Standardized excitable elements for scalable engineering of far-fromequilibrium chemical networks

Author list:

Samuel W. Schaffter¹, Kuan-Lin Chen¹, Jackson O'Brien², Madeline Noble¹, Arvind Murugan², Rebecca Schulman^{1,3,4,*}

Author affilications:

¹Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, 21218;

²The James Franck Institute and Department of Physics, University of Chicago, Chicago, IL, 60637;

³Computer Science, Johns Hopkins University, Baltimore, MD, 21218;

⁴Chemistry, Johns Hopkins University, Baltimore, MD, 21218

*Corresponding Author:

Rebecca Schulman Department of Chemical & Biomolecular Engineering, Johns Hopkins University Shaffer Hall 200B 3400 N Charles St, Baltimore, MD, 21218 Phone: 410-516-8457 email: rschulm3@jhu.edu

Keywords:

DNA nanotechnology, DNA computing, in vitro transcription



ABSTRACT

Engineered far-from-equilibrium synthetic chemical networks that pulse or switch states in response to environmental signals could precisely regulate the kinetics of chemical synthesis or self-assembly pathways. Currently, such networks must be extensively tune to compensate for the different activities of and unintended reactions between a network's different chemical elements. Elements with standardized performance would allow rapid construction of networks with designed functions. Here we develop standardized excitable chemical elements, termed genelets, and use them to construct complex *in vitro* transcriptional networks. We develop a protocol for identifying >15 interchangeable genelet regulatory elements with uniform performance and minimal crosstalk. These elements can be combined to engineer feedforward and feedback modules whose dynamics are predicted by a simple kinetic model. We show modules can be rationally integrated and reorganized into networks that produce tunable temporal pulses and act as multi-state switchable memories. Standardized genelet elements, and the workflow to identify more, should make engineering complex far-from-equilibrium chemical dynamics routine.

INTRODUCTION

The chemistry of life operates far from equilibrium^{1,2}. Complex cellular processes such as stress response^{3,4} and morphogenesis^{5,6} are orchestrated by genetic regulatory networks (GRNs) ^{6,7} that continuously require energy in order to rapidly and repeatedly change chemical dynamics in response to time-varying environmental stimuli. Each gene in a GRN can be considered an excitable unit, whereby ultrasensitivity and amplification enables a small change in a regulating molecule's concentration to lead to a dramatic change in output concentration for efficient signal propagation⁸. Synthetic chemistries⁹⁻²⁰ composed of excitable units that operate far-fromequilibrium like GRNs could direct complex chemical dynamics^{1,2,21–23} and pattern formation^{24–26} or chemical regulation of nanostructures^{27,28}. Cellular GRNs typically integrate 10-100 excitable units to direct complex dynamics and thus a key step toward emulating GRN dynamics in synthetic systems is scaling up synthetic far-from-equilibrium chemistries^{6,7}. The circuit elements^{10,12,19}, or nodes, for such synthetic chemistries are typically designed for specific modules with additional nodes designed as needed to increase complexity^{14,27,29,30}. However, iterative design is often needed to find working nodes, and repeatedly appending nodes constrains future design space. A library of orthogonal circuit elements with standardized performance would make it possible to rapidly construct modules and integrate them into larger dynamic networks (Fig. 1a).

We sought to develop such a workflow for *in vitro* transcriptional circuits^{12,16,19,20,28,31} composed of short transcriptional templates called genelets. Each genelet consists of an input (I) domain that combines a DNA activator binding site and an incomplete T7 RNA polymerase (T7 RNAP) promoter site, and an output (O) domain that encodes an RNA transcript. A genelet's output is transcribed when a DNA activator binds the genelet's input domain (Fig. 1b). Genelets regulate one another by transcribing RNAs that repress or activate a target node (Fig. 1c) by controlling the ability of DNA activators to bind their target genelets. A regulating RNA (rR or rC) changes a genelet's state when its concentration is higher than the concentration of its corresponding DNA regulator (dA or dB). DNA regulators are present in excess of their genelets, facilitating ultrasensitivity through molecular titration to produce excitable behavior akin to

neuronal firing³². Regulating RNAs are also degraded by RNase H, enabling signal turnover (Fig. 1d,e). Genelets therefore emulate the powerful aspects of cellular GRNs using only short DNA strands and two enzymes. However, scalable genelet network engineering has been limited by spurious side reactions and inconsistent regulation rates across sequence domains^{12,30}.

Here we develop a scalable method of building genelets with standardized performance. We identify >15 standardized circuit elements, or regulatory domains, and show that these domains can be interchangeably combined into genelets that activate and/or repress one another. We engineer a suite of feedforward and feedback modules and demonstrate that these modules can be integrated into mesoscale networks that execute tunable temporal dynamics or serve as switchable memories. Importantly, these engineered networks behave as a general model of genelet behavior predicts, and both feedback and feedforward modules perform similarly within mesoscale networks and in isolation, indicating circuit element orthogonality. Finally, we show our workflow can identify additional standardized domains. Standard genelet components will enable predictive engineering of far-from-equilibrium chemical networks making it possible to routinely program complex chemical synthesis pathways, reaction-diffusion patterns, or regulation of materials.

RESULTS

A modular design for bidirectionally regulated genelets

We first sought to create a library of standard genelet nodes that could be bidirectionally regulated, exhibit predictable and uniform behavior, and whose members have minimal crosstalk. To build this library, we first developed a standard genelet design in the form of sequence domains with specific lengths and complementarity (Fig. 1b). We then built a genelet library by choosing different, modular sequences for these domains that exhibited uniform performance.

We began with a genelet design that sequesters most of each node's long input domain within a hairpin which reduces the propensity for crosstalk between network components³⁰. Such a node can be repressed by upstream transcripts: A genelet:activator complex presents a 3' ssDNA toehold, and a strand displacement reaction between an RNA repressor and the complex mediated by this toehold (the repressor toehold, or RTH) removes the DNA activator. However, a 3' ssDNA overhang can facilitate promoter independent transcription by T7 RNAP, which can cause a genelet to turn itself off (autoinhibition)³⁰. To mitigate autoinhibition, we moved the ssDNA RTH domain to the 5' end of the activator (Supplementary Section 2). We termed this the HPC50 node design (Supplementary Fig. 1).

We next sought to modify the HPC50 design to allow upstream transcripts to not only repress but also activate downstream genelets. One way to create such bidirectionally regulated nodes would be to sequester a genelet's DNA activator in a double-stranded DNA (dsDNA) complex^{12,16,28}. An upstream RNA transcript could then release the DNA activator from this complex to turn a node on (Supp. Section 3.1). However, such a transcript must share sequence elements with the DNA activator; we found that transcripts that could release HPC50 DNA activators from this complex also bound to the genelets, preventing the activators from binding (Supp. Figure 6, Supp. Section 3.1).

We therefore devised a new scheme for bidirectional regulation, the HPC5 node design, in which a DNA blocker strand binds to the genelet to prevent or reverse DNA activator binding (Fig.



1b). Transcription can occur when the activator is bound, but not when the blocker is. Upstream signals can upregulate or downregulate genelet transcription by sequestering either the genelet's

Figure 1 | The HPC5 genelet toolbox. (a) Network engineering workflow. (b) Each genelet consists of an input (I) domain (G1 here), that controls transcription activity, and an output (O) domain, S1 here. The T7 RNAP promoter sequence (pink) of a genelet is not double-stranded, so little transcription occurs (OFF). Transcription occurs from a genelet:activator complex (ON) because the activator completes the promoter sequence. A DNA blocker prevents both transcription and DNA activator binding (BLK). Output domains that repress input domain *i* are labeled R_i and those that coactivate input domain *i* are labeled C_i . DNA and RNA species are depicted as solid or dashed lines and have a d or r prefix, respectively. Fluorophore (F) and quencher (Q) modifications are used to track genelet state. (c) Upstream transcripts reverse activation (repression) or blocking (coactivation). Coactivation passes through OFF to get to ON. (d,e) An RNA repressor turns a node OFF. The DNA blocker was omitted but it could displace the activator if free (d). An RNA coactivator removes the blocker to allow activator binding, which turns a node ON (e). Both repression

and coactivation are reversed via RNase H degradation. Reactions with free activators and blockers are shown in Supplementary Fig. 2.

DNA blocker or DNA activator (Fig. 1c). An RNA repressor sequesters the DNA activator to downregulate expression (Fig. 1d), while an RNA coactivator sequesters the DNA blocker to upregulate expression through DNA activator binding (Fig. 1e). We confirmed that the DNA blocker prevented activator binding and DNA and RNA coactivators both facilitated genelet activation. We also verified that coactivation could be reversed by adding blocker or degrading the RNA coactivator (Supplementary Section 3.2).

Creating a library of interchangeable regulatory domains with standardized behavior

Using the HPC5 design, we sought to create a combinatorial library of modular genelet input and output sequences that could be interchangeably assembled into networks. We used NUPACK $3.2.2^{33}$ to design sequences for 36 new genelet input-output sequences predicted to be thermodynamically stable in their designed hairpin structures and to have minimal non-designed secondary structure interactions with each other or the G1 HPC5 node (Supplementary Section 4.1). We measured rates of activation of two sample sequences from this library and found they differed by 5-fold (Supplementary Fig. 14). This difference is consistent with previously reported rate constants of 4-way branch migration reactions^{34,35} which ranged from 10^3-10^4 M⁻¹s⁻¹, so 150 nM DNA activator would be expected to activate >0.9 fraction of 25 nM genelets anywhere between 15-100 minutes. To obtain standardized behavior across nodes, we sought to identify sequences that each activated relatively fast, which we defined as reaching >0.9 fraction ON in <1 h. We developed a screening assay to quickly and inexpensively measure the DNA activation rates of the remaining 34 designed sequences (Supplementary Section 4.2). Sixteen satisfied these screening criteria (Fig. 2b and Supplementary Fig. 15).

We next tested the coactivation and repression kinetics of these 16 nodes. Five input sequence domains randomly selected from these nodes all turned on and off rapidly in response to their respective DNA coactivators and repressors. But two of these nodes coactivated and/or repressed much more slowly when their RNA coactivators or repressors were transcribed (Fig. 2c and Supplementary Fig. 18), which suggested the RNA might be incompletely transcribed or misfolded³⁶. To eliminate nodes that exhibited slow transcriptional regulation, we developed a faster and quicker screening assay to measure coactivation and repression kinetics (Fig. 2c and Supplementary Section 4.3). We used this assay to determine that eight of 11 remaining untested nodes reached >0.9 fraction ON/OFF in less than an hour in response to both DNA and transcribed RNA regulators (Fig. 4d and Supplementary Section 4.4). Five could only be regulated quickly in one direction, suggesting these should be used when only that direction of regulation is required⁷. In total, this screening process therefore yielded 11 distinct node sequences that could be quickly coactivated, repressed, or both.

To test whether these 11 nodes' input and output domains (Fig. 1b) could be used interchangeably, we measured the activation kinetics of 28 unique input and output combinations. Twenty-seven of these took <1 h to switch their targets to >0.9 ON/OFF. These experiments also suggested that the choice of the output domain of a genelet did not generally affect the rate of

regulation of its input domain: nodes with the same input domain generally switched on or off in similar times (Supplementary Section 4.5).



Figure 2 | Design and screening protocol for identifying sequences for standardized HPC5 genelet domains. (a) A library of genelet node sequences, termed HTT variants, are designed computationally (using NUPACK) by minimizing undesired secondary structure interactions between library sequences. (b) HTT genelet variants that reach >0.9 fraction ON in <1 h after addition of their DNA activator (in the absence of enzymes) are identified from the designed library. (c) The rates of coactivation and repression by transcribed RNA regulators of HTT variants identified in (b) are then measured using constitutively active transcription templates for RNA coactivators and repressors (solid lines). HTT genelet variants that can reach >0.9 fraction ON in <1 h are selected as node sequences. Dashed curves show rates of coactivation/repression in response to DNA coactivators/repressors. (d) The number of nodes that passed

each stage of the screening protocol. See Supplementary Section 4 for a detailed description of screening protocols. (e) The standardized HPC5 genelet input and output domains can be interchanged without significant changes in regulatory kinetics. See Supplementary Section 4.5 for individual kinetic trajectories, experimental conditions, and additional input and output domain combinations.

Engineering temporal genelet expression programs

We next investigated whether the standardized nodes we identified could be integrated into regulatory networks whose dynamics could be quantitatively predicted and programmed. To guide network design, we developed a general model of genelet dynamics consisting of the ordinary differential equations describing the mass action kinetics of the designed reactions involved in genelet regulation, transcription, and RNA degradation for the set of genelets making up a network (Supplementary Section 5). Because each of the library's nodes has similar performance, the model assumes the kinetic rate constants are the same regardless of a node's sequence.

We first used the model to design modules that produce well-defined pulses in signal expression through coordinated genelet coactivation and repression, as in the incoherent type 1 feedforward loop (IFFL) motif common in cellular GRNs^{6,7,37,38} and biochemical information processing^{39,40}. In a genelet IFFL (Fig. 3a), an output node (G1) should pulse when the activator of the upstream G2 nodes is added.

Simulations indicated that the relative rates and strengths of G1 coactivation and repression determine pulse shape and timing (Supplementary Fig. 6.1). Varying the concentrations of the IFFL components produced trends in pulse kinetics predicted by these simulations (Supplementary Fig. 29). However, we also found that the BLK G3R1 node turned on in the absence of input in the IFFL1, which was not predicted. In isolation, BLK G3R1 turned itself on at a rate proportional to the dB3 concentration, suggesting promoter-independent transcription of unbound DNA blockers⁴¹. We modified the single-stranded 3' ends of the blockers to be 2' methylated RNA⁴², which eliminated BLK G3R1 autoactivation (Supplementary Fig. 32). We also found the G8 and G10 nodes exhibited autoactivation, so we also changed their single-stranded 3' ends to 2' methylated RNA (Supplementary Section 7). Using methylated blockers, we were able to reliably tune IFFL behavior and responsiveness as predicted in simulations: when G2 nodes were activated, increasing both pulse height and width, and when G2 was not activated, the network did not respond (Fig. 3b).

A key feature of a library of standardized nodes is that the same module could be built from any set of nodes in the library. To test whether nodes could be interchanged in this way, we designed two additional IFFL modules using other genelet input domains from our library and evaluated their dynamics using the same concentrations of nodes used for IFFL1. All three modules (IFFL1, IFFL2, and IFFL3) exhibited similar pulse dynamics that qualitatively matched the predictions of the general model (Fig. 3b-d).

We next asked whether we could connect multiple IFFL modules in a composite network, the IFFL1_2^{3,7} (Fig. 3e), which the general model predicted would produce sequential pulses of the two IFFL modules' outputs when triggered (Fig. 3f and Supplementary Fig. 35). To test this hypothesis, we initially connected the IFFL2 module to an upstream node (G3C5). This new connection caused the G3C5_IFFL2 network's output to pulse even in the absence of input. We hypothesized this spurious pulsing was caused by leak transcription from BLK G5 genelets

(Supplementary Section 8.4). Incorporating a leak transcription from BLK G5 nodes of 5% the rate of ON G5 nodes in the general model recapitulated the G3C5_IFFL2's spontaneous activation. This revised model predicted that increasing the concentrations of the DNA blockers in the IFFL2 module should prevent spontaneous activation (Supplementary Fig. 36). We also identified



Figure 3 | Incoherent feedforward loops (IFFL) orchestrate temporal pulses in genelet activation. (**a**) The IFFL1 network. Here and elsewhere, a yellow star on a node indicates that the node was fluorescently modified to track its activation levels during experiments. (**b**) Normalized activation levels of IFFL1 nodes predicted by the general genelet model (left) and in experiments (right) for networks with 125, 250, or 500 nM dA1. Dashed lines in plots show activation levels when dA2 was not added. The general genelet model is described Supplementary Section 5. (**c**-**d**) Two IFFL modules assembled from different nodes (above plots). Normalized activation levels of network nodes with 125, 250, or 500 nM of the pulsing node's activator. Dashed lines in plots show activation levels when dA5 (c) and dA9 (d) were not added. (**e**) The IFFL1_2 network, consisting of the IFFL1 and IFFL2 modules connected in series. (**f**) Normalized activation levels of IFFL1_2 network nodes with different concentrations of G3C5. Detailed

methods are in Supplementary Sections 6.2 and 8.2. IFFL1, IFFL2, IFFL3, and IFFL1_2 design notes and sequences are in Supplementary Sections 6.3, 6.5, 6.6, and 8.1.

undesired hybridization between rR1 and rR4 that altered the IFFL1 pulse dynamics in the IFFL1_2; changing a single base in rR4 removed this interaction (Supplementary Fig. 39), so that after this change, the IFFL1_2 produced two sequential pulses similar to those predicted by the general model (Fig. 3f). This rR4 sequence variant was used in all subsequent networks. Using the general model as a guide (Supplementary Fig. 37), we found we could delay the IFFL2's pulse relative to the IFFL1 pulse in the IFFL1_2 by decreasing the concentration of the G3C5 node that connects the IFFL modules together or increasing the dB5 concentration (Fig. 3g and Supplementary Fig. 40).

Engineering multi-stability in mutually repressive networks

We next sought to determine whether we could use standard genelet nodes to design feedback networks with predictable behavior. As a test, we chose to construct multi-stable memory networks where transient signals induce sustained changes in signal expression until new signals are introduced^{3,4,7,43}. Such networks employ a mutually repressive feedback architecture in which all nodes repress one another^{6,7} and multi-stability requires the feedback strengths between nodes to be fairly uniform, as unbalanced feedback strengths between two nodes in a mutually repressive network can lead to mono-stability (Supplementary Fig. 44). Increasing the number of states, and therefore the number of nodes in a mutually repressive architecture exacerbates the possibility of creating a network with less than the desired number of stable states (Supplementary Figs. 44 and 45) from variation in reaction rates or crosstalk between components. Switchable mutually repressive bi-stable modules have been developed and characterized for a number of synthetic GRN analogs^{14,19,44}, including genelets^{19,30}. We sought to test whether standardized genelets could facilitate straightforward engineering of switchable networks with more than two states.

We began by designing a tri-stable network (TSN) composed of three interlocked mutually repressive bi-stable modules (BSM1-3) (Fig. 1a and Supplementary Section 9.1). Kinetic simulations using the general genelet model predicted that for suitable component concentrations and the rate constants assumed in our general model, this network would maintain each of three stable states indefinitely after being initialized in them (Supplementary Figs. 46-48). Consistent with these predictions, the designed network maintained each initial state for over 8 hours in experiments (Fig. 4b).

Simulations suggested that inducer RNAs³⁰ that bind to and inactivate specific RNA repressors could switch the network's state by turning a repressed node ON, which would subsequently shut off the nodes that were active in the original state (Supplementary Figs. 49-50 and Fig. 4c). We asked whether an inducer RNA design developed for KWW genelets³⁰ that relies on partial hybridization of the inducer to its target RNA repressor could be adopted to create a standard inducer RNA design for HPC50 genelets (Supplementary Figure 42). Indeed, standard inducer designs complementary to 16 bases of their target repressors could induce all six possible



single state changes in the TSN (Fig. 4d) and could induce two (Fig. 4e) and three (Supplementary Figs. 52) state changes when added in sequence.

Figure 4 | A tri-stable network (TSN) composed of three mutually repressive bi-stable modules (BSMs). (a) The TSN architecture. G1D, G3D, and G4D are reporting nodes that monitor network state. As all TSN interactions are repressive, DNA blockers were omitted and HPC50 genelets (denoted with *), which lack blocking toeholds (BTH in Fig. 1c), were used. (b) Normalized activation levels of reporting nodes after TSN initialization in each of its three stable states. (c) Schematic of the 6 possible TSN single state changes and how they can be triggered by inducer RNAs that inhibit specific RNA repressors. rI1, rI3, and rI4 direct switches to States 1, 2, and 3, respectively. (d) Normalized activation levels of reporting nodes during each possible state change. Inducer RNAs were added after 1 h in the initial state (green arrows) to final concentrations of 10 μ M. Switching into State 2 occurs faster than predicted in simulations, possibly due to a high rR3 degradation rate or rI3 actively removing rR3 from dA3. (e) Normalized activation levels of reporting nodes during sequential state changes. Inducer RNA concentrations are in Supplementary

Section 9.2. TSN network design notes and sequences are in Supplementary Section 9.1. Experimental methods are in Supplementary Section 9.2.

Scalable network engineering through module integration

We next asked whether we could integrate feedforward and feedback modules into larger composite networks whose dynamics could be predicted by the behaviors of their component modules. We first designed a network in which the two possible outputs of a bi-stable module (BSM4) triggered two different temporal pulses^{3,4} (Fig. 5a) produced by one of two different IFFL modules (Fig. 5b), termed the bi-stable incoherent IFFL1 OR 2 (BS_IFFL1|2) network. When initialized in each of its two states, the BS_IFFL1|2 network triggered the respective IFFL, producing output pulses similar to those produced by the IFFL modules in isolation (Fig. 5c and Fig. 3). The BS_IFFL1|2 network's state could be switched by standard inducer RNAs, and switching dynamics were similar to those of the BSM4 in isolation (Supplementary Fig. 54). These state changes triggered the designed IFFL modules (Fig. 5d).

We next sought to integrate the BS IFFL1|2 into even larger networks by adding both upstream and downstream connections. The IFFL pulses might transiently trigger other downstream processes, whereas activation of upstream nodes might induce changes in BSM4 state. Thus, to demonstrate both upstream and downstream integration of the BS IFFL1|2 network, we programmed the pulse of IFFL1 to trigger a state change in the BSM4 (Fig. 5e). We first added an induction module³⁰ (IM) composed of two nodes whose respective outputs are the two inducer RNAs that change the network's state. Transient activation of each IM node triggered the designed state change (Supplementary Fig. 55). We then introduced a second output node (G1C9) into IFFL1 that, by producing an input to the IM, triggers the network to switch from State 1 to State 2 (Fig. 5f). We termed this network the induction BS IFFL1/2 with IFFL1 feedback (I BS IFFL1/2 FB1). Using simulations, we identified concentrations of G1C9 and the IM nodes predicted to cause I BS IFFL1|2 FB1 initialized in State 1 to transiently activate G9I8, thus inducing a switch into State 2 as designed (Supplementary Fig. 56). In experiments, the network orchestrated the designed dynamics (Fig. 5g), and the timing of the autonomous switch out of State 1 almost perfectly matched the timing predicted by the general model. The time at which each IFFL's output pulse reached its maximum height deviated from the times predicted by the simulation by less than 40 minutes (Supplementary Fig. 57). We also were able to tune the time spent in State 1 by changing the concentration of G1C9, as predicted by the general genelet model (Supplementary Fig. 58).

The I_BS_IFFL1|2_FB1 network used 10 of the 11 standardized nodes identified during screening and exhibited dynamics closely mirroring those predicted by the general genelet model, providing compelling evidence for the orthogonality of these standardized nodes. Finally, to ask whether more orthogonal domains might be added to the standard node library, we followed the screening workflow presented in Figure 2 to identify six additional standardized domains (Supplementary 11.1). We used three of these domains to construct another IFFL module whose pulse dynamics were similar to IFFL1, 2, and 3 (Supplementary Section 11.2).



Figure 5 | Engineering mesoscale networks by integrating modules and programming additional interactions. (a) Schematic of a network in which different pulses are triggered after the network enters each of two stable states. (b) The BS_IFFL1|2 network, which implements the circuit in (a), consists of a bi-stable module (BSM4) coupled to two feedforward loops (IFFL1 or IFFL2). *s denote HPC50 genelet nodes, which lack blockers and BTH domains (Supplementary Section 1.1). (c,d) Normalized activation levels of BS_IFFL1|2 network reporting nodes after initialization of the network in its two stable states (c) and when there are state changes (d). Inducer RNAs were added (green arrows) to a final concentration of 10 μ M. The activation of G1S1 in State 2 may be due to G7 nodes not staying fully repressed (c). (e) The I_BS_IFFL1|2_FB1 network. The IFFL1 feeds back to an induction module (IM) to trigger a change to State 2. (f) Desired I_BS_IFFL1|2_FB1 behavior. (g) Normalized activation levels of I_BS_IFFL1|2_FB1 reporting nodes predicted by the general genelet model (left) and measured in experiments (right) after the network is initialized in State 1. BS_IFFL1|2 and I_BS_IFFL1|2_FB1 design notes and sequences are in Supplementary Sections 10.1 and 10.2, respectively. Experimental methods are in Supplementary Section 10.3.

DISCUSSION

Integration and recombination of parts with standard performance makes it possible to engineer powerful and dynamic behavior in a variety of physical systems. For example, computers are composed of integrated electrical circuits assembled from such parts. Here we create standardized parts for building dynamic biochemical networks by selectively screening for sequences with similar regulatory behavior and by designing against undesired interactions between components (crosstalk, autoinhibition, and autoactivation). Circuits could be assembled both by reorganizing the genelet nodes/modules tested here and by recombining the input and output domains to create new nodes. We expect that new nodes would behave reliably, as each of the 42 input/output combinations we tested had similar kinetics (Supplementary Section 12). Genelet module dynamics were fairly independent of the choice of nodes used to build the modules: we successfully constructed four unique incoherent feedforward loop (IFFL) modules and four unique bistable switch modules (BSMs). Genelet modules could readily be integrated and recombined into mesoscale networks that rival smaller viral GRNs in complexity and function⁴⁵. The I BS IFFL1|2 FB1 network integrated four functional modules and was composed of 10 orthogonal input domains and 15 total nodes. This network represents a significant increase in complexity compared to similar in vitro circuits that have been developed^{14,28-30}; it uses 5-fold more nodes than any genelet network without the HPC design.

The dynamics of the networks constructed from our standardized nodes were well predicted by a general model of genelet behavior. To facilitate future model-driven network design we developed a Python package for rapid simulation of networks with different topologies, initial conditions, and species concentrations that automatically generates the appropriate standardized genelet sequences for a given topology⁴⁶.

The standardized engineering toolbox we present could significantly expand the frontier of autonomous chemical systems. For example, sequential IFFL modules could be used to orchestrate hierarchical chemical synthesis⁴⁷ and the I_BS_IFFL1|2_FB1 could select or switch between multiple synthesis pathways in response to environmental cues, emulating the responsive chemical regulation of cellular metabolism. Further, genelet networks that rival the complexity of smaller viral genomes could be compartmentalized to create artificial cells⁴⁸. Such protocells could autonomously process and transmit chemical information to organize, maintain, or transform chemical reaction-diffusion patterns^{24,26}. The engineering pipeline and standardized circuit elements presented here should make building such complex chemical systems reliable and routine.

METHODS

Oligonucleotides, enzymes, and other reagents

DNA and RNA sequences for all networks and network elements are listed in the Supplementary Information. All oligonucleotides were purchased from Integrated DNA Technologies, Inc (IDT). Most genelet DNA strands used in the final networks were HPLC purified by IDT, unless otherwise stated in the Supplementary Information. Synthetic RNA oligonucleotides were ordered from IDT unpurified. Ribonucleotide triphosphates (NTPs) were purchased from ThermoFisher Scientific. T7 RNAP was purchased in bulk (300,000 units) from Cellscript (270 U/µL, C-T7300K). Yeast inorganic pyrophosphatase (YIPP) was purchased from New England Biolabs (NEB) (0.1 U/µL). RNase H was purchased from ThermoFisher Scientific (5 U/µL). Bovine Serum Albumin (BSA) was purchased from Sigma Aldrich (Cat# A3858). All genelets were annealed (held at 90°C for 5 minutes and then cooled to 20°C at 1°C/min) in NEB RNAPol reaction buffer (Cat# M0251S) with non-template and template strands at equimolar concentrations. For initially blocked genelets, the DNA blocker strand was present at 1.5x the concentration of the non-template strand. Supplementary Section 1 presents the 11 standardized input and output genelet sequences and describes the method for assembling new network connections using these domains.

Reaction conditions and data acquisition

Unless otherwise stated, network reactions were conducted at 37°C in NEB RNAPol reaction buffer supplemented with MgCl₂ (final concentration 30 mM), NTPs (ATP, UTP, CTP, GTP final concentration 7.5 mM each), and BSA (final concentration 0.1 mg/ml). BSA was included in the transcription mix to prevent the other enzymes from sticking to the walls of the reaction tubes. In addition to T7 RNA polymerase and RNase H, yeast inorganic pyrophosphatase (YIPP) was also included in reactions $(1.35 \times 10^{-3} \text{ U/}\mu\text{L})$ to extend the duration of the transcription reactions⁴⁹. Nearly all of the experiments in this study were conducted with T7 RNAP from a single bulk batch of T7 RNA polymerase purchased from Cellscript to minimize the need for recalibration due to batch-to-batch variation in enzyme activity^{12,19,28}. The bulk batch of T7 RNA polymerase was split into smaller aliquots (each aliquot was enough for 20-30 experiments) to minimize enzyme degradation from repeated removal from the freezer. Concentrations of the DNA components and enzymes used are tabulated in the experimental methods sections for each network in the Supplementary Information. All kinetic data was obtained in a quantitative PCR machine (Agilent Mx3005P) equipped with the standard filters: FAM/SYBR Green I (492nm-516nm), HEX/JOE/VIC (535nm-555nm), Cy3 (545nm-568nm), ROX/Texas Red (585nm-610nm), Cy5 (635nm-665nm). TYE665 and Cy5 were tracked with the Cy5 filter, HEX with the HEX filter, TEX615 with the ROX filter, and FAM with the FAM filter. Fluorescence measurements were taken every minute during the reactions. See Supplementary Section 13 for details on fluorescence data normalization procedures for all the different network experiments.

ACKNOWLEDGEMENTS

The authors thank Elisa Franco, Eiji Nakamura, Moshe Rubanov, and Pepijn Moerman for insightful conversations and comments on the manuscript. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1232825. This work was supported by the Department of Energy under Grant No. DE-SC001 0426 for materials and supply costs.

AUTHOR CONTRIBUTIONS

S.W.S and R.S. designed the research. S.W.S conducted most of the experiments and simulations. K.C. performed the experiments presented in Supplementary Section 11. M.N.

performed preliminary experiments for the study. J.O. and A.M. conducted the multi-stability simulations and analysis. S.W.S and R.S. wrote the paper with feedback from the other authors.

COMPETING INTERESTS

The authors declare no competing interests

DATA AVAILABILITY

All data and simulation code are available from the corresponding author upon reasonable request.

SOFTWARE AVAILABILITY

The general genelet model code, including scripts for the main text simulations, is available at: <u>https://github.com/sschaff6/general-genelet-model.git</u>.

REFERENCES

- 1. van Esch, J. H., Klajn, R. & Otto, S. Chemical systems out of equilibrium. *Chem. Soc. Rev.* **46**, 5474–5475 (2017).
- 2. van Roekel, H. W. H. *et al.* Programmable chemical reaction networks: emulating regulatory functions in living cells using a bottom-up approach. *Chem. Soc. Rev.* **44**, 7465–7483 (2015).
- 3. Schultz, D., Wolynes, P. G., Jacob, E. B. & Onuchic, J. N. Deciding fate in adverse times: Sporulation and competence in Bacillus subtilis. *Proc. Natl. Acad. Sci. USA* **106**, 21027–21034 (2009).
- 4. Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L. & Adhya, S. Switches in bacteriophage lambda development. *Annu. Rev. Genet.* **39**, 409–429 (2005).
- Davidson, E. H. *et al.* A genomic regulatory network for development. *Science* 295, 1669–1678 (2002).
- Peter, I. S. & Davidson, E. H. Assessing regulatory information in developmental gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **114**, 5862 (2017).
- 7. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461 (2007).
- 8. Ferrell, J. E., Jr & Ha, S. H. Ultrasensitivity part III: cascades, bistable switches, and oscillators. *Trends in Biochemical Sciences* **39**, 612–618 (2014).

- Ackermann, J., Wlotzka, B. & McCaskill, J. S. In vitro DNA-based predator–prey system with oscillatory kinetics. *Bull. Math. Biol.* 60, 329–354 (1998).
- 10. Montagne, K., Plasson, R., Sakai, Y., Fujii, T. & Rondelez, Y. Programming an in vitro DNA oscillator using a molecular networking strategy. *Mol. Syst. Biol.* **7**, 466 (2011).
- Semenov, S. N. *et al.* Rational design of functional and tunable oscillating enzymatic networks. *Nat. Chem.* 7, 160–165 (2015).
- 12. Kim, J. & Winfree, E. Synthetic in vitro transcriptional oscillators. *Mol. Syst. Biol.* 7, 465 (2011).
- 13. Montagne, K., Gines, G., Fujii, T. & Rondelez, Y. Boosting functionality of synthetic DNA circuits with tailored deactivation. *Nat. Commun.* **7**, 13474 (2016).
- Padirac, A., Fujii, T. & Rondelez, Y. Bottom-up construction of in vitro switchable memories. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3212–E3220 (2012).
- 15. Helwig, B., van Sluijs, B., Pogodaev, A. A., Postma, S. G. J. & Huck, W. T. S. Bottom-up construction of an adaptive enzymatic reaction network. *Angew. Chem. Int. Ed.* **57**, 14065–14069 (2018).
- 16. Subsoontorn, P., Kim, J. & Winfree, E. Ensemble bayesian analysis of bistability in a synthetic transcriptional switch. *ACS Synth. Biol.* **1**, 299–316 (2012).
- 17. Postma, S. G. J., te Brinke, D., Vialshin, I. N., Wong, A. S. Y. & Huck, W. T. S. A trypsin-based bistable switch. *Tetrahedron* **73**, 4896–4900 (2017).
- Genot, A. J. *et al.* High-resolution mapping of bifurcations in nonlinear biochemical circuits. *Nat. Chem.* 8, 760 (2016).
- 19. Kim, J., White, K. S. & Winfree, E. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Mol. Syst. Biol.* **2**, 68 (2006).
- 20. Kim, J., Khetarpal, I., Sen, S. & Murray, R. M. Synthetic circuit for exact adaptation and fold-change detection. *Nucleic Acids Res.* **42**, 6078–6089 (2014).
- 21. Mattia, E. & Otto, S. Supramolecular systems chemistry. *Nat. Nanotechnol.* 10, 111–119 (2015).

- 22. Whitesides, G. M. Reinventing chemistry. Angew. Chem. Int. Ed. 54, 3196–3209 (2015).
- Lehn, J.-M. Perspectives in chemistry—steps towards complex matter. *Angew. Chem. Int. Ed.* 52, 2836–2850 (2013).
- Zadorin, A. S. *et al.* Synthesis and materialization of a reaction–diffusion French flag pattern. *Nat. Chem.* 9, 990 (2017).
- 25. Gines, G. et al. Microscopic agents programmed by DNA circuits. Nat. Nanotechnol. 12, 351 (2017).
- 26. Dupin, A. & Simmel, F. C. Signalling and differentiation in emulsion-based multi-compartmentalized in vitro gene circuits. *Nat. Chem.* **11**, 32–39 (2019).
- Green, L. N. *et al.* Autonomous dynamic control of DNA nanostructure self-assembly. *Nat. Chem.* 11, 510–520 (2019).
- Franco, E. *et al.* Timing molecular motion and production with a synthetic transcriptional clock. *Proc. Natl. Acad. Sci. USA* **108**, E784–E793 (2011).
- 29. Meijer, L. H. H. *et al.* Hierarchical control of enzymatic actuators using DNA-based switchable memories. *Nat. Commun.* **8**, 1117 (2017).
- 30. Schaffter, S. W. & Schulman, R. Building in vitro transcriptional regulatory networks by successively integrating multiple functional circuit modules. *Nature Chemistry* **11**, 829–838 (2019).
- 31. Franco, E., Giordano, G., Forsberg, P.-O. & Murray, R. M. Negative autoregulation matches production and demand in synthetic transcriptional networks. *ACS Synth. Biol.* **3**, 589–599 (2014).
- Kim, J., Hopfield, J. & Winfree, E. Neural network computation by in vitro transcriptional circuits. in Advances in Neural Information Processing Systems 17 (eds. Saul, L. K., Weiss, Y. & Bottou, L.) 681– 688 (MIT Press, Boston, MA, 2005).
- Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170– 173 (2011).

- 34. Dabby, N. Thesis: Synthetic molecular machines for active self-assembly: prototype algorithms, designs, and experimental study. (California Institute of Technology, 2013).
- Groves, B. *et al.* Computing in mammalian cells with nucleic acid strand exchange. *Nat. Nanotechnol.* **11**, 287–294 (2016).
- 36. Isambert, H. The jerky and knotty dynamics of RNA. *Methods* 49, 189–196 (2009).
- Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA* **100**, 11980–11985 (2003).
- Goentoro, L., Shoval, O., Kirschner, M. W. & Alon, U. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol Cell* 36, 894–899 (2009).
- Zhang, C., Tsoi, R., Wu, F. & You, L. Processing Oscillatory Signals by Incoherent Feedforward Loops.
 PLoS Comput Biol 12, e1005101–e1005101 (2016).
- 40. Rahman, A., Tiwari, A., Narula, J. & Hickling, T. Importance of Feedback and Feedforward Loops to Adaptive Immune Response Modeling. *CPT Pharmacometrics Syst Pharmacol* **7**, 621–628 (2018).
- Krupp, G. RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene* **72**, 75–89 (1988).
- Lapham, J. & Crothers, D. M. RNase H cleavage for processing of in vitro transcribed RNA for NMR studies and RNA ligation. *RNA* 2, 289–296 (1996).
- Fang, D. & Zhu, J. Dynamic balance between master transcription factors determines the fates and functions of CD4 T cell and innate lymphoid cell subsets. *Journal of Experimental Medicine* 214, 1861–1876 (2017).
- 44. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, (2000).
- 45. Mahmoudabadi, G. & Phillips, R. A comprehensive and quantitative exploration of thousands of viral genomes. *eLife* **7**, e31955 (2018).

- 46. Schaffter, S. W. General Genelet Model: https://github.com/sschaff6/general-genelet-model.git.(2020).
- 47. O'Reilly, R. K., Turberfield, A. J. & Wilks, T. R. The evolution of DNA-templated synthesis as a tool for materials discovery. *Acc. Chem. Res.* **50**, 2496–2509 (2017).
- Lyu, Y. *et al.* Protocells programmed through artificial reaction networks. *Chem. Sci.* **11**, 631–642 (2020).
- 49. Cunningham, P. & Ofengand, J. Use of inorganic pyrophosphatase to improve the yield of in vitro transcription reactions catalyzed by T7 RNA polymerase. *BioTechniques* **9**, 713–714 (1990).