentally challenge modularity as a property of engineered biological circuits. Here, we summarize some of these studies and potential engineering solutions that have been proposed to enable modular composition of biological circuits.

Introduction

When analyzing or designing a complex system, it is often convenient to view the system as being composed of subsystems that have been previously characterized in isolation. If the salient input/output properties of the subsystems are known and it is clear how these are connected to each other, then the behavior of the complex system can be inferred by composing the behaviors of its subsystems. In this compositional approach, it is critical that the input/output behavior of any subsystem characterized in isolation does not change upon composition (Figure 1A). Whether modular composition holds is determined by where the boundaries between modules are drawn and by what inputs and outputs are considered. Indeed, there have been a number of studies focused on this general question, which proposed formalisms to analyze interconnected systems. These include the behavioral systems framework developed by Willems [1] and the framework proposed by Paynter to connect systems based on energy transactions [2]. Modularity is widely used in engineering, especially in electrical and computer systems. The design of these systems largely relies on composing parts into devices, devices into circuits, and circuits into subsystems by "forgetting" the complexities internal to each device, circuit, or subsystem, once these are composed or inserted into a larger system (Figure 1B). We can often ignore the internal structure of modules once we assemble these together because modules have been conveniently engineered such that their input/output behavior is robust to interconnections [3]. One example that demonstrates the power of exploiting modular composition is modern computers. In terms of physical implementation, a computer is composed mostly of many transistors. The problem of arranging these transistors so that the result is a functioning computer is made tractable through a hierarchical series of abstractions. At each layer, functional modules are created from the composition of submodules in a lower layer of abstraction. The internal working of each module is abstracted away, allowing the design of each layer to proceed without worrying about how the lower layers work (Figure 1B). For example, the NAND gates are designed to apply only small loads to the gates connected to their inputs, and to be minimally affected by the loads applied by the gates connected to their outputs [3].

In biology, there are numerous examples of molecular core processes that are combined in different ways to create function in biological cellular networks. However, the extent to which modularity can be used as a paradigm to understand a natural system or to *de novo* engineer a synthetic one is still not completely known. For example, the complex circuitry that controls bacterial chemotaxis [4] can be viewed as the composition of three main subsystems as depicted in Figure 1C. The sensing module measures the extracellular concentration of a ligand, the computation module decides what response is appropriate, and the actuation module switches between driving the bacterium in a straight line and making it tumble, with a switching frequency determined by the output of the computation module. This is useful to understand the function of the overall system as the composition of each of the three modules, each carrying out its own function largely independent of the others [5]. In systems biology, an intriguing notion of modularity has been introduced by John Gerhart and Marc Kirschner to help explain how viable phenotypic

[a]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA, 02139, USA grunberg@mit.edu

[b]Corresponding author. Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA, 02139, USA ddv@mit.edu

variation arises from genetic changes. Certain "conserved core processes" such as transcription and translation of DNA into proteins, and signal transduction via covalent modification of proteins, can be viewed as "modules" that largely keep unchanged functionality through evolution and are rewired in different ways through evolutionary changes [6, 7, 8].

In the remainder of this article, we focus on modularity in engineered biological systems.

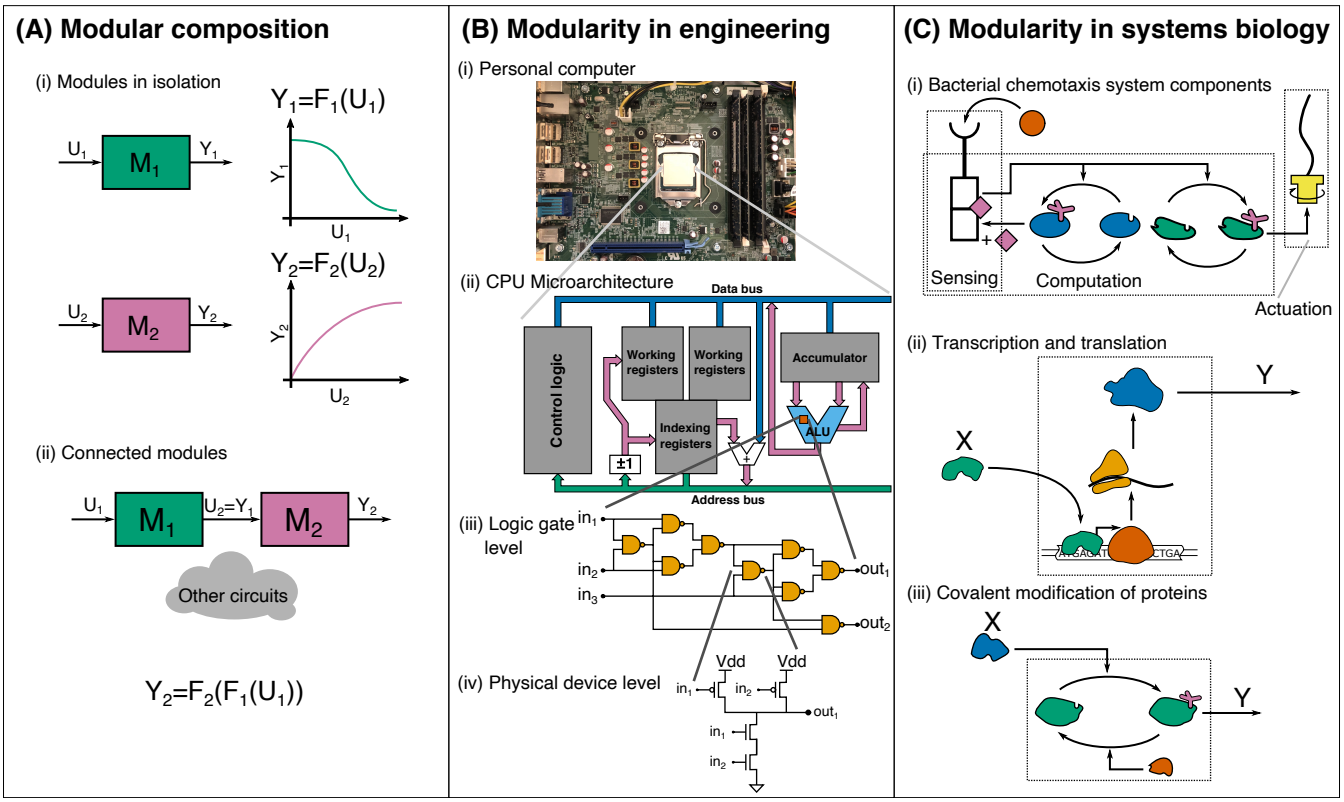

Figure 1: **Modularity.** (A) Modular composition. (i) The green and purple plots show the input/output relationships, F_1 and F_2, of modules M_1 and M_2 from inputs U_1, U_2 to outputs Y_1, Y_2, respectively. (ii) M_1 and M_2 are connected, setting $U_2 = Y_1$, and independent of the presence of other circuits, the input/output relationship from U_1 to Y_2 is the composition of F_1 and F_2. (B) Modularity in engineering. The heirarchical structure of a personal computer. (i) Some of the components of a personal computer. (ii) The microarchitecture of a CPU (here the Zilog z80 CPU [9]). Gray blocks represent groups of registers, capable of storing values, connected together with blocks that perform operations, such as the arithmetic logic unit (ALU) via data buses. The orange square inside the ALU denotes an adder. (iii) The construction of a one-bit adder from NOT-AND (NAND) gates. Each orange symbol denotes a NAND gate and the black lines indicate wires connecting them. (iv) One possible construction of a NAND gate from metal-oxide-semiconductor transistors. (C) Modularity in biology. (i) A simplified schematic of the circuitry that dictates the bacterial chemotaxis response [4, 8, 7]. The arrows indicate chemical interactions between species. Three major subsystems are shown: sensing, computation, and actuation. Two examples of "conserved core processes" [6]. (ii) Transcription and translation of DNA into proteins. X regulates the transcription of DNA into mRNA by RNA Polymerase (orange), and the mRNA is translated into a protein (blue) by ribosomes (yellow) [10]. (iii) A signal transduction pathway constructed from a covalent modification cycle. For phosphorylation, X mediates the addition of a phosphoryl group (purple) to the green protein, producing Y, while the orange protein mediates the removal of the phosphoryl group [10].

Lack of modularity of engineered genetic circuits

In engineered genetic circuits, modularity can fail at different levels. Here, we focus on genetic devices, that is, on systems that produce an output protein, such as a transcription factor, in response to a regulatory input molecule, such as another transcription factor or a small signaling molecule. We refer to such devices as modules. These devices are critical building blocks of engineered genetic circuits, are usually characterized in "isolation", and their input/output characteristics are used to infer the behavior of multiple composed such devices [11]. We thus illustrate ways in which the input/output response of a device can change when the device's context changes.

The first, and most widely studied, instance of modularity failing is when the input/output response of a device is affected by additional regulators, different from the intended regulatory inputs (Figure 2A). These phenomena are broadly referred to as "off-target effects", which occur any time a regulatory molecule does not bind to its target with perfect specificity [12, 13]. As a consequence, the input/output relationship of a given device is affected by unintended regulators (Figure 2A). These unintended interactions cause a change of the input/output response of module M_1 when module M_2 is present, and vice-versa. Indeed, when they are used concurrently, one cannot rely on their input/output behaviors characterized in isolation to predict their combined response to inputs.

A second instance of modularity failing is when a device's output protein becomes "loaded" due to it being an input to another device. In Figure 2B, Module 1 responds to regulatory input U_1 with TF Y_1 and Module M_2 "measures" Y_1 by having it activate the production of protein Y_2, a fluorescent protein, for example. However, when Module M_3 is added, which also takes Y_1 as an input regulator, the free level of Y_1 decreases due to "sequestration" of TF Y_1 by the promoter in M_3. As a consequence the level of Y_1 decreases, and with it, its measurement Y_2. Therefore, the input/output response of M_1 changes depending on the presence of M_3. The issue of loads on TFs is pervasive in the engineering of genetic circuits and its effects have been widely analyzed experimentally [14, 15]. In [16], a theoretical systems framework was proposed to capture the effect of loads upon module interconnection. Specifically, the authors proposed to describe each module by adding to its input U and output Y, an additional input S, called the retroactivity to the output, and an additional output R, called the retroactivity to the input (Figure 2B). Here, S captures the load applied to the module by a downstream system and R captures the load that the module applies to its upstream system. If these signals are explicitly accounted for during composition of genetic circuits, then modular composition can be restored [17].

Every genetic device in a cell uses resources to perform transcription and translation. These resources are limited, and while being sequestered by one module they are unavailable to others. This decrease in the available resources leads to a situation where each module's input/output response is different when other modules are present. In fact, it has been demonstrated experimentally that the expression level of any two proteins can become coupled due to resource competition [18]. This coupling can further cause surprising changes on the emergent behavior of a genetic circuit [19]. As depicted in Figure 2C, modules M_1 and M_2 compete for resources and thus the output of M_2 (purple protein) drops when the input to M_1 is increased. In particular, it has been experimentally demonstrated that protein levels of Y_1 and Y_2 are constrained on an isocost line [18]. This violates the property of modular composition. The block diagram in Figure 2C depicts this problem, in which each of modules M_1 and M_2 apply a load, shown as a red (retroactivity) arrow, to the host cell resources. Modules M_1 and M_2 thus become undesirably coupled through these load arrows.

In a related phenomenon, genetic circuits on plasmid cause a growth rate defect and this, in turn, cause a change in the dynamics of the genetic circuits. As shown in Figure 2D (i), when synthetic genes are expressed, transcriptional and translational resources are used, thereby possibly reducing the production rate of genes that control growth rate. This reduces the growth rate of the cell [20, 21]. In turn, growth rate affects synthetic circuits by, for example, changing the dilution rate of proteins [22] and by changing translation rate [23]. Thus, each device can affect the growth dynamics of the cell and, in turn, the growth rate of the cell affects each device's input/output response, leading to unintended interactions among modules (Figure 2D(ii)).

Finally, the input/output response of a genetic device can be affected by the genetic context. For example, the expression level of one gene in response to its regulators can be affected by the orientation of other surrounding genes [24]. Some of these effects are shown in Figure 2E in which two genetic devices, M_1 and M_2, are shown on plasmids. In one case, the gene of M_2 is placed so that it diverges from the gene of M_1, and in the other case the gene of M_2 is placed in tandem with that of M_1. The response of M_1 to its input U_1 is different in these two cases, suggesting an unintended interaction between M_1 and M_2. In particular, in [24] it is explained that DNA supercoiling is a major factor responsible for these differences.

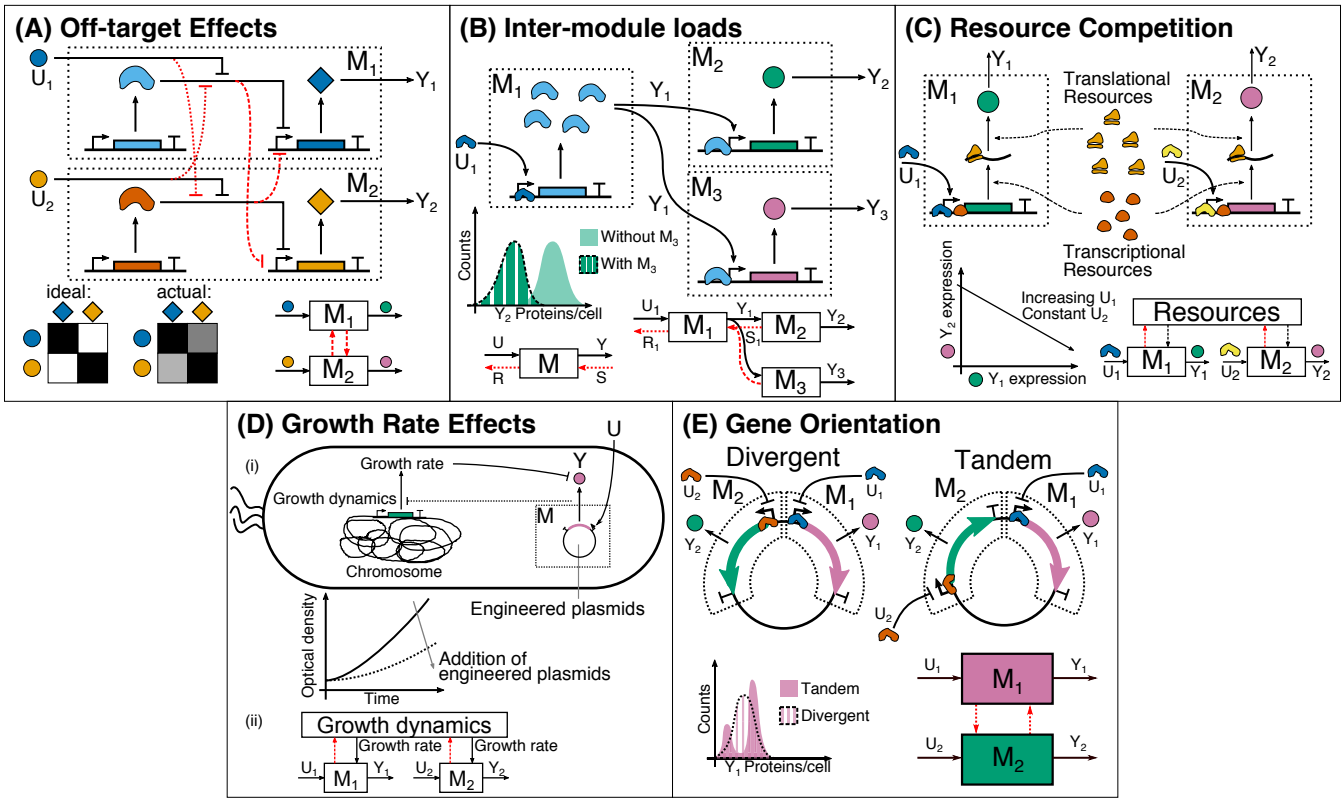

Figure 2: **Lack of modularity of engineered genetic circuits.** (A) Off-target effects. Modules M_1 and M_2 produce outputs Y_1 and Y_2 in response to inputs U_1 and U_2, respectively. The engineered interactions are indicated with solid lines and the arrows at a right angle indicate promoters, and the "T's" represent terminators. The unintended regulations (dashed lines) due to off-target effects cause the actual input/output response of the modules working together to deviate from the corresponding responses characterized in isolation. (B) Inter-module loads. M_1 is cascaded with M_2 as shown by the solid arrow labeled U_2. When M_3 is also connected to the output of M_1, the output of M_2 changes as shown in the plot. The red arrows capture loading as a signal, retroactivity, that a downstream module applies to the upstream module. (C) Resource competition. Modules M_1 and M_2 produce outputs Y_1 and Y_2 in response to inputs U_1 and U_2, respectively. The modules share the same pools of transcriptional (orange) and translational (yellow) resources. The plot shows Y_2 versus Y_1 for different values of U_1. The red arrows in the block diagram indicate the load applied to resources by M_1 and M_2. (D) Growth rate effects. (i) Certain genes that are key for growth (green) are repressed by sequestration of translational resources by a synthetic gene (purple) on a plasmid. The genetic module on the plasmid produces output protein Y in response to input U. The optical density of the cell culture over time with and without the presence of the input U is plotted. (ii) A block diagram showing the implications of growth effects on modularity. The red arrows indicate that modules M_1 and M_2 affect growth rate. (E) Gene orientation. Module M_1 response is affected by the orientation of a gene in a nearby module M_2. The plot shows the steady state distribution of Y_1 in response to induction of U_1 when U_2 is also induced. The dashed red arrows show the unintended interactions between the modules due to DNA supercoiling [24].

Methods for restoring modularity

One approach to mitigate off-target effects is the creation of orthogonal libraries of transcription factors and corresponding target promoters. For example, directed evolution was used to create a "Marionette" strain of *E. coli* containing 12 transcription factors, each of which regulates a corresponding promoter based on the level of a ligand cognate to the TF [13]. This was done by repeatedly mutating the promoters and TF coding sequences, then selecting for the most effective mutants based on positive selection and negative selection. All selection was done with only one gene at a time being induced so that only off-target effects due to cross reactivity of ligands and TFs was reduced, and no other coupling between modules, such as resource competition, was minimized.

Inter-module loads can be attenuated by placing load drivers between any pair of modules that need to be connected to each other as shown in Figure 3B. Load drivers are devices that impart a small load to their upstream module, and can keep the same output despite large loads applied by downstream modules. This prevents the loads applied by downstream modules ($\mathsf{M_3}$) from changing the signal being transmitted by an upstream module ($\mathsf{M_1}$). Such load drivers have been constructed using single covalent modification cycles [25] or by using cascaded cycles [26], with the latter design carrying significant advantages over the former. Specifically, cascaded devices have the ability to decouple the design requirements of having a small retroactivity to the input while attenuating the retroactivity to the output [27]. Single-stage devices have not demonstrated this ability.

In order to address the resource sharing and growth defects issues, a number of solutions have recently appeared that use feedback control in different architectures. At a high level, solutions can be grouped into three different such architectures: a global regulation strategy, a local regulation strategy, and a host-favoring strategy (Figure 3C). Resource competition, specifically competition for translational resources, has been mitigated by a local feedback regulation strategy in [28, 29]. One example of this is the post-transcriptional feedback circuit realized through small interfering RNA (sRNA) in [28] (Figure 3D (i)). By cotranslating the protein of interest in any given module with an activator for an sRNA that targets the protein's mRNA, any module can be made robust to perturbations in the amount of available ribosomes in the cell [28]. This decouples the operation of independent devices, thus aiding modularity. Transcriptional feedback regulation was also considered in [29], however this approach requires the input and output proteins of each module to be codesigned such that the output binds to and sequesters the input, posing a limitation on the connectivity of modules. A different solution considered a global regulation strategy, which tightly regulates the available ribosomes such that loads from one module do not affect the global resource pool. Such a controller was constructed to control the level of available o-ribosomes, a type of synthetic ribosome (Figure 2C (ii)) [30]. This way, the coupling between two independent modules using o-ribosomes could be mitigated. Alternative solutions also appeared, which proposed the construction of orthogonal resources, one for each genetic device [31].

In [21], the authors proposed to mitigate the effects of genetic modules on growth rate by a feedback system that senses the cellular burden (usage of translational resources), and appropriately suppresses the synthetic gene in a host-favoring strategy (Figure 3E). When the growth rate drops, the burden driven promoter increases its transcriptional activity, activating dCas9 based repression circuitry which transcriptionally represses the synthetic gene. Such a circuit ensures that when a synthetic load is introduced in the cell, its expression level is kept sufficiently low so to not significantly alter growth rate.

When the main concern of an application is to produce a synthetic protein, such as a metabolite, while not affecting growth, the host-favoring controller would be mostly appropriate. By contrast, if the objective is to construct a multi-module system, such as an array of sensors or a classifier, an approach where a local controller is used may be preferable as the controllers do not affect the emergent function of the system, while artificially repressing genes to keep burden low may disrupt overall circuit function. Global controllers may be used for the same purpose as local controllers, but the use of o-ribosomes limits the tunability of each device [32]. Other global controller designs may consider the possibility of directly controlling the endogenous ribosome pool. However, this remains a significant challenge due to the complexity of the endogenous ribosome regulatory circuitry.

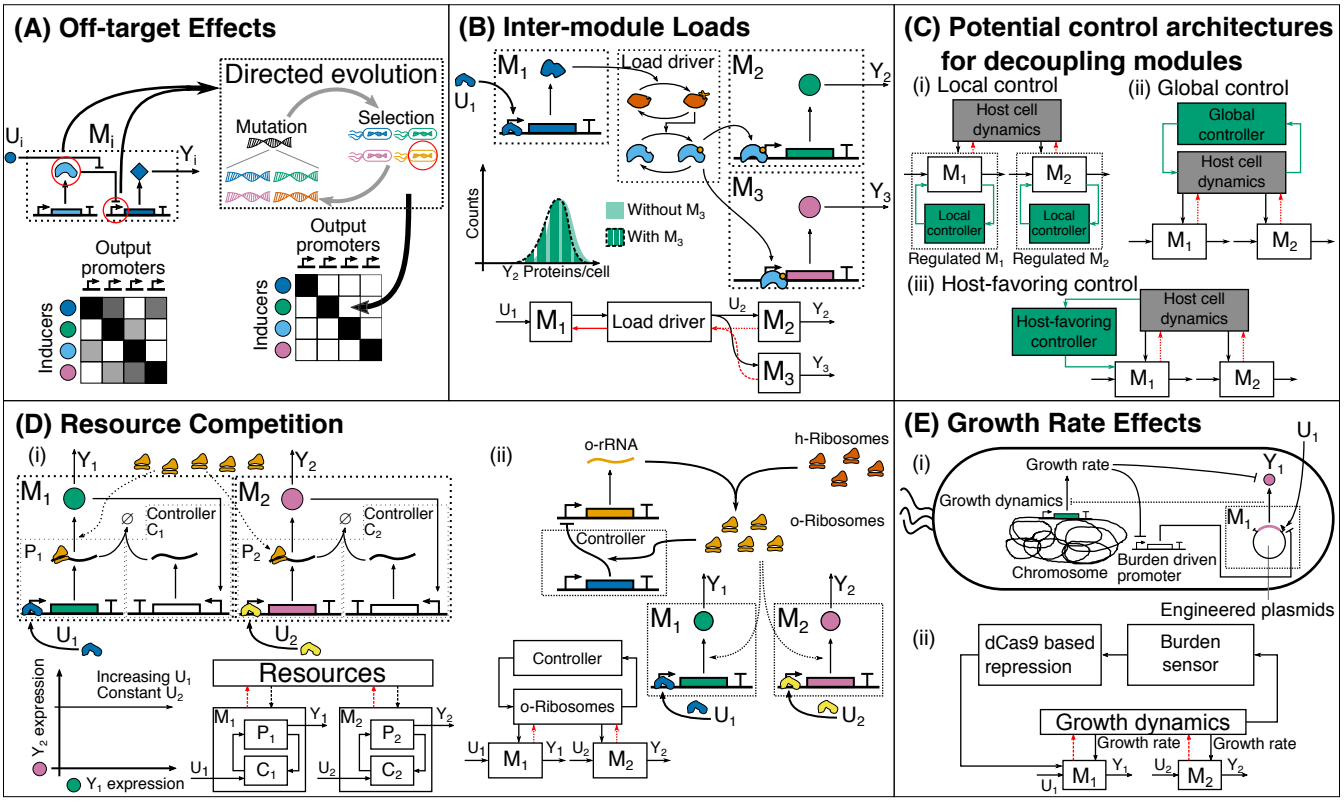

Figure 3: **Solutions for rescuing modularity.** (A) Off-target effects. Directed evolution can be used to optimize transcription factors and promoter pairs (circled in red). Transcription factors and promoters for each module are repeatedly mutated and the mutants that exhibit low off-target effects are selected for [13]. After directed evolution the interaction map between input ligands and output promoter activity of each module is diagonal. (B) Inter-module loads. A load driver placed between module M_1 and its downstream modules M_2, M_3 attenuates the effects of retroactivity. Black arrows indicate reactions between molecular species. The output of M_1 (dark blue) mediates the covalent modification of the green protein, which itself mediates the covalent modification of the light blue protein. The active form of the light blue protein is the load driver's output, and acts as a transcriptional input to M_1 and M_3 (the loads). The black and red arrows in the block diagram indicate engineered connections and loading effects (retroactivity), respectively. The plot shows the population distribution of the Y_2 green protein counts with and without module M_3 when the load driver is present (compare to the decrease in Y_2 in Figure 2B in the absence of a load driver). (C) Potential control architectures for decoupling resource and growth coupled modules. The green components of the diagram indicate the controller. (i) A local controller added to each module makes the individual modules robust to changes in the host cell's dynamics (resources or growth rate). (ii) A global controller regulates parts of the host cell's dynamics, thereby preventing modules from affecting it. (iii) A host-favoring controller regulates one module in response to changes to the host cell's dynamics, thereby attenuating the effect that a module has on it. (D) Resource competition. Modules M_1 and M_2 produce outputs Y_1 and Y_2, respectively, in response to inputs U_1 and U_2, respectively. The same pool of translational resources (orange) is shared between the modules. (i) Local controller. Each module is composed of the plant P_i and the controller C_i. The plot shows the outputs Y_2 and Y_1 as U_1 is increased. The red arrows in the block diagram indicate the sequestration of ribosomes by M_1 and M_2 [28]. (ii) Global controller. The modules are engineered to use o-ribosomes. A controller regulates the production of o-ribosomes in response to loading by M_1 and M_2 by repressing the production of o-rRNA when excess o-ribosomes are available [30]. (E) Growth rate effects. (i) A burden-driven promoter has a transcription rate that increases when the growth rate of the *E. coli* is reduced. Genes that repress the expression of the synthetic construct are placed under the control of the burden-driven promoter [21]. (ii) The block diagram shows this host-favoring control architecture applied to module M_1.

Conclusions

In this review, we have considered recent work concerning modularity as a tool for the analysis and design of genetic circuits. Traditional engineering disciplines suggest that modular composition may be a promising approach also for engineering biology, and significant progress has been made towards engineering genetic parts and devices that are modular. Nevertheless, it is still unclear whether enforcing modularity in engineered biological circuit is the best way to tackle the complexity of designing increasingly sophisticated systems. Other approaches that explicitly account for "unwanted" interactions and make these part of how systems are interconnected may be a promising alternative. This, however, may require to re-invent the rules for interconnecting systems and for designing network dynamics.

Conflict of interest statement

Nothing declared.

Acknowledgements

Support for this work was provided by the National Science Foundation through CMMI Award Number 1727189. The authors thank Yili Qian for his helpful suggestions on the manuscript.

References

(••) Indicates references of outstanding interest.

[1] Willems J. C and Polderman J. W. *Introduction to mathematical systems theory: a behavioral approach*, volume 26. Springer Science & Business Media, 2013.

[2] Paynter H. M. *Analysis and design of engineering systems*. MIT press, 1961.

[3] Shepherd P. *Integrated Circuits*. Macmillan International Higher Education, 1996.

[4] Rao C. V, Kirby J. R, and Arkin A. P. **Design and diversity in bacterial chemotaxis: A comparative study in escherichia coli and bacillus subtilis**. *PLOS Biology*, 2(2):e49–, 02 2004.

[5] Lauffenburger D. A. **Cell signaling pathways as control modules: Complexity for simplicity?** *Proceedings of the National Academy of Sciences*, 97(10):5031, 05 2000.

[6] Gerhart J and Kirschner M. **The theory of facilitated variation**. *Proceedings of the National Academy of Sciences*, 104(suppl 1):8582, 05 2007.

[7] Kirschner M and Gerhart J. C. *The plausibility of life: Resolving Darwin's dilemma*. Yale University Press, 2006.

[8] Alon U. **Biological networks: The tinkerer as an engineer**. *Science*, 301(5641):1866–1867, 2003.

[9] Zaks R. *Programming the Z80*. Sybex Incorporated, 1979.

[10] Del Vecchio D and Murray R. M. *Biomolecular feedback systems*. Princeton University Press Princeton, NJ, 2015.

[11] Nielsen A. A. K, Der B. S, Shin J, Vaidyanathan P, Paralanov V, Strychalski E. A, Ross D, Densmore D, and Voigt C. A. **Genetic circuit design automation**. *Science*, 352(6281):aac7341, 04 2016.

[12] Halleran A. D and Murray R. M. **Cell-free and in vivo characterization of lux, las, and rpa quorum activation systems in e. coli**. *ACS Synthetic Biology*, 7(2):752–755, 02 2018.

[13] Meyer A. J, Segall-Shapiro T. H, Glassey E, Zhang J, and Voigt C. A. **Escherichia coli "marionette"strains with 12 highly optimized small-molecule sensors**. *Nature Chemical Biology*, 15(2):196–204, 2019.

(••) The authors engineered a strain of *E. coli* which has 12 orthogonal transcription factors, each targeting its own promoter and with its own inducer.

[14] Jayanthi S, Nilgiriwala K. S, and Del Vecchio D. **Retroactivity controls the temporal dynamics of gene transcription**. *ACS Synthetic Biology*, 2(8):431–441, 08 2013.

[15] Lee T.-H and Maheshri N. **A regulatory role for repeated decoy transcription factor binding sites in target gene expression**. *Molecular Systems Biology*, 8(1):576, 2012.

[16] Del Vecchio D, Ninfa A. J, and Sontag E. D. **Modular cell biology: retroactivity and insulation**. *Molecular Systems Biology*, 4(1):161, 01 2008.

[17] Gyorgy A and Del Vecchio D. **Modular composition of gene transcription networks**. *PLOS Computational Biology*, 10(3):e1003486–, 03 2014.

[18] Gyorgy A, Jiménez J, Yazbek J, Huang H.-H, Chung H, Weiss R, and Del Vecchio D. **Isocost lines describe the cellular economy of genetic circuits**. *Biophysical Journal*, 109(3):639–646, 2015.

[19] Qian Y, Huang H.-H, Jiménez J, and Del Vecchio D. **Resource competition shapes the response of genetic circuits**. *ACS Synthetic Biology*, 6(7):1263–1272, 07 2017.

[20] Borkowski O, Bricio C, Murgiano M, Rothschild-Mancinelli B, Stan G.-B, and Ellis T. **Cell-free prediction of protein expression costs for growing cells**. *Nature Communications*, 9(1):1457, 2018.

[21] Ceroni F, Boo A, Furini S, Gorochowski T. E, Borkowski O, Ladak Y. N, Awan A. R, Gilbert C, Stan G.-B, and Ellis T. **Burden-driven feedback control of gene expression**. *Nature Methods*, 15:387 EP –, 03 2018.

(••) The authors engineered a controller which maintains the host cell's growth rate in the presence of a synthetic burden.

[22] Tan C, Marguet P, and You L. **Emergent bistability by a growth-modulating positive feedback circuit**. *Nature Chemical Biology*, 5(11):842–848, 2009.

[23] Klumpp S and Hwa T. **Bacterial growth: global effects on gene expression, growth feedback and proteome partition**. *Current Opinion in Biotechnology*, 28:96–102, 2014.

[24] Yeung E, Dy A. J, Martin K. B, Ng A. H, Del Vecchio D, Beck J. L, Collins J. J, and Murray R. M. **Biophysical constraints arising from compositional context in synthetic gene networks**. *Cell Systems*, 5(1):11–24.e12, 2017.

[25] Nilgiriwala K. S, Jiménez J, Rivera P. M, and Del Vecchio D. **Synthetic tunable amplifying buffer circuit in e. coli**. *ACS Synthetic Biology*, 4(5):577–584, 05 2015.

[26] Mishra D, Rivera P. M, Lin A, Del Vecchio D, and Weiss R. **A load driver device for engineering modularity in biological networks**. *Nature Biotechnology*, 32:1268 EP –, 11 2014.

[27] Shah R and Del Vecchio D. **Signaling architectures that transmit unidirectional information despite retroactivity**. *Biophysical Journal*, 113(3):728–742, 2017.

[28] Huang H.-H, Qian Y, and Del Vecchio D. **A quasi-integral controller for adaptation of genetic modules to variable ribosome demand**. *Nature Communications*, 9(1):5415, 2018.

(••) The authors engineered a controller which can make a gene robust to changes in the available ribosomes.

[29] Shopera T, He L, Oyetunde T, Tang Y. J, and Moon T. S. **Decoupling resource-coupled gene expression in living cells**. *ACS Synthetic Biology*, 6(8):1596–1604, 08 2017.

[30] Darlington A. P. S, Kim J, Jiménez J. I, and Bates D. G. **Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes**. *Nature Communications*, 9(1):695, 2018.

[31] Liu C. C, Jewett M. C, Chin J. W, and Voigt C. A. **Toward an orthogonal central dogma**. *Nature chemical biology*, 14(2):103–106, 01 2018.

[32] Rackham O and Chin J. W. **A network of orthogonal ribosome·mRNA pairs**. *Nature Chemical Biology*, 1(3):159–166, 2005.