ARTICLES
https://doi.org/10.1038/s41557-022-01001-3
nature chemistry

Check for updates

Standardized excitable elements for scalable engineering of far-from-equilibrium chemical networks

Samuel W. Schaffter[1], Kuan-Lin Chen[1], Jackson O'Brien[2], Madeline Noble[1], Arvind Murugan[2] and Rebecca Schulman[1,3,4] ✉

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. Currently, such networks must be extensively tuned to compensate for the different activities of and unintended reactions between a network's various chemical components. Modular elements with standardized performance could be used to rapidly construct networks with designed functions. Here we develop standardized excitable chemical regulatory elements, termed genelets, and use them to construct complex in vitro transcriptional networks. We develop a protocol for identifying >15 interchangeable genelet elements with uniform performance and minimal crosstalk. These elements can be combined to engineer feedforward and feedback modules whose dynamics match those predicted by a simple kinetic model. Modules can then be rationally integrated and organized into networks that produce tunable temporal pulses and act as multistate switchable memories. Standardized genelet elements, and the workflow to identify more, should make engineering complex far-from-equilibrium chemical dynamics routine.

Cellular processes such as stress response[1,2] and morphogenesis[3] are orchestrated by genetic regulatory networks (GRNs)[3,4] that continuously consume energy[5,6] to operate far-from-equilibrium. Energy dissipation allows GRNs to repeatedly produce adaptive, dynamic chemical responses to environmental stimuli that change over time. Genes in a GRN typically have nonlinear, often excitable, input–output responses in which small changes in the level of an input induce large changes in the amount of output produced. These properties allow GRNs to efficiently propagate signals[7] and, in principle, to orchestrate arbitrarily complex dynamic responses using networks which couple a sufficient number of modular genes[8]. Synthetic chemistries[9–20] with these same properties could be used for programming time-varying chemical signals[5,6], directing pattern formation[21–23] or regulating nanostructure assembly or reconfiguration[24,25].

Cellular GRNs that direct complex cellular processes often comprise 10–100 genes. A key step toward emulating the complex dynamics of these cellular GRNs in synthetic chemical networks is to scale up the number of circuit nodes that can operate together[3,4]. Such scale-up has been performed by adding newly designed nodes to existing circuits[14,26,27]. However, iterative design cycles are often required to find nodes that work as desired within an existing network. And nodes created specifically to add to existing networks often cannot be interchanged or rearranged to form new networks, meaning that building a new type of network can require starting such cycles of component and network design anew. As electronic circuit design methods suggest, a library of orthogonal nodes with standard input–output characteristics could make it possible to use a model-driven approach to rapidly construct many different networks from the same set of interchangeable units (Fig. 1a). For example, a range of complex chemical logic circuits can be routinely assembled with standardized biomolecular gates[28–32].

We sought to develop such a method for building far-from-equilibrium chemical networks composed of short transcriptional templates, called genelets[12,16,19,20,25,27,33]. Each genelet has an input (I) domain, which combines a DNA activator binding site and an incomplete T7 RNA polymerase (T7 RNAP) promoter site, and an output (O) domain, which encodes an RNA transcript. A genelet's output is transcribed when a DNA activator is bound to the genelet's input domain, that is, an activated (ON) state (Fig. 1b). Genelets regulate one another by transcribing RNAs that control the ability of other nodes' DNA activators to bind their respective 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 (an activator dA or blocker dB). DNA regulators are present in excess of their genelets, facilitating nonlinear, excitable input–output behaviour[12,19] akin to the sigmoidal response commonly used for nodes of artificial neural networks[34]. RNA is degraded by RNase H, enabling signal turnover, and thereby sustaining a network's far-from-equilibrium operation (Fig. 1c–e). However, it has been challenging to build large genelet networks with predictable behaviour because of spurious side reactions between components and differences in the response characteristics of nodes with different sequences[12,27].

Here we develop a scalable method for building genelet networks by constructing a library of interchangeable genelets with similar performance, that is, input–output behaviour and response times. We identify >15 standardized circuit elements, or regulatory domains, and show that these domains can each be combined into genelets that activate and/or repress one another. With this library we rapidly engineer a suite of reliable feedforward and feedback modules and then integrate these modules into mesoscale networks that produce pulses with tunable amplitudes and delays or act as switchable memories. Network design is informed by a general

[1]Department of Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA. [2]Department of Physics, University of Chicago, Chicago, IL, USA. [3]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [4]Department of Chemistry, Johns Hopkins University, Baltimore, MD, USA. ✉e-mail: rschulm3@jhu.edu

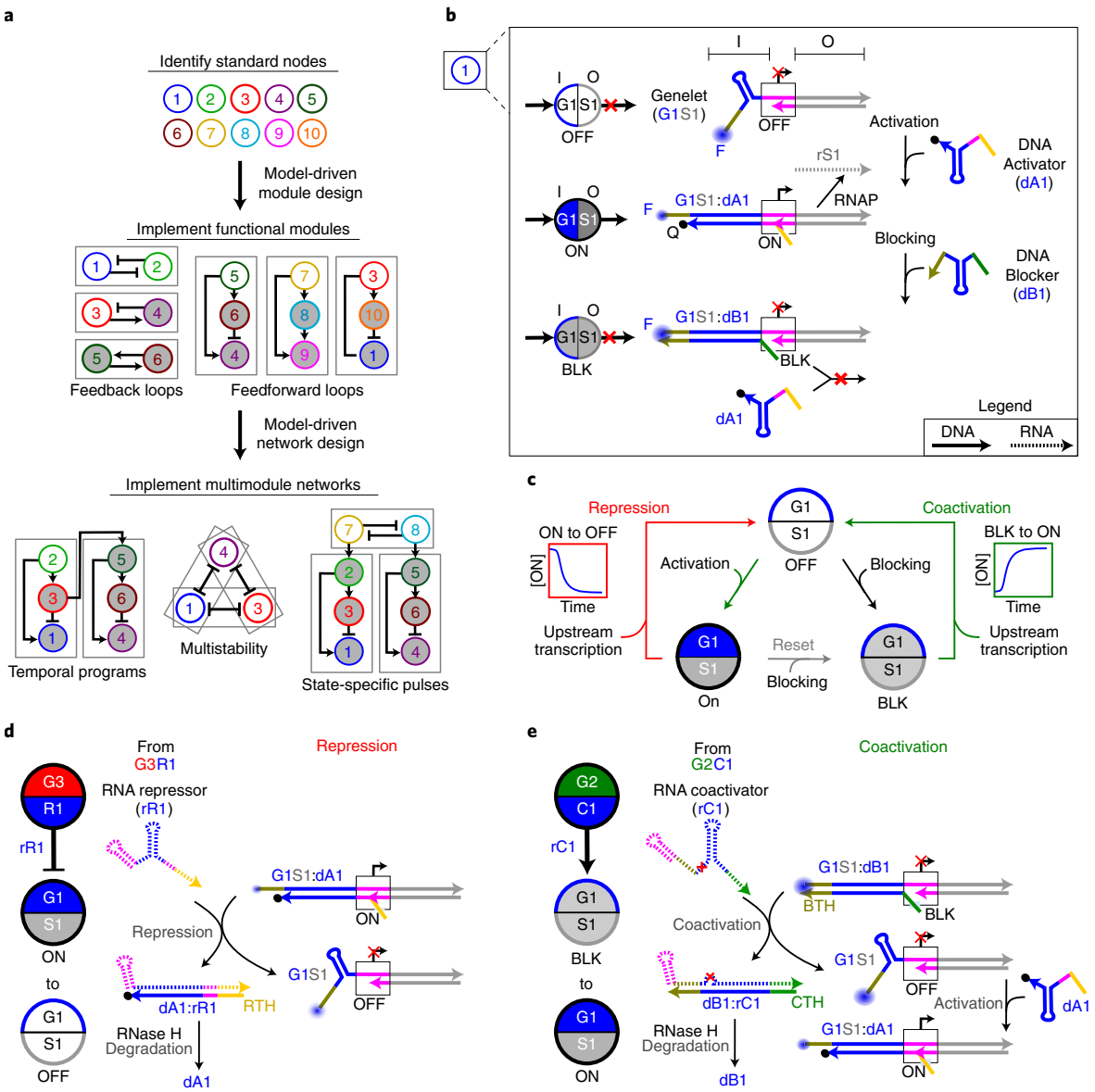


Fig. 1 | The hairpin clamp (HPC5) genelet toolbox. a, Network engineering workflow. **b**, Node states and species. Each genelet consists of an input (I) domain (G1 here), which controls transcription activity, and an output (O) domain, S1 here. The T7 RNAP promoter sequence (pink) of a genelet is not fully 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). Throughout this work, output domains that repress input domain *i* are labelled R*i* and those that coactivate input domain *i* are labelled C*i*, where $i = 1, 2, \ldots, N$. DNA species are depicted with solid lines and have d prefixes; RNA species are depicted as dashed lines and have r prefixes. Fluorophore (F) and quencher (Q) modifications are used to measure genelet state. **c**, Node state transitions. 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 degradation of RNA bound to a DNA activator or blocker by RNase H. Regulation pathways in the presence of both free activators and blockers are shown in Supplementary Fig. 2.

model of genelet behaviour that reliably predicts network dynamics. Finally, we expand the library of standardized domains using a standard protocol. These results suggest a reliable and predictive way of constructing chemical networks exhibiting a broad range of dynamic responses to time-varying chemical inputs. Such networks could potentially regulate self-assembly, biochemical synthesis or reactions, and orchestrate chemical detection or recognition.

Results

A modular design for bidirectionally regulated genelets. We sought to create a library of interchangeable and predictable genelet nodes with similar performance that could be bidirectionally regulated. We hypothesized that such a library would allow us to use a general model of genelet behaviour to rapidly design and implement regulatory networks with diverse dynamics. To build this standard library, we first sought to develop a genelet design that would minimize the impacts of unintended side reactions and genelet crosstalk. We began with a genelet design that sequesters most of each node's long input domain within a hairpin (hairpin clamp (HPC)). Hiding most of the recognition sequence within a hairpin minimizes crosstalk and permits the design of multiple modular units[19,27]. HPC nodes are repressed by upstream transcripts that react with a

genelet–activator complex via a 3′ ssDNA toehold (the repressor toehold (RTH)) to remove the DNA activator from the complex. However, the 3′ ssDNA overhang can facilitate promoter-independent transcription by T7 RNAP, which can lead nodes to spuriously turn off by themselves, a phenomenon termed autoinhibition[27]. To build genelets without this autoinhibition behaviour, we moved the ssDNA RTH domain of the HPC design to the 5′ end of the activator to create the HPC5o design (Supplementary Information, section 2 and Supplementary Fig. 1).

We then sought to modify the HPC5o design to allow upstream transcripts to activate downstream genelets in addition to repressing them. We first considered a design in which a genelet is off when its DNA activator is sequestered in a double-stranded DNA (dsDNA) complex[12,16,25]. An upstream RNA transcript could then release the DNA activator from this complex to turn the genelet on (Supplementary Information, section 3.1). However, a transcript that could release the DNA activator must share sequence elements with that activator; we found that these transcripts also competed with the activators to bind to the genelets, thereby preventing activation (Supplementary Fig. 7 and Supplementary Information, section 3.1).

We thus devised an alternative scheme for bidirectional regulation, the HPC5 design, in which a DNA blocker strand binds more strongly than a DNA activator to a genelet's input domain (Fig. 1b). Transcription can occur from a genelet bound to an activator, but not from a genelet bound to a blocker. Upstream RNA signals then downregulate genelet transcription by sequestering DNA activators and upregulate genelet transcription by sequestering DNA blockers (Fig. 1c–e). We confirmed that a DNA blocker prevented a DNA activator from binding to a genelet's input domain and that DNA and RNA coactivators could both facilitate genelet coactivation. We also verified that coactivation could be reversed by degrading the RNA coactivator or adding blocker in excess of the coactivator (Supplementary Information, section 3.2). Finally, we validated that HPC5 genelet transcription could be rapidly turned on by adding coactivators and off by adding repressors, and that the fraction of genelet in the ON state, measured via fluorescence (Fig. 1b), correlated linearly with transcription rate (promoter activity), as seen in previous studies[19] (Supplementary Information, section 3.4).

Standardized, interchangeable regulatory domains. Using the HPC5 design, we sought to create a library of modular genelet input and output sequences that could be interchangeably assembled into networks. We used NUPACK 3.2.2[35] to design sequences for 36 genelet input–output sequences predicted to both be thermodynamically stable in their designed hairpin structures and have minimal non-designed interactions with one another or with the G1 HPC5 node (Supplementary Information, section 4.1). In addition to orthogonality, we also sought sequences that could be fully activated and/or repressed at similar rates, such that a general model of genelet behaviour, which assumes the same reaction rates for every node, could reliably guide network design. DNA activators and/or blockers are used in excess of their respective genelets, so full activation or blocking is expected. The speed of genelet regulation should be governed by the rates of four-way branch migration. Previously reported rate constants for four-way branch migration[36,37] range from 10^3 to $10^4\,M^{-1}s^{-1}$, suggesting that >0.9 fraction of 25 nM genelet should activate in response to 150 nM of DNA activator in 15–100 min. Thus, we sought to identify sequences that each activated to >0.9 fraction ON in <1 h. We reasoned that this criterion represented the distribution of rates for different sequences, and that sequences that switched at these rates would be similar enough to follow the predictions of our general genelet model. Because four-way branch migration rates have not been extensively studied for diverse sequences, we first tested the assumption that four-way branch migration rates for HPC5 genelets would generally occur within the expected range. We developed a quick and inexpensive screening assay to measure the activation rates of our 36 designed genelets with excess of their DNA activators (Supplementary Information, section 4.2). Surprisingly, many activator sequences reacted with their genelets either incompletely or much more slowly than predicted: only 16 reached >0.9 fraction ON in <1 h (Fig. 2b and Supplementary Fig. 20). Incomplete or slow activation of a genelet–activator pair might be caused by undesired secondary structures within a genelet or activator or the existence of reactions other than the designed four-way branch migration activation process, and we deemed such sequences unreliable for network construction. To explore sequence criteria that led to incomplete or abnormally slow activation, we designed chimeric activators by swapping subdomains of fast and slow sequences and measured their activation rates with cognate genelets. These experiments indicated that the hairpin stem sequence of the activators was important in determining reaction rate (Supplementary Fig. 21) but we were not able to identify specific sequence patterns that reliably indicated a slow-activating sequence (Supplementary Fig. 22).

We next tested the coactivation and repression kinetics of the 16 nodes that passed DNA activation screening (Fig. 2c and Supplementary Information, section 4.3). We monitored this switch by adding DNA coactivator or repressor in excess of their DNA blocker or activator, so that these factors should completely switch their target nodes ON/OFF. All 16 nodes reached >0.9 fraction ON/OFF in <1 h when switched by DNA coactivators and repressors. When we monitored the rate of switching in response to the transcription of RNA coactivators or repressors, however, we found that many nodes switched incompletely or at least 2-fold more slowly than in response to the corresponding DNA sequence. We found six of these nodes could be switched >0.9 fraction ON and OFF in <1 h in response to transcribed RNA regulators and five could be rapidly regulated in one direction but not the other (Fig. 2d and Supplementary Information, section 4.4). Varying the concentration of the RNA regulator transcription template did not speed up the regulation of slow nodes, suggesting that incomplete RNA transcription or RNA misfolding[38], rather than differences in transcription rates, caused slow regulation (Supplementary Fig. 26).

To test whether the input and output domains (Fig. 1b) of the 11 nodes we identified could be used interchangeably, we measured the rates at which 28 unique input–output combinations coactivated or repressed downstream reporting genelets. Twenty-seven of these combinations took <1 h to switch their targets to >0.9 ON/OFF, suggesting the choice of input domain did not greatly influence the rate of downstream regulation (Fig. 2e and Supplementary Information, section 4.5). In these experiments, regulation is the result of coupled transcription and strand displacement. To characterize the variation in transcription rate across different nodes, we measured the transcription rates of nodes with two different input domains, G1 and G4, and found the transcription rates were within a factor of two of one another (Supplementary Information, section 3.4). A limitation of the HPC design is that four-way branch migration is slower than three-way branch migration with similar toehold lengths[39]. However, we found a 10-fold higher genelet concentration could reduce switching time 4-fold; this is comparable to the speed-up predicted to occur if one increased these strand displacement rate constants by two orders of magnitude (Supplementary Information, section 4.6).

Engineering temporal genelet expression programs. We next investigated whether the standardized nodes we identified could be integrated into regulatory networks whose dynamics could be predicted and programmed using a general model of genelet dynamics. The model consisted of ordinary differential equations describing the mass action kinetics of the designed genelet reactions for a given network architecture and assumed the kinetic rate constants were the same for each node (Supplementary Information, section 5).

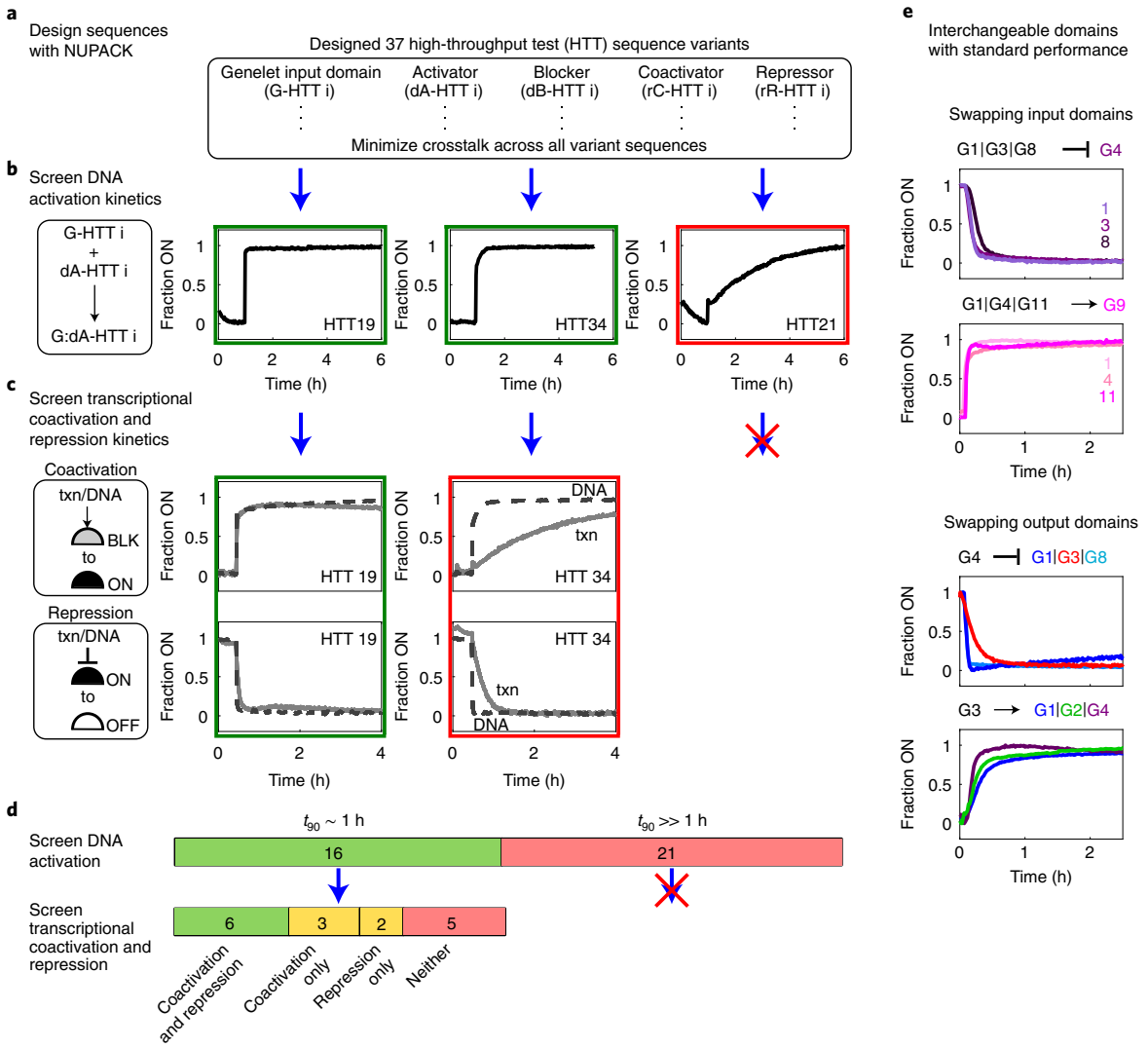


Fig. 2 | Design and screening protocol for identifying sequences for standardized HPC5 genelet domains. a, A library of genelet node sequences, termed HTT variants, are designed computationally (using NUPACK[35]) by minimizing undesired secondary structure interactions between library sequences. **b**, HTT genelet variants that reach >0.9 fraction ON in <1h after addition of their DNA activators (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 <1h 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. Supplementary Information, section 4 contains detailed descriptions of screening protocols. **e**, The standardized HPC5 genelet input and output domains can be interchanged without notable changes in regulatory kinetics. See Supplementary Information, section 4.5 for individual kinetic trajectories, experimental conditions, and additional input and output domain combinations.

We first used the model to design incoherent type 1 feedforward loop (IFFL)[3,4,40] modules that produce pulses of genelet activation via coordinated coactivation and repression. In a genelet IFFL (Fig. 3a), an output node (in IFFL1, G1) should pulse in response to the activators of the input nodes (in IFFL, dA2). Simulations indicated that the relative rates and strengths of G1 coactivation and repression determine pulse shape and timing (Supplementary Information, section 6.1), which we verified in experiments by varying the concentrations of IFFL components (Supplementary Fig. 37).

However, we found that the BLK G3R1 node in IFFL1 turned on even without an input, which was not predicted by our model. In isolation, BLK G3R1 also turned itself on, a phenomenon we called autoactivation. Autoactivation occurred at a rate proportional to its blocker (dB3) concentration, suggesting dB3 could be transcribed via promoter-independent transcription[41]. A dB3 strand with a single-stranded 3′ end composed of 2′ methylated RNA rather than DNA[42], which should prevent promoter-independent transcription initiated at the 3′ end, eliminated BLK G3R1 autoactivation (Supplementary Fig. 40). An IFFL with methylated dB3 produced pulses whose heights and durations could be reliably tuned by varying DNA activator concentrations according to model predictions (Fig. 3b). After identifying G3 autoactivation, we tested the propensity for autoactivation of other nodes and found that G8 and G10 nodes also exhibited autoactivation. We also eliminated this autoactivation by replacing the single-stranded DNA at the 3′ ends of their blockers with methylated RNA (Supplementary Information, section 7).

To test whether genelet nodes could be interchanged in IFFLs, we built two additional IFFL modules using other genelet domains from our library and measured their pulse dynamics using the same concentrations as in IFFL1. All three modules (IFFL1, IFFL2 and

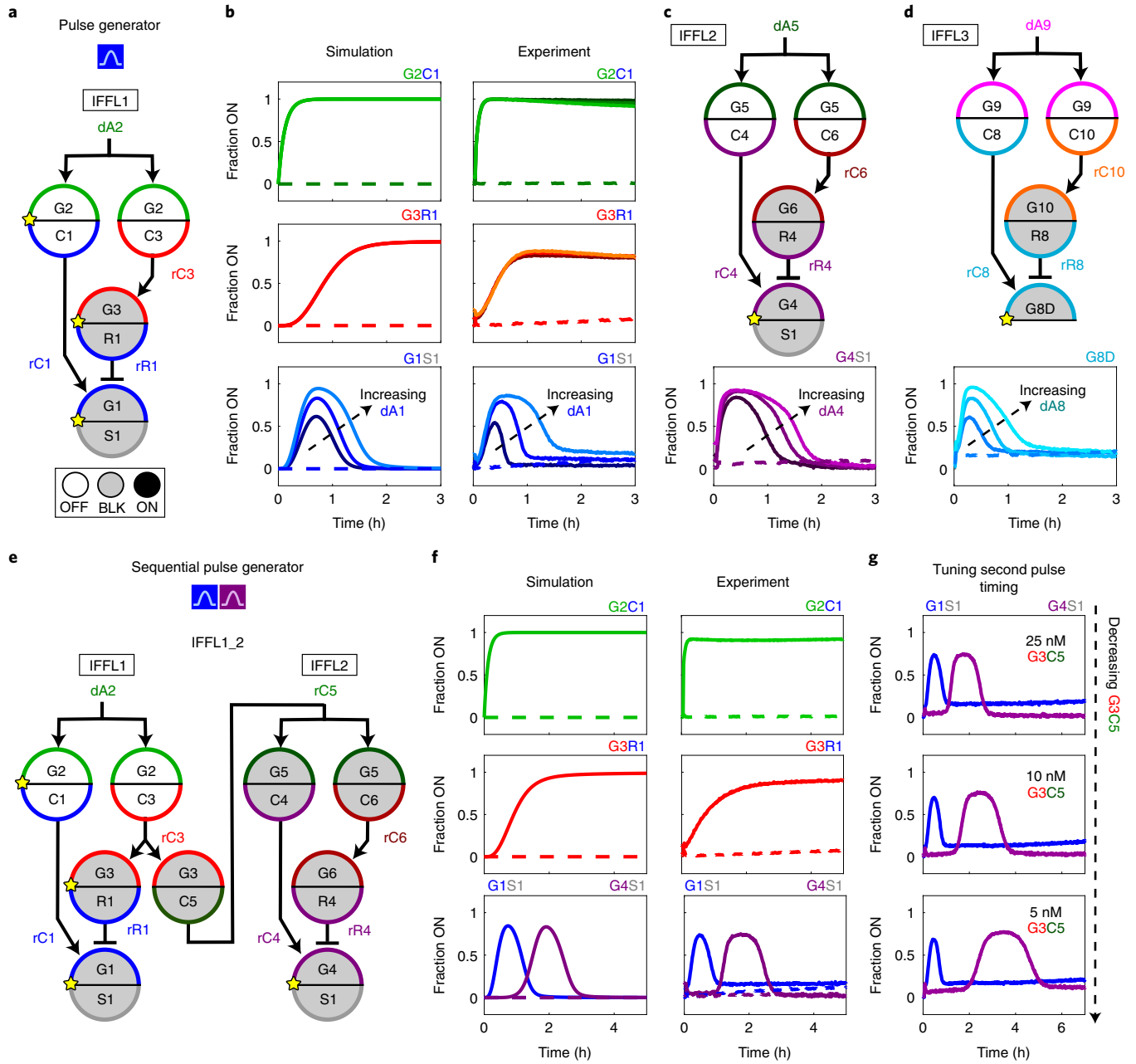


Fig. 3 | IFFLs orchestrate temporal pulses in genelet activation. a, The IFFL1 module. Here and elsewhere, a yellow star on a node indicates that the node was fluorescently modified to measure its activation levels during experiments. **b**, Normalized activation levels of IFFL1 nodes predicted by the general genelet model (left) and in experiments (right) for networks with 125, 250 or 500 nM dA1. Dashed curves in plots show activation levels when dA2 was not added. The general genelet model is described in Supplementary Information, section 5. **c**,**d**, Two IFFL modules, IFFL1 (**c**) and IFFL2 (**d**), 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 curves in plots show activation levels when dA5 (**c**) and dA9 (**d**) were not added. **e**, The IFFL1_2 network, consisting of IFFL1 and IFFL2 modules connected in series. **f**, Normalized IFFL1_2 node activation levels predicted by the general genelet model (left) and from experiments (right). **g**, Normalized activation levels of IFFL1_2 network nodes with different concentrations of G3C5. Detailed methods are in Supplementary Information, sections 6.2 and 8.2. IFFL1, IFFL2, IFFL3 and IFFL1_2 design notes and sequences are in Supplementary Information, sections 6.3, 6.5, 6.6 and 8.1.

IFFL3) exhibited pulse dynamics in agreement with model predictions (Fig. 3b–d).

We next asked whether we could connect multiple IFFL modules into a composite network that our model predicted would produce sequential pulses[1,4] (IFFL1_2, Fig. 3e,f). As a step toward constructing this network, we connected IFFL2 to an upstream node (G3C5). The resulting network, G3C5_IFFL2, produced a pulse in the absence of an input. We hypothesized this spurious pulsing was caused by leak transcription from the BLK G5 genelets, possibly due to transient DNA activator binding (Supplementary Information, section 8.3). Incorporating a 5% leak transcription rate from BLK G5 nodes in the model recapitulated the G3C5_IFFL2's spontaneous activation. Guided by this model, we eliminated spurious activation by increasing the IFFL2 blocker concentrations, which serves as a threshold for upstream transcriptional leak (Supplementary Figs. 43 and 44), and built an IFFL1_2 network that generated sequential

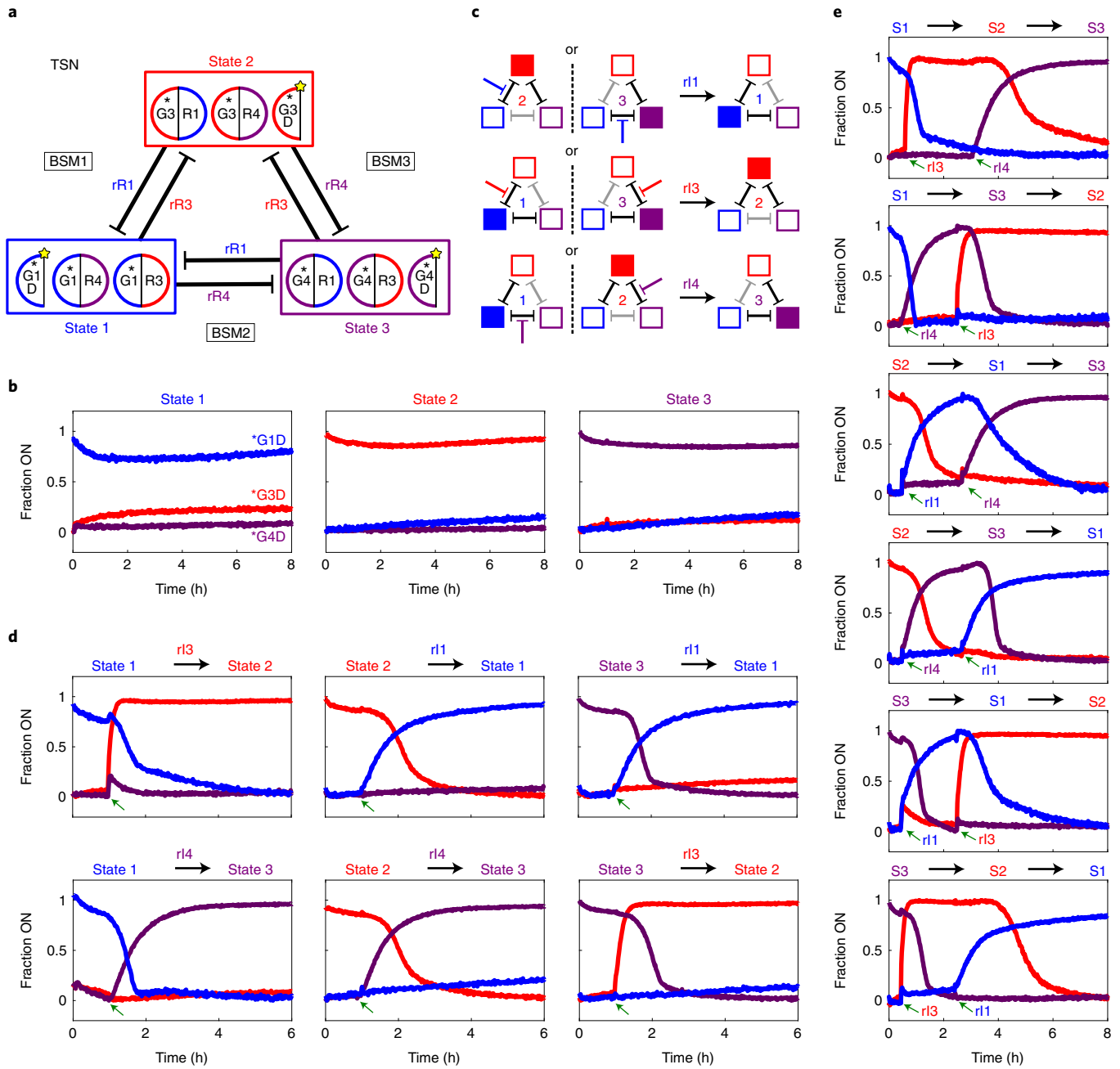


Fig. 4 | A TSN composed of three mutually repressive BSMs. a, The TSN architecture. G1D, G3D and G4D are reporting nodes that monitor network state. As all TSN interactions are repressive, DNA blockers were omitted and HPC5o genelets (denoted with *), which lack blocking toeholds (BTH in Fig. 1c), were used. **b**, Normalized activation levels of reporting nodes after TSN initialization in each of its three stable states. **c**, Schematic of the six 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 μM. Switching into state 2 occurs faster than predicted in simulations, possibly due to a high rR3 degradation rate or to rI3 actively removing rR3 from dA3. **e**, Normalized activation levels of reporting nodes during sequential state changes. Inducer RNA concentrations are in Supplementary Information, section 9.2. TSN network design notes and sequences are in Supplementary Information, section 9.1. Experimental methods are in Supplementary Information, section 9.2.

IFFL pulses (Fig. 3f). We further found we could use the general genelet model (Supplementary Fig. 45) to devise multiple means to delay the second pulse: decreasing the G3C5 concentration (Fig. 3g) or increasing the dB5 concentration (Supplementary Fig. 46).

Engineering multistability with mutually repressive networks. We next investigated whether standard genelet domains could be reorganized into feedback networks with predictable behaviour. We chose to construct multistable memory networks in which transient signals can induce sustained changes in signal expression[1,2,4]. Memory is a hallmark of far-from-equilibrium networks, as a continuous energy supply is required to keep a network in one of several stable steady states[3,4]. A mutually repressive feedback network can be multistable[3,4] so long as the feedback strength of all

connections is fairly uniform (Supplementary Figs. 50–52). Simulations predict that as the number of mutually repressive states increases, feedback strengths must become more and more uniform to achieve multistability. Thus, nodes with standardized performance are critical for constructing large multistable networks (Supplementary Figs. 50 and 51). Many synthetic GRN analogues[14,19,43], including genelets[19,27], have implemented switchable, mutually repressive bistable modules. To test the interchangeability and predictability of our standardized nodes, we sought to engineer a switchable tristable network (TSN).

We designed a TSN composed of three interlocked mutually repressive bistable modules (BSM1–3) (Fig. 4a and Supplementary Information, section 9.1) that simulations predicted could maintain each of three stable states after being initialized in them (Supplementary Figs. 52–54). Consistent with these predictions, the network maintained each initial state for over 8 h in experiments (Fig. 4b). To demonstrate switchable memory, we designed inducer RNAs that bound to 16 bases of their target RNA repressors[27] (Supplementary Information, section 9.1). Simulations indicated these inducer RNAs could switch network state by turning a repressed node ON, which would subsequently shut off the active nodes of the initial state (Supplementary Figs. 55–56 and Fig. 4c). The inducers triggered all six possible single state changes (Fig. 4d), and sequential inducer additions orchestrated two (Fig. 4e) and three (Supplementary Fig. 58) state changes.

Scalable network engineering through module integration. We next asked whether we could integrate feedforward and feedback modules into larger composite networks whose dynamics were predicted by the behaviours of their component modules. We first designed a network in which the two different outputs of a bistable module (BSM4) triggered different IFFL pulses[1,2] (Fig. 5a,b), termed BS_IFFL1|2. When initialized in each of its two states, BS_IFFL1|2 produced output pulses similar to those produced by IFFL1 and IFFL2 in isolation (Figs. 5c and 3). The BS_IFFL1|2 network's state could be switched by standard inducer RNAs (Fig. 5d), at rates similar to BSM4 alone (Supplementary Fig. 60).

We next incorporated upstream and downstream connections into the BS_IFFL1|2 network (Fig. 5e). We first added an induction module[27] (IM) composed of two nodes that produce the inducer RNAs that switch the BSM4 state (Supplementary Fig. 61). We then introduced a second output node (G1C9) into IFFL1. This second output produces a coactivator that triggers the IM to switch BSM4 from state 1 to state 2 (Fig. 5f). We termed the resulting network I_BS_IFFL1|2_FB1. Using simulations, we identified G1C9 and IM node concentrations predicted to cause I_BS_IFFL1|2_FB1 initialized in state 1 to transiently activate G9I8, which induces a switch into state 2 (Supplementary Fig. 62). In experiments, the pulsing and switching occurred as predicted (Fig. 5g), with the timing of the autonomous switch out of state 1 closely matching model predictions. The time at which each IFFL's output pulse reached its maximum height deviated from model predictions by ~40 min (Fig. 5g and Supplementary Fig. 63). We also tuned the time spent in state 1 by changing the G1C9 concentration (Supplementary Fig. 64). These results provide compelling evidence for the orthogonality of individual nodes and larger modules.

The I_BS_IFFL1|2_FB1 network used 10 of the 11 standardized nodes identified during screening. To expand the node library, we followed our screening workflow (Fig. 2) to identify six additional standardized domains (Supplementary Information, section 11.1). We used three of these domains to successfully construct an IFFL4 module (Supplementary Information, section 11.2).

Discussion

Integration and recombination of parts with standard performance makes it possible to engineer powerful and dynamic behaviour into 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 designing against undesired interactions between components (crosstalk, autoinhibition, and autoactivation) and selectively screening for sequences with similar regulatory behaviour. Modules could be rapidly assembled from genelet library nodes and networks could be rapidly assembled from modules without extensive characterization of new connections. We expect that new node and module combinations would also behave reliably, as each of the 42 input–output combinations we tested (Supplementary Information, section 12) met our screening criteria.

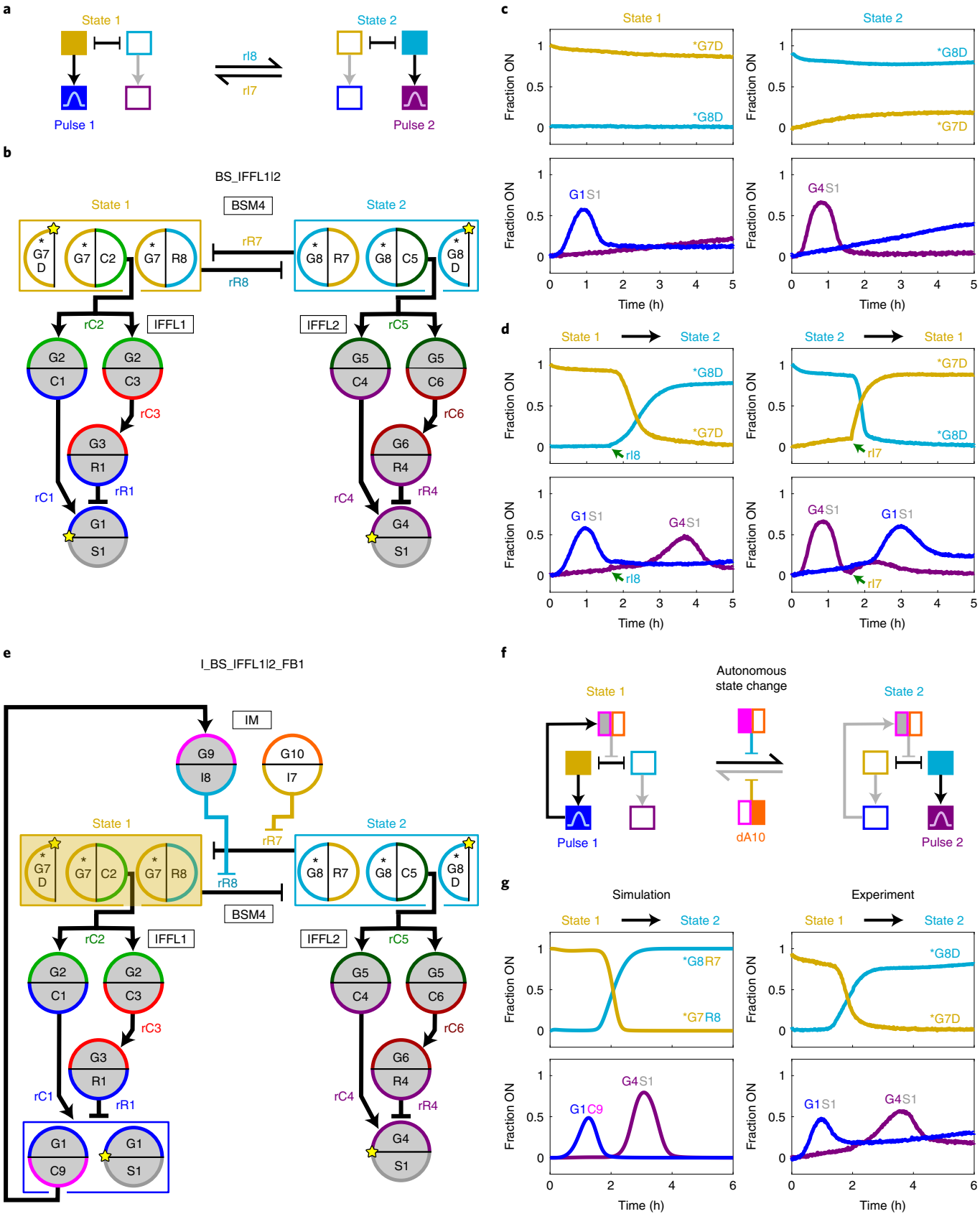
This standardized far-from-equilibrium network engineering toolbox should markedly expand the frontier of autonomous chemical systems by enabling the construction of mesoscale networks that rival smaller viral GRNs in complexity and function[44]. The I_BS_IFFL1|2_FB1 network could select or switch between multiple synthesis pathways in response to environmental cues, emulating the responsive chemical regulation of cellular metabolism, and sequential IFFL modules could be used to orchestrate hierarchical chemical synthesis[45]. Genelet networks might also be compartmentalized to create artificial cells[23] that could autonomously process and transmit chemical information to organize, maintain or transform chemical reaction–diffusion patterns[21,23].

To facilitate the assembly of other networks, we developed an open-source Python package for simulating networks with different topologies, initial conditions and species concentrations that automatically generates the appropriate standardized genelet sequences for a given topology[46]. We tested this software by designing a range of far-from-equilibrium networks, such as oscillators, persistence detectors, and logic gates, that could be assembled using our genelet library (Supplementary Information, section 13). To facilitate the transfer of network designs from simulations to experiments, we also compiled an experimental troubleshooting guide (Supplementary Information, section 14).

Previous estimates of the number of non-interacting HPC genelet domains[27], combined with the screening success we observed here, suggest the possibility of building a library of 30–50 non-interacting sequences with similar regulatory behaviour (Supplementary Information, section 15). While these innovations greatly increase the sizes of circuits that can be assembled—the I_BS_IFFL1|2_FB1 is 5-fold larger than any genelet network

Fig. 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 bistable module (BSM4) coupled to two feedforward loops (IFFL1 and IFFL2). Asterisks denote HPC5o genelet nodes, which lack blockers and BTH domains (Supplementary Information, 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 μM. 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. IFFL1 feeds back to an induction module (IM) to trigger a change to state 2. **f**, Desired I_BS_IFFL1|2_FB1 behaviour. **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 Information, sections 10.1 and 10.2, respectively. Experimental methods are in Supplementary Information, section 10.3.

developed without the HPC design[14,25–27]—other chemistries, regulation mechanisms or innovations such as compartmentalization[23,47] or localization[48] will be required to build networks the size seen in cellular genomes (Supplementary Information, section 15). Critical challenges to realizing such large networks are waste accumulation[12,19] and loss of enzyme activity[27], both of which limit the number of cycles a far-from-equilibrium system can undergo in batch reactions. These challenges scale with increasing



network size. Thus, the design of systems such as continuous-flow reactors or compartments[11,47,49] that can replenish or manage chemical fuel will be required before we can meaningfully construct synthetic networks that reach these scales.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-022-01001-3.

Received: 16 February 2021; Accepted: 16 June 2022;
Published online: 4 August 2022

References

1. 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).
2. Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L. & Adhya, S. Switches in bacteriophage lambda development. *Annu. Rev. Genet.* **39**, 409–429 (2005).
3. Peter, I. S. & Davidson, E. H. Assessing regulatory information in developmental gene regulatory networks. *Proc. Natl Acad. Sci. USA* **114**, 5862 (2017).
4. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).
5. van Esch, J. H., Klajn, R. & Otto, S. Chemical systems out of equilibrium. *Chem. Soc. Rev.* **46**, 5474–5475 (2017).
6. 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).
7. Ferrell, J. E.Jr & Ha, S. H. Ultrasensitivity part III: cascades, bistable switches, and oscillators. *Trends Biochem. Sci.* **39**, 612–618 (2014).
8. McAdams Harley, H. & Shapiro, L. Circuit simulation of genetic networks. *Science* **269**, 650–656 (1995).
9. 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).
11. 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).
14. Padirac, A., Fujii, T. & Rondelez, Y. Bottom-up construction of in vitro switchable memories. *Proc. Natl Acad. Sci. USA* **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).
18. 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. Zadorin, A. S. et al. Synthesis and materialization of a reaction–diffusion French flag pattern. *Nat. Chem.* **9**, 990 (2017).
22. Gines, G. et al. Microscopic agents programmed by DNA circuits. *Nat. Nanotechnol.* **12**, 351 (2017).
23. Dupin, A. & Simmel, F. C. Signalling and differentiation in emulsion-based multi-compartmentalized in vitro gene circuits. *Nat. Chem.* **11**, 32–39 (2019).
24. Green, L. N. et al. Autonomous dynamic control of DNA nanostructure self-assembly. *Nat. Chem.* **11**, 510–520 (2019).
25. Franco, E. et al. Timing molecular motion and production with a synthetic transcriptional clock. *Proc. Natl Acad. Sci. USA* **108**, E784–E793 (2011).
26. Meijer, L. H. H. et al. Hierarchical control of enzymatic actuators using DNA-based switchable memories. *Nat. Commun.* **8**, 1117 (2017).
27. Schaffter, S. W. & Schulman, R. Building in vitro transcriptional regulatory networks by successively integrating multiple functional circuit modules. *Nat. Chem.* **11**, 829–838 (2019).
28. Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196–1201 (2011).
29. Song, T. et al. Fast and compact DNA logic circuits based on single-stranded gates using strand-displacing polymerase. *Nat. Nanotechnol.* **14**, 1075–1081 (2019).
30. Kishi, J. Y., Schaus, T. E., Gopalkrishnan, N., Xuan, F. & Yin, P. Programmable autonomous synthesis of single-stranded DNA. *Nat. Chem.* **10**, 155–164 (2017).
31. Shah, S. et al. Using strand displacing polymerase to program chemical reaction networks. *J. Am. Chem. Soc.* **142**, 9587–9593 (2020).
32. Chen, Z. et al. De novo design of protein logic gates. *Science* **368**, 78 (2020).
33. 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).
34. Kim, J., Hopfield, J. & Winfree, E. in *Advances in Neural Information Processing Systems 17* (eds Saul, L. K., Weiss, Y. & Bottou, L.) 681–688 (MIT Press, 2005).
35. Zadeh, J. N. et al. NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
36. Dabby, N. *Synthetic Molecular Machines for Active Self-assembly: Prototype Algorithms, Designs, and Experimental Study*. PhD thesis, California Institute of Technology (2013).
37. Groves, B. et al. Computing in mammalian cells with nucleic acid strand exchange. *Nat. Nanotechnol.* **11**, 287–294 (2016).
38. Isambert, H. The jerky and knotty dynamics of RNA. *Methods* **49**, 189–196 (2009).
39. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).
40. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl Acad. Sci. USA* **100**, 11980–11985 (2003).
41. Krupp, G. RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene* **72**, 75–89 (1988).
42. 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).
43. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
44. Mahmoudabadi, G. & Phillips, R. A comprehensive and quantitative exploration of thousands of viral genomes. *eLife* **7**, e31955 (2018).
45. 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).
46. Schaffter, S. W. *General Genelet Model* (2020); https://github.com/sschaff6/general-genelet-model.git
47. Dubuc, E. et al. Cell-free microcompartmentalised transcription–translation for the prototyping of synthetic communication networks. *Curr. Opin. Biotechnol.* **58**, 72–80 (2019).
48. Chatterjee, G., Dalchau, N., Muscat, R. A., Phillips, A. & Seelig, G. A spatially localized architecture for fast and modular DNA computing. *Nat. Nanotechnol.* **12**, 920–927 (2017).
49. Laohakunakorn, N. et al. Bottom-up construction of complex biomolecular systems with cell-free synthetic biology. *Front. Bioeng. Biotechnol.* **8**, 213 (2020).

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 (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 U) from Cellscript (200 U μl^{-1}, C-T7300K). Yeast inorganic pyrophosphatase was purchased from New England Biolabs (NEB) (0.1 U μl^{-1}). RNase H was purchased from ThermoFisher Scientific (5 U μl^{-1}). Bovine serum albumin (BSA) was purchased from Sigma Aldrich (catalogue number A3858). All genelets were annealed (held at 90 °C for 5 min and then cooled to 20 °C at 1 °C per min) in NEB RNAPol reaction buffer (catalogue number M0251S) with non-template and template strands at equimolar concentrations. For initially blocked genelets, the DNA blocker strand was present at 1.5× the concentration of the non-template strand. Supplementary Information, 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 $MgCl_2$ (final concentration, 30 mM), NTPs (ATP, UTP, CTP, GTP; final concentration, 7.5 mM each), and BSA (final concentration, 0.1 mg ml^{-1}). 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 was also included in reactions (1.35×10^{-3} U μl^{-1}) to extend the duration of the transcription reactions[50]. To minimize the need for recalibration due to batch-to-batch variation in enzyme activity[12,19,25], T7 RNAP was purchased in bulk from Cellscript and three batches were used throughout the study. The bulk batches of T7 RNA polymerase were 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 polymerase chain reaction machine (Agilent Mx3005P) equipped with the standard filters: FAM/SYBR Green I (492–516 nm), HEX/JOE/VIC (535–555 nm), Cy3 (545–568 nm), ROX/Texas Red (585–610 nm), Cy5 (635–665 nm). TYE665 and Cy5 were measured 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 Information, section 16 for details on fluorescence data normalization procedures for all the different network experiments.

Data availability

The data associated with this manuscript are available at: https://doi.org/10.7281/T1/UBSZF1.

Code availability

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

References

50. 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).

Acknowledgements

The authors thank E. Franco, E. Nakamura, M. Rubanov and P. 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 number DGE-1232825 to S.W.S. This work was principally supported by the Department of Energy under award number DE-SC001 0426 to R.S. K.C. was supported by National Science Foundation award number EFRI-1830893 and Army Research Office award W911NF2010057. This work was also supported by the University of Chicago Materials Research Science and Engineering Center, which is funded by the National Science Foundation under award number DMR-2011854 to J.O. and A.M. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

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 Information, sections 11, 3.4 and 4.6. M.N. performed preliminary experiments for the study. J.O. and A.M. conducted the multistability 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.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41557-022-01001-3.

Correspondence and requests for materials should be addressed to Rebecca Schulman.

Peer review information *Nature Chemistry* thanks James Carothers and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.