ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

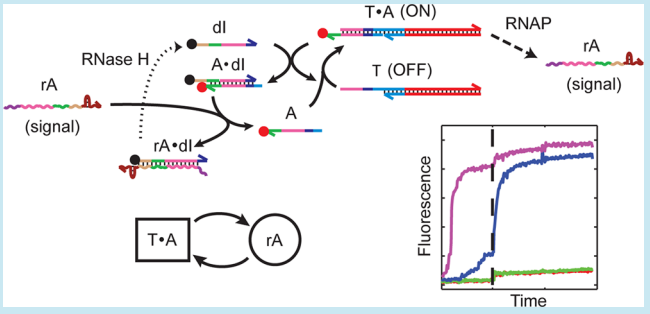
Ensemble Bayesian Analysis of Bistability in a Synthetic Transcriptional Switch

Pakpoom Subsoontorn,[†,⊥,¶] Jongmin Kim,[†,∥,⊥] and Erik Winfree*[,‡,§,∥]

Departments of [†]Biology, [‡]Computation and Neural Systems, [§]Computer Science, and [∥]Bioengineering, California Institute of Technology, Pasadena, California, 91125, United States

Supporting Information

ABSTRACT: An overarching goal of synthetic and systems biology is to engineer and understand complex biochemical systems by rationally designing and analyzing their basic component interactions. Practically, the extent to which such reductionist approaches can be applied is unclear especially as the complexity of the system increases. Toward gradually increasing the complexity of systematically engineered systems, programmable synthetic circuits operating in cell-free *in vitro* environments offer a valuable testing ground for principles for the design, characterization, and analysis of complex biochemical systems. Here we illustrate this approach using *in vitro* transcriptional circuits ("genelets") while developing an activatable transcriptional switch motif and configuring it as a bistable autoregulatory circuit, using just four synthetic DNA strands and three essential enzymes, bacteriophage T7 RNA polymerase, *Escherichia coli* ribonuclease H, and ribonuclease R. Fulfilling the promise of predictable system design, the thermodynamic and kinetic constraints prescribed at the sequence level were enough to experimentally demonstrate intended bistable dynamics for the synthetic autoregulatory switch. A simple mathematical model was constructed based on the mechanistic understanding of elementary reactions, and a Monte Carlo Bayesian inference approach was employed to find parameter sets compatible with a training set of experimental results; this ensemble of parameter sets was then used to predict a test set of additional experiments with reasonable agreement and to provide a rigorous basis for confidence in the mechanistic model. Our work demonstrates that programmable *in vitro* biochemical circuits can serve as a testing ground for evaluating methods for the design and analysis of more complex biochemical systems such as living cells.



KEYWORDS: *in vitro, synthetic biology, systems biology, reductionism, transcriptional circuits*

The ability to understand and reliably engineer complex molecular systems as sophisticated as biological cells will open up profound possibilities in biotechnology and nanotechnology.[1−7] Historically, forward and reverse engineering of complex systems has benefited from the principle of reductionism – the idea that complex phenomena can be understood by breaking them down into the simplest components, characterizing these components and how they interact with one another, and then conceptually building the system back up from these components to provide an explanation for the phenomena. For forward engineering, reductionism amounts to the idea that each system component can be rationally constructed and tested and that mechanistic models, built upon component-level understanding, can correctly guide how to put components together into systems with desirable behaviors. For simple biochemical systems, the reductionist approach has worked reasonably well. In contrast, when applied to complex biochemical systems such as living cells, the reductionist approach has been argued to have reached its limit due to intractable complexity, unknown architecture, and paucity of quantitative data.[8,9] While there may be limits for classical reductionist approaches in practice, it is our belief, however, that the current difficulties in fact stem from relative inexperience in successfully applying the reductionist approach combined with systems biology techniques to complex biochemical systems, because the available systems to study are either too simple (i.e., dead) or too complex (i.e., alive), leaving a "complexity gap" exactly where reductionist techniques could most effectively be tested and refined. This calls for establishing a training ground where one can incrementally push the limits of reductionism. Within such a training ground, each round of design and characterization should result in a small deviation from the predicted system behavior, so that the deviation's cause can be traced, the improved understanding formalized, and the complexity of the system continually increased.

Synthetic biology provides one such training ground by building and analyzing synthetic circuits that exhibit behaviors analogous to those of biological systems, while synthetic simplicity facilitates quantitative analysis as well as systematic

Received: March 23, 2012
Published: June 13, 2012

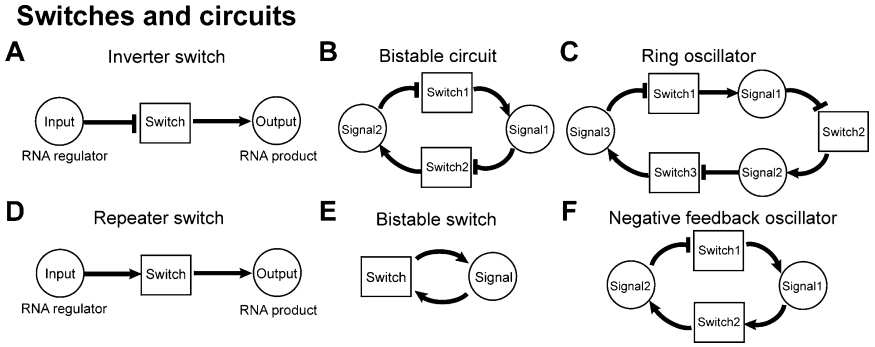


Figure 1. Systematic construction of a transcriptional inverter, a repeater, and circuits composed of these modular switch motifs. An inverter switch (A) and a bistable circuit (B) appeared in ref 37; a repeater switch motif (D) and oscillators (C,F) appeared in ref 38; the autoregulatory switch (E) is presented here for the first time as an independent circuit, although it appeared as a subcircuit of an oscillator in ref 38. Blunt ends indicate inhibition, while arrowheads indicate signal production or activation.

engineering design.[10−12] While *in vivo* synthetic biochemical systems have made significant progress, the daunting complexity of the cellular context remains a major obstacle.[13] *In vitro* (cell-free) synthetic biology has emerged as a powerful platform for studying and harnessing biochemical processes, independent from complex ancillary processes for cell viability and growth,[14,15] thus reducing the unintended interactions between the circuit of interest and its environment. These *in vitro* synthetic biochemical systems show different degrees of complexity ranging from cell extracts equipped with a complete set of cellular machineries[16−24] to simple reconstituted systems with only a few molecular species.[25−31] There is an apparent complexity-programmability trade-off, from simple reconstituted circuits that are difficult to rewire[28] to readily programmable circuits within relatively complex cell extracts.[18,24] Simple synthetic biochemical systems embedded in a relatively simple environment, yet readily programmable and capable of demonstrating complex dynamics, would provide an ideal test-bed for the reductionist approach in molecular biology.

We previously introduced programmable DNA-based synthetic biochemical circuits that are enzyme-free[32−35] or that require only a few purified enzymes[36−39] to operate. While the former operate in a simpler environment (DNA is the only macromolecule), such enzyme-free circuits require complex and specific DNA "fuels" for each type of implemented reaction, greatly increasing the complexity of synthesis and preparation. In contrast, the latter so-called *in vitro* transcriptional circuits provide a model system that shares architectural similarity with natural gene regulatory circuits and captures key features such as the use of production and degradation of macromolecules to represent signals (or to construct other molecular machinery) and the use of a universal small molecule fuel (NTP) to power all its functions.

Initial work on *in vitro* transcriptional circuits demonstrated individual switches functioning as inverters and showed their integration into circuits with bistable and oscillatory dynamics[37,38] (Figure 1, top). Subsequent work demonstrated a transcriptional repeater, a transcriptional switch whose RNA output level is a sigmoidal activation function of its RNA regulator (Figure 1, bottom). An activator-regulated switch (repeater) offers greater flexibility and simplicity for synthetic circuit design and can allow for faster timing and less noise relative to only inhibitor-regulated switches (inverters).[40] This repeater design motif was successfully integrated as part of a negative-feedback oscillator[38] that, in turn, has been used to drive other molecular processes such as DNA tweezers.[41,39]

The basic DNA switch, a simplified gene or "genelet", has a modular architecture that allows for independent design of the RNA product and the RNA regulator within a switch. Hence, one can "wire" several switches together to compose a complex regulatory network, in principle, by simply designing the RNA output of one switch to be the RNA regulator of another switch. Moreover, individual switch characteristics such as switching thresholds and maximum output levels can be set by the concentrations of switch components rather than by molecular characteristics of binding domains, facilitating tuning of circuit behavior. The state of each switch (transcription rate) and the levels of signals (RNA concentrations) define the instantaneous state of the circuit. As the circuit dynamics relies only on RNA transcription and degradation, our *in vitro* circuits operate in a relatively simple environment with NTP fuel and only three essential enzymes, T7 RNA polymerase (RNAP) and two ribonucleases (RNases). Despite this simplicity, arbitrarily large regulatory networks with complex dynamical behaviors can be implemented in principle.[36]

In this article, we describe the systematic design process for the repeater based on thermodynamic and kinetic considerations and experimentally validate each elementary reaction required for a functional repeater. We then construct and analyze a self-activating switch, a single repeater wired to itself, that can exhibit bistability. Furthermore, we show that the experimental results are consistent with a simple mathematical model, whose parameters are set by Bayesian inference from a subset of experimental data (the training set) and which were predictive of general trends in an independent set of experimental data (the test set). While our previous work employed a single least-squares-fit parameter set to explain the observed experimental data, the use here of the ensemble Bayesian inference approach allows us to assess the level of confidence in our estimated parameters. Moreover, by systematically dividing the experimental results into the training set and the test sets in this work, we can demonstrate that the mechanistic understanding of elementary reactions combined with system-level characterizations could be used to predict system behavior, illustrating successful application of a combined reductionist (bottom-up) and systems biology (top-down) approach. While this is not surprising in and of itself, in light of the potential to systematically construct more complex circuits it presents *in vitro* synthetic biology as a promising route for exploring the complexity gap that can serve as a testbed to improve our mastery of complex biochemical systems.

dx.doi.org/10.1021/sb300018h | *ACS Synth. Biol.* 2012, 1, 299−316

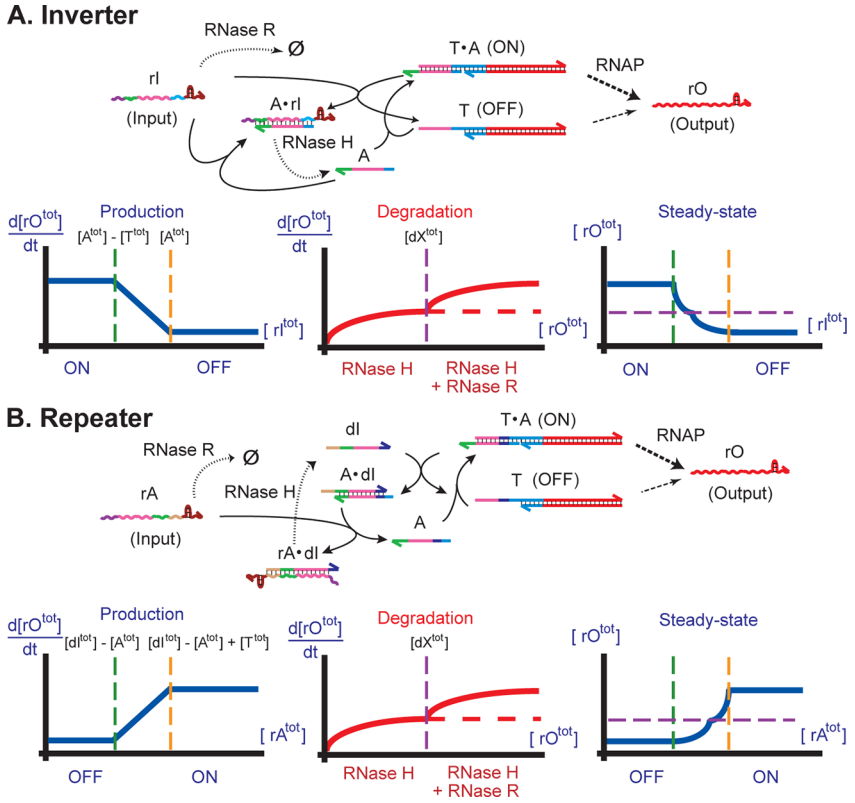


Figure 2. Designed reaction mechanisms. (A, top) An inverter switch. T is the DNA template with an incomplete promoter region, A is the DNA activator, rI is the RNA inhibitor (input), and rO is the RNA transcript (output). Sequence domains are color-coded to represent either complementary or identical sequences; in this and following figures, DNA is represented by straight lines, while RNA is represented by squiggly lines. Hybridization reactions are marked by black arrows. Transcription by RNAP and degradation by RNase H and RNase R are marked by black dashed arrows and dotted arrows, respectively. (A, bottom) The schematic production curve of rO as a function of rI, the degradation curve of rO as a function of rO, and the steady-state curve of rO as a function of rI (constructed from the composition of the production and degradation curves). Green dashed lines mark $[rI^{tot}] = [A^{tot}] - [T^{tot}]$, where all free A species are consumed by rI, yet the switch is still fully ON; orange dashed lines mark $[rI^{tot}] = [A^{tot}]$, where all A species are consumed such that the switch is fully OFF. The purple dashed line marks $[rO^{tot}] = [dX^{tot}]$ where dX is the regulatory target of rO: below this level rO is mostly bound to dX such that the degradation curve is dominated by RNase H, while above this level RNase R degrades free rO. A dashed red line in the degradation plot illustrates the case when only RNase H is present. Note that the dX strand or RNase reactions on rO are not shown. (B, top) A repeater switch. T and A are the same as above, dI is the DNA inhibitor used to set the switching threshold, rA is the RNA activator (input), and rO is the RNA transcript (output). (B, bottom) Green dashed lines mark $[rA^{tot}] = [dI^{tot}] - [A^{tot}]$, where all free dI species are consumed by rA, yet the switch is fully OFF; orange dashed lines mark $[rA^{tot}] = [dI^{tot}] - [A^{tot}] + [T^{tot}]$, where enough A species are freed from dI and available to hybridize with T, such that the switch is fully ON. The purple dashed line marks $[rO^{tot}] = [dX^{tot}]$, where dX is the regulatory target of rO as above. Note that annihilation, interfering, recovering, recapturing reactions, dX strand, and RNase reactions on rO are not shown (see Figure 4).

RESULTS

Overview of Design and Analysis. We take a systematic approach to the design and analysis of synthetic biochemical circuits. Much as in other forward engineering approaches, our first step was the system specification, describing desirable system-level behaviors (e.g., input-output response characteristics) and the required component-level interactions ("System Specification and Reaction Mechanisms"). Then, individual components (here, DNA sequences for switches) were designed based on known biophysics of nucleic acids so that they satisfy the thermodynamic and kinetic constraints ("Sequence Design"). Each reaction mechanism was independently tested ("Reaction Mechanism Validation"). All components were then composed into a complete system (a positive autoregulatory circuit) with desired system behaviors (bistable dynamics) demonstrated experimentally ("System-Level Characterization of Bistability"). Finally, a mechanistic model was built from the characterized component-level interactions, fit to experimental data, and used to predict system-level behavior for new experiments, with confidence intervals estimated for both model parameters and predictions ("Quantitative Modeling and Prediction").

System Specification and Reaction Mechanisms. The repeater shares several components and functional mechanisms with the previously implemented inverter.[37] Therefore, we first describe an inverter (Figure 2A). An inverter is an *in vitro* transcriptional switch whose RNA output production rate is a sigmoidal inhibitory function of its RNA input concentration. An inverter consists of two components, a DNA template ("T") and a DNA activator ("A"). The DNA template T consists of a single-stranded regulatory domain, a partially single-stranded T7 RNAP promoter, and a double-stranded region encoding an RNA output ("rO"). This partially single-stranded promoter is transcribed about 100-fold less efficiently than a complete promoter[42,37] and thus is designated as an OFF state. The single-stranded DNA activator A is complementary to the missing promoter region of T. Upon hybridization of T and A ("activation"), the resulting T·A complex has a complete promoter except for a nick and was found to be transcribed well, approximately half as

dx.doi.org/10.1021/sb300018h | ACS Synth. Biol. 2012, 1, 299–316

efficiently as a fully double-stranded promoter. Therefore, T·A is designated as an ON state. Note that there is a single-stranded domain of A that extends beyond the double-helical domain of the T·A complex, called the "toehold" domain. The toehold is important to ensure that regulation of the inverter by the RNA inhibitor ("rI") is fast. Because the A·rI complex is thermodynamically more stable than the T·A complex (it has more base pairs) and the toehold domain of A is available for initiating a hybridization reaction with rI, rI can strip off A from the T·A complex ("inhibition") by means of a toehold-mediated branch migration and strand displacement reaction.[43,44] Typically, A is in excess of T such that the input rI will first react with free A ("annihilation") and then strip off A from the ON-state switch, and the remaining rI (if any) will be free in solution. Assuming strong and irreversible stoichiometric hybridization reactions occurring at a much faster time scale than the production and degradation of RNA signals, the response of the switch to the input rI is a piecewise linear graph, such that the production rate of rO is a piecewise linear inhibitory function of rI (Figure 2A, left). In practice, the input-output response is sigmoidal rather than piecewise linear due to constant turnover of RNA signals on a similar time scale as hybridization (cf. ref 37).

The degradation rate of RNA signals plays an important role in circuit dynamics by setting the time-constants of signal propagation. At the same time, the shape of the degradation curve in combination with the production curve determines the steady-state RNA levels. In our system, RNA signals are degraded by two different RNases: *Escherichia coli* RNase H and RNase R.[45,46] RNase R degrades single-stranded RNA, which here would be both free rI and free rO. RNase H degrades RNA that is hybridized to DNA, which here would be rI within the A·rI complex and rO that is bound to its downstream regulatory target, "dX", i.e., rO within the dX·rO complex. Now consider the degradation curve of rO that results from these two enzyme activities. At low rO concentrations, most rO molecules exist within a dX·rO complex and therefore get degraded by RNase H; at high rO concentrations, free rO, the amount of rO in excess of the total dX concentration, can be degraded by RNase R. Taking hybridization reactions into account, the degradation curves for RNase H and RNase R each constitute a typical Michaelis−Menten saturation curve but with different origins: $[rO^{tot}] = 0$ for RNase H and $[rO^{tot}] = [dX^{tot}]$ for RNase R. Thus, the overall degradation function has a kink located at the total concentration of dX, where RNase R kicks in (Figure 2A, middle). When the maximum rates and switching thresholds for production and degradation curves are approximately matched, the resulting steady-state RNA output level shows a sigmoidal inhibitory response with respect to RNA inputs, although the transition region contains a kink (Figure 2A, right). In principle, the kink within the transition region may not exist depending on the total concentration of dX.

A repeater utilizes some of the same modular design motifs as an inverter (Figure 2B). The repeater design used indirect activation by an RNA activator, rather than using RNA to directly bind and activate a ssDNA promoter for a couple of reasons (see "Sequence Design" for details). Like an inverter, a repeater also contains a DNA template T and a DNA activator A. However, the repeater has an extra component, a DNA inhibitor ("dI"). Much like the RNA inhibitor rI of an inverter, the DNA inhibitor dI can bind to free A ("DNA annihilation") or strip off A from the ON-state T·A complex to produce the A·dI complex and the OFF-state switch T ("inhibition"). The input of the repeater, an RNA activator, "rA", is a single-stranded RNA that can bind to free dI ("RNA annihilation") or strip off dI from A·dI to form the rA·dI complex ("release"). Then, the released A can bind back to T and turn the switch ON. Unlike the inverter, the initial concentrations of T and A should be about the same, while the initial concentration of dI should be in excess of A to provide the activation threshold for rA. The input rA will first react with free dI and then strip off dI from the A·dI complex, and the remaining rA (if any) will be free in solution.

Overall, the fraction of ON-state switch in a repeater and, consequently, the output production rate, is a sigmoidal activation function of its RNA input, rA (Figure 2B, left). The degradation function of a repeater is identical to that of an inverter (Figure 2B, middle). Therefore, the resulting steady-state RNA output level shows a sigmoidal activation response with respect to RNA input with a (possible) kink within the transition region (Figure 2B, right). As in the case of the inverter, the repeater's regulatory domain and output domain are modular, allowing for independent sequence choices for the RNA input and RNA output. Furthermore, the inverter and repeater switches can be implemented together as modular components within a circuit;[38,39] however, while a given RNA signal species can regulate multiple downstream switches, due to the difference in domain structures for activators and inhibitors, it must either act as an activator for all of its target switches or as an inhibitor for all of its target switches. With this caveat, which could probably be overcome with minor modifications to our design, our modular construction can be applied systematically to implement arbitrarily complex circuits.

Sequence Design. The challenge of designing sequences for a functional repeater lies in accommodating desirable hybridization reactions outlined above and avoiding side reactions and crosstalk. The sequence design for a repeater needs to satisfy both thermodynamic and kinetic constraints: the RNA activator rA should have the desired regulatory influence on the state of a repeater switch through sequence-specific hybridization reactions, ideally on a faster time scale than that of its own production and degradation.

Sequences of the repeater components were chosen to minimize alternative folding[47] and spurious interactions.[48,49] Utilizing the modularity of synthetic switch designs, several domains were adapted from previous work,[37] e.g., the binding domain of the OFF switch template to the activator (27 bases) and the toehold of the activator (9 bases). The 3′ end hairpin structure (16 bases) of the RNA output improves transcription efficiency and also decreases self-coded extension of RNA transcripts by RNAP.[50] The sequences of T, A, dI, rA, and their complexes are shown in Figure 3. (Note that the sequence of the RNA output, rO, is identical to the RNA input, rA, because this design will be used as a self-activating switch in the system-level experiments below; however, for the time being we will focus on reaction mechanisms that are generic to the repeater motif as it could be used in arbitrary circuits.)

We chose to implement a repeater using indirect activation by an RNA activator, rather than using RNA to directly bind and activate a ssDNA promoter for two reasons. First, although weak transcription from an RNA/DNA hybrid promotor may be possible,[51] RNA output from the T7 promoter has to start with "GG",[52] which cannot fit into the missing promoter region, even if we shift the nick position slightly. (Not using the first few nucleotides of the transcript for promoter binding would relieve this sequence constraint but would result in a

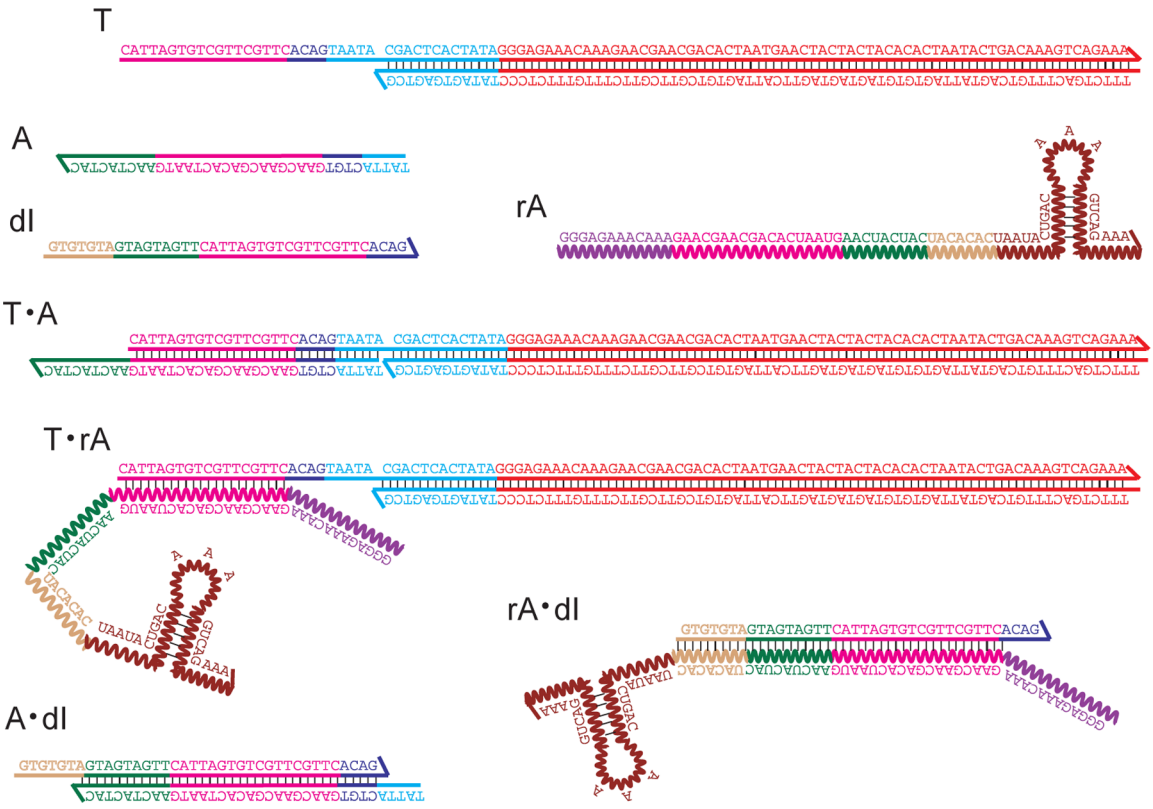


Figure 3. Sequences of DNA and RNA species of a self-activating switch. The sequence domains are color-coded to indicate identical or complementary sequences. The switch template T consists of two single-stranded species T-nt and T-t annealed together (see Methods for details). Although we simply colored the output domain of switch template coding for the RNA output ("rO" in Figure 2) red, this sequence is identical to the RNA input "rA" because we constructed a self-activating switch.

single-stranded tail at the nick in the promoter, potentially disrupting transcription further.) Second, in case we use an RNA activator directly, the abortive transcripts (typically the first 8−13 nt of RNA transcript), which are more abundant than full-length transcripts,[53] could bind spuriously to the promoter region, competing with the intended RNA−DNA interaction at the promoter. The abortive transcripts are also produced in our current switch design; however, we did not use those initial RNA sequences or their complements for the signal propagation, minimizing interference by abortive transcripts.

Because we implemented indirect activation by RNA inputs, the RNA activator rA and the DNA activator A share common sequence domains. Thus, a T·rA complex is expected to form when both T and rA are available ("interfering"). However, it is not desirable if an excess of rA interferes with the hybridization reaction between T and A. Thus, we implemented a staggered design, where the A·dI complex leaves 5 bases of A in single-stranded form (colored light blue in Figure 3) and the rA·dI complex leaves 4 bases of dI in single-stranded form (colored dark blue in Figure 3), resulting in a total 9 base-pair difference between the proper ON state, T·A, and the interfering complex, T·rA. These modifications allow A to strip off T from T·rA ("recovering") and dI to strip off rA from T·rA ("recapturing") quickly and thermodynamically favorably (see Figure 4).

We chose the sequences and lengths of the binding domains for T, A, dI, and rA such that the predicted standard free energy changes for hybridization reactions ensure that each desired reaction is thermodynamically favorable. Specifically, the standard free energies of formation (ΔG°) for complexes are predicted to be in the following order: rA·dI < A·dI < T·A < T·rA, where lower values are more favorable (Figure 4, inset). These energies were estimated using the standard nearest-neighbor model with DNA−DNA parameters[54] for the formation of T·A and A·dI complexes and RNA−DNA parameters[55] for the formation of T·rA and rA·dI complexes. Since the two strands that form the template T do not separate from each other in any of the considered reactions, the free energy of T itself may be (and was) ignored. (See Methods for details.) Note that starting complexes for strand displacement reactions (T·rA, T·A, and A·dI for recovering and recapturing, inhibition, and release, respectively) contain single-stranded overhangs (toeholds) to which the incoming strands (A and dI, dI, and rA, respectively) can bind to initiate branch migration, thus ensuring that the desired reactions are not only thermodynamically favorable but also kinetically fast.[43,44]

Reaction Mechanism Validation. To test whether our sequence design strategies resulted in the desired reaction pathways, a gel-based assay was used that can validate the existence of reaction pathways but gives only qualitative information on reaction kinetics. The hybridization reactions and strand displacement reactions for switch components are validated by running different combinations of T, A, dI, and rA in a non-denaturing gel (Figure 4). For the gel analysis, we used 50 nM of T, 500 nM of A, 700 nM of dI, and 500 nM of rA such that A is in excess of T, dI is in excess of A, and the sum of A and rA is in excess of dI. The two single-stranded species comprising T were annealed prior to mixing with other single-stranded species. The indicated components were simultaneously mixed in a test tube at room temperature and were allowed to sit for at most 5 min before being subjected to a non-denaturing gel.

dx.doi.org/10.1021/sb300018h | *ACS Synth. Biol.* 2012, 1, 299−316

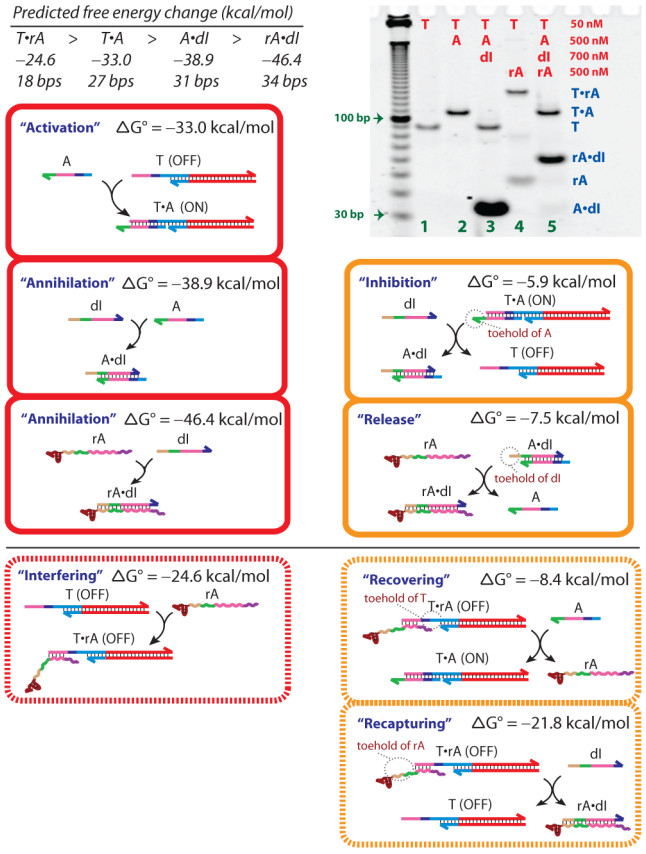


Figure 4. Validation of DNA and RNA hybridization reactions in a repeater switch. Schematic representations of simple hybridization reactions (red boxes) and of strand displacement reactions (orange boxes) are shown. The three reactions with dashed outlines are not included in our mathematical model (see Box 1). Note that all reactions are thermodynamically driven and that toehold-mediated branch migrations provide fast kinetic pathways for the strand displacement reactions. (The inhibition and release reactions, being examples of "toehold exchange",[44] are reversible despite being biased forward; however, the reverse reactions are not modeled.) A non-denaturing gel was used to analyze the results of such hybridization and strand displacement reactions in the absence of enzymes (top right). The red labels indicate the constituents loaded in each lane. The leftmost lane contains a 10 base-pair ladder. The blue labels on the right indicate corresponding single-stranded species and complexes. Single-stranded A and dI bands do not appear because they have run off of the gel.

When T and A were mixed together, A bound to T, resulting in a well-defined band of T·A complex (lanes 1 vs 2: activation). When T, A, and an excess of dI were mixed together, only T and A·dI complex were observed, implying that dI can bind to A and also strip off A from T·A (lanes 2 vs 3: annihilation and inhibition). Of note, the single-stranded A and dI migrated faster than the 30-bp marker and were not visualized in the gel. Also, when rA and T were mixed together, rA bound to T and formed an interfering complex T·rA, which migrated slower than T·A complex (lane 4: interfering). However, when all four components were mixed together with the sum of A and rA concentrations in excess of dI, the T·rA complex was no longer visible, while T·A and rA·dI complexes were observed, implying that A and dI can strip off rA from T·rA and dI can bind to rA (lanes 4 vs 5: recovering, recapturing, and annihilation). Moreover, it is also implied that rA can bind to dI and strip off dI from the A·dI complex and release A, which in turn binds to T, resulting in a T·A complex (lanes 3 vs 5: annihilation, release, and activation). We did not observe a noticeable A·dI band in lane 5 (roughly 200 nM would be expected), indicating that the concentration of rA may have been underestimated. The apparent completion of all reactions within the allotted 5 min prior to loading on the gel is qualitatively consistent with the expected kinetics for DNA hybridization and toehold-mediated strand displacement reactions. (The reported measurements are in the range of 10^5 to 10^6/M/s.[43,56,57])

Three types of enzyme reactions were separately characterized (Figure 5). As desired, RNAP could efficiently transcribe rA from an ON-state template, T·A, while transcription was barely detectable from an OFF-state template, T, even after 3.5 h of incubation (Figure 5, top). In this case, since dI was not provided in the transcription reaction, we do not expect to observe autoregulation. RNase H could degrade rA when both rA and dI were present, but no degradation of rA was observed without dI (Figure 5, middle). Thus, RNase H degrades RNA that are hybridized with DNA, but not RNA free in solution. RNase R could degrade about 4 μM rA within 210 min; however, 1 μM of rA was left over after the same amount of time when 4 μM of rA was mixed with 1 μM of dI (Figure 5, bottom). Thus, RNase R degrades single-stranded RNA, but not RNA within RNA−DNA hybrid complexes.

Taken together, our sequence design resulted in the desired fast hybridization and strand displacement reactions among switch components, with enzyme reactions providing production and degradation of RNA regulatory signals. Note, however, that it is not clear whether there is a well-defined time-scale separation for enzyme reactions and hybridization or strand displacement reactions; this time-scale separation was assumed for the piecewise-linear response of the switch to the input signal (cf. Figure 2B). As with the response characteristics of inverters,[37] we expect that the sharp switch response of the repeater will likely be somewhat smeared out in the case of fast enzyme-mediated turnover of signals. Therefore, when setting enzyme and DNA concentrations to obtain desired circuit operation, one must strike a balance between the speed of the switch's response (high enzyme concentrations) and the sharpness of its transition (low enzyme concentrations).

System-Level Characterization of Bistability. We have thus far verified each elementary reaction required for a functional repeater. To demonstrate the functionality of a repeater in the context of a circuit, we chose to implement, with a single self-regulating repeater switch, a bistable latch whose function is similar to that of the two-node network developed in ref 37. The construction of a self-activating switch is straightforward as it only requires the RNA output rO to be identical to the RNA input rA, as shown in Figure 3, and therefore, we call the RNA output "rA" henceforth. The target of the RNA output, dX, is now the DNA inhibitor, dI. Beyond demonstrating an interesting circuit, the construction serves as an initial test of the sufficiency of the non-quantitative mechanism characterization for predicting system-level behavior.

We explored the self-activating switch behavior under two different enzyme settings, with both RNase H and RNase R or with RNase H only. As a single-strand specific RNA exonuclease, RNase R is promising for reducing incomplete degradation product ("waste") and potentially helping establish predictable system behaviors. On the other hand, RNase R presented experimental challenges including not being commercially available at the time of experiment. Further, excluding RNase

dx.doi.org/10.1021/sb300018h | ACS Synth. Biol. 2012, 1, 299−316

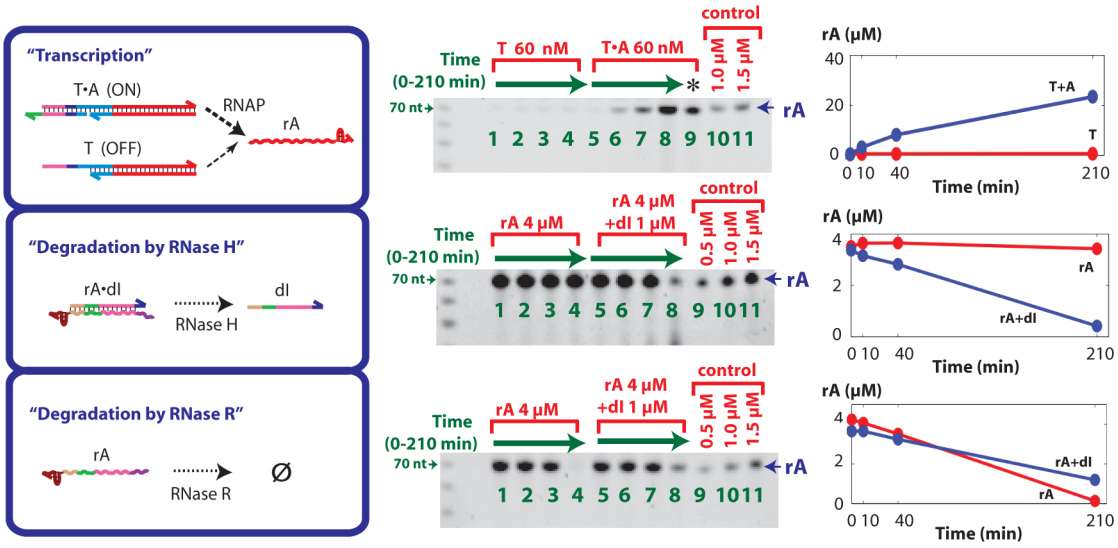


Figure 5. Validation of enzyme reactions in a repeater switch: (left) reaction diagrams, (middle) denaturing gel analysis, (right) measured RNA level from gels. For the denaturing gels, the leftmost lanes contain 10 base ladders and the templates or substrates for enzymes are indicated by red labels on the top of the gels. Samples in lanes 1−4 (and also for lanes 5−8) were taken from the same reaction tube at different time points (0, 10, 40, and 210 min). The sample in lane 9 of the gel analyzing the RNAP reaction (*) contained a half volume of that loaded in lane 8 to avoid SYBR gold signal saturation. Control lanes have purified rA at the indicated concentrations. Nominal enzyme concentrations are as follows: [RNAP] = 33.4 nM, [RNaseH] = 0.168 nM, and [RNaseR] = 0.336 nM.

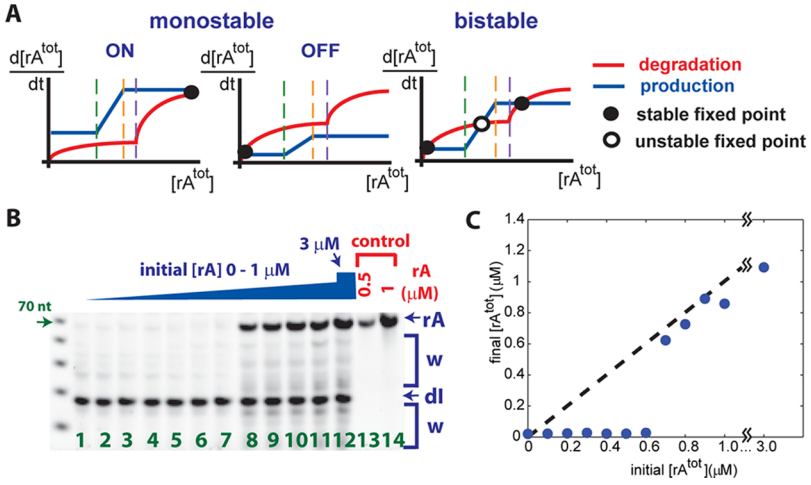


Figure 6. Bistability of a self-activating switch using both RNase H and RNase R. (A) Schematic production and degradation curves and dynamics of a self-activating switch. Depending on the locations and stabilities of fixed points, two types of monostable behaviors or a bistable switch behavior are expected. Green dashed lines mark $[rA^{tot}] = [dI^{tot}] - [A^{tot}]$, where the switch is fully OFF; orange dashed lines mark $[rA^{tot}] = [dI^{tot}] - [A^{tot}] + [T^{tot}]$, where the switch is fully ON; purple dashed lines mark $[rA^{tot}] = [dI^{tot}]$, where RNase R kicks in (cf. Figure 2B). (B) A denaturing gel analysis to test bistability of a self-activating switch. Lanes 1−12 show the results from 12 separate reactions. The reaction conditions are as follows: $[T^{tot}]$ = 50 nM, $[A^{tot}]$ = 90 nM, $[dI^{tot}]$ = 1000 nM, [RNAP] = 66 nM, [RNaseH] = 0.7 nM, and [RNaseR] = 0.23 nM, with variable initial rA concentrations ($[rA^{tot}]$ = 0, 0.1, 0.2, ..., 1, and 3 μM). After 120 min, all reactions were stopped and subjected to denaturing gel analysis. Control lanes have purified rA at the indicated concentrations. Brackets marked "w" indicate incomplete transcription products and degradation products. (C) The measured final concentrations of rA from the denaturing gel in panel B. Black dashed line marks the points where the initial and final rA concentrations coincide.

R eliminates the kink within the degradation function and possibly allows for a simpler quantitative analysis of switch dynamics.

First, let us consider the case with both RNase H and RNase R in the system. The production and degradation curves shown in Figure 2B can now be interpreted as the production and degradation curves of the RNA activator rA as functions of its own concentrations. The intersections of production and degradation curves thus indicate the steady states where total rA concentrations remain constant, and where and how they cross will dictate whether the self-activating switch shows monostable or bistable behavior. If the production and degradation curves cross at a single stable fixed point, a monostable ON state or a monostable OFF state will be achieved (Figure 6A, left and center). If an unstable fixed point exists, where a slight increase in rA drives further rA production and a slight decrease in rA drives further rA reduction (Figure 6A, right), the system will show bistable behavior with the threshold set at this unstable fixed point. In that case, the switch states approach either of the two stable steady states, completely ON or OFF, depending on the initial concentration of rA.

Next, we consider the case with RNase H but without RNase R. Since the degradation curve by RNase H alone is insensitive

dx.doi.org/10.1021/sb300018h | ACS Synth. Biol. 2012, 1, 299−316

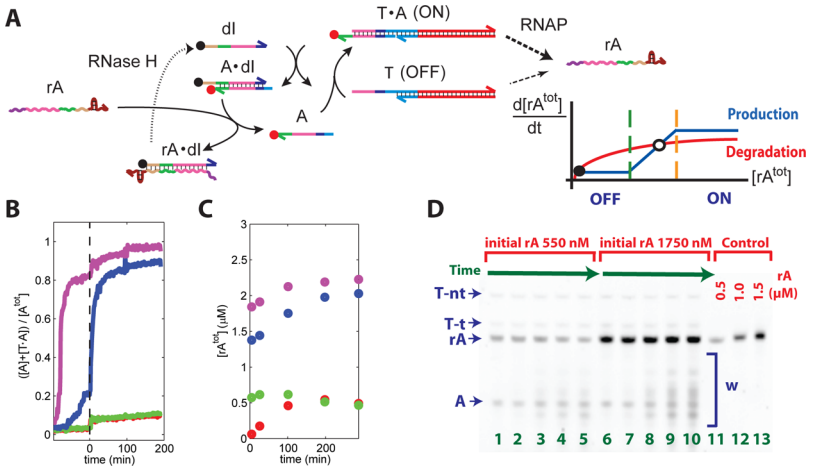


Figure 7. Bistability of a self-activating switch with only RNase H. (A) Schematic reaction diagram (cf. Figure 2B). Here, the activator A is labeled with Cy5 fluorophore (red circle) and the inhibitor dI is labeled with IowaBlack-RQ quencher (black circle) for real-time monitoring of switch states by fluorescence. A possible arrangement of production and degradation curves is shown as an inset (cf. Figure 6A). Green dashed line marks $[rA^{tot}] = [dI^{tot}] - [A^{tot}]$; orange dashed line marks $[rA^{tot}] = [dI^{tot}] - [A^{tot}] + [T^{tot}]$. (B) Time-courses measured by fluorescence. The reaction conditions were 48 nM of $[T^{tot}]$, 145 nM of $[A^{tot}]$, 1500 nM of $[dI^{tot}]$, 16.7 nM of [RNAP], and 1.68 nM of [RNaseH] with the initial rA concentrations at 0, 550, 1350, and 1750 nM (colored red, green, blue, and magenta, respectively). The enzymes were added at time 0 (black dashed line). Normalized fluorescence signals were used to determine the fraction of A that is not bound to dI. (C) Time courses of $[rA^{tot}]$ measured by denaturing gel. The colors correspond to those of fluorescence traces in panel B. (D) One of the denaturing gels used for measurements shown in panel C: lanes 1−5 correspond to samples marked green in panel C; lanes 6−10 correspond to magenta in panel C. The rA concentrations were measured with respect to the purified rA bands in the control lanes.

to the change of rA concentrations beyond the total concentration of its binding target dI, even when an unstable steady state exists, the production and the degradation curve may not cross at the high steady state (Figure 7A, inset). In this scenario, the behavior of the switch state may still be bistable, but in the high state the rA concentration will grow without bound, which makes the system more latch-like.

For both enzyme settings, whether the switch is bistable or monostable depends not so much on the physical nature of molecular components but rather on the features set by continuously tunable concentrations, such as the output amplitude and the activation threshold.

We first experimentally demonstrated bistability of the self-activating switch operating in the presence of both RNase H and RNase R (Figure 6). When the self-activating switch was subject to transcription and degradation reactions with a wide range of initial rA concentrations, the final rA concentrations reached two distinct states after 2 h: almost 0 nM or higher than 600 nM. Thus, the self-activating switch showed bistable behavior as expected.

However, the experimental results deviated from an idealized circuit behavior in two ways: a low switching threshold and non-constant ON-state outputs. First, the switching threshold lay between 600 and 700 nM of rA in the experiment while the expected threshold from idealized production and degradation curves lay between 910 and 960 nM of rA for the case shown in Figure 6B (cf. green and orange dashed lines in Figure 6A). This discrepancy may be explained by the "burst phase" in enzyme kinetics,[58] which could initially produce extra RNA transcripts and effectively decrease the apparent switching threshold for initial rA concentrations. Second, the ON-state output levels were not constant, in contrast to almost identical OFF-state output levels. For the largest input (3 μM rA), the initial excess RNA was indeed reduced to the presumed steady-state level of roughly 1.1 μM, but for ON-state inputs of 1 μM or less the final output concentrations were roughly the same as the input concentrations. This could be explained by slow production and degradation kinetics near the steady state. On the other hand, the measured rA output levels were as low as 600 nM when more than 960 nM of rA would be required to maintain a fully ON state in the idealized model. RNase R was present to clean up the incomplete single-stranded degradation products by RNase H. Nevertheless, incomplete degradation products accumulated over time (Figure 6B, brackets marked "w"). Thus, it is possible that some portions of these incomplete degradation products still function much like an intact rA, e.g., rA without the hairpin region at the 3′ end, suggesting that under these circumstances the *effective* amount of activator may have in fact been closer to 960 nM.

To test bistability in the absence of RNase R, we monitored the kinetics of the switch response through real-time fluorescence measurement (Figure 7). In these experiments, the 3′ end of A is labeled with a Cy5 dye and the 5′ end of dI is labeled with an IowaBlack-RQ quencher such that the fluorescence is low when the A·dI complex is formed due to fluorescence quenching,[59] while the fluorescence is high when A is free or within the T·A complex (Figure 7A). Of note, fluorophore-quencher interactions can stabilize the resulting complex, A·dI.[59] (In preliminary work, we observed mild to severe differences between fluorophore-labeled strands and their unlabeled twins, depending on the fluorophore used.) We initiated the reaction with four different rA concentrations and measured how the system approached different steady states (Figure 7B). The fluorescence traces reached either a maximum signal, which implied a completely ON state, or a minimum signal, which implied a completely OFF state. The fluorescence monitoring indicated that the self-activating switch quickly approached steady state and that both fully ON and OFF cases were stable. At the same time, the dynamics of the RNA signal was determined by taking samples at different time points and measuring the total rA concentrations by gel (Figure 7C,D). The high fluorescence traces (magenta and blue) corresponded to greater

dx.doi.org/10.1021/sb300018h | *ACS Synth. Biol.* 2012, 1, 299−316

than 2 μM final rA levels, but the low fluorescence traces (green and red) corresponded to about 0.5 μM final rA levels. Thus, we achieved bistable behavior in both switch and RNA signals, with the caveat that the high rA output level is (in principle) not bounded.

Quantitative Modeling and Prediction. By virtue of its programmability with quantifiable elementary reactions operating in a simple environment, *in vitro* transcriptional networks offer a promising testbed for evaluating modeling and analysis methods for synthetic and systems biology. Despite its importance, there are a limited number of studies that address the non-identifiability of high-dimensional multiparameter nonlinear models that typically arise in complex models of living organisms[60−62] and fewer still that examine this issue in the context of synthetic biology.[63,64] Here, using the self-activating switch as an example, we provide evidence that Bayesian inference can be used effectively to find parameter set ensembles for detailed biochemical models and make reliable predictions with confidence intervals, especially when the mathematical model was constructed on the basis of detailed mechanistic understanding of elementary reactions.

In the context of *in vitro* self-activating switch, the following questions can be investigated by mathematical modeling and analysis. First, are the elementary reactions described in Figures 4 and 5 sufficient to explain the system behavior? Second, how fine-tuned do the rate constants have to be to achieve bistability? Third, would the model biochemical reaction network be able to "learn" from a small subset of training data to predict other experimental results? To address these questions, we constructed a simple mathematical model for the self-activating switch that uses four ordinary differential equations. The dynamics of each DNA and RNA species were derived from elementary hybridization and strand displacement reactions described in Figure 4, assuming Michaelis−Menten enzyme kinetics for enzyme reactions described in Figure 5 (see Box 1).[66,67] Note that the concentration of "interfering complex" T·rA was assumed to be negligible, and thus, "interfering", "recovering", and "recapturing" reactions were not included. Initial simulation results showed that a bistable switch response can be achieved with plausible rate parameters similar to a previous work.[37]

To test whether our mathematical model, with parameters tuned to match a relatively small number of experiments, is predictive of system behavior for a wider range of parameter choices, we need separate training and test data sets. A sufficiently rich training set is needed to constrain parameters effectively. The training set was obtained by measuring final rA concentrations for self-activating switches initialized with a wide range of rA concentrations and three different dI concentrations as switching thresholds (Figure 8A, red, blue, and green circles). As expected from the idealized production and degradation curves in Figure 7A, adjusting dI concentrations shifted the threshold of switching accordingly. Below the apparent switching threshold, final rA concentrations were about 0.5 μM, while well above the threshold, final rA concentrations exceeded 2 μM.

Two separate test data sets were prepared to assess how well our mathematical model, after constraining the model parameters using only the training data, can generalize to novel experimental conditions. The first test set was the time-course experiment presented in Figure 7, where both the switch state and rA concentrations were monitored (Figure 8B). The second test data set was prepared to analyze the sensitivity of bistability with respect to each experimental variable by systematically varying the template ($[T^{tot}]$), activator ($[A^{tot}]$), inhibitor ($[dI^{tot}]$), and enzyme concentrations ([RNAP] and [RNaseH]); for each chosen value of an experimental variable, the reaction was initiated with either a high or low amount of rA (730 nM vs 0 nM), and the final rA concentration was recorded (Figure 8C).

To find suitable values for the set of 11 parameters in our model (Box 1), we used a Monte Carlo Bayesian inference approach that results in an ensemble of parameter sets that are compatible with the given data within noise bounds.[60] Because relationships among parameters and experimental observables are inherently non-unique,[68] in order to make useful predictions, the model's behavior with all parameter combinations consistent with the available data needs to be analyzed,[60] ideally generating model predictions with quantifiable confidence intervals.

Here, we briefly describe the Monte Carlo Bayesian inference approach to find an ensemble of parameter sets, loosely following ref 60. (See Supporting Information for details.) We define the cost function as

$$E = \sum_{n=1}^{N} ([\mathrm{rA}]_{\mathrm{exp}}^{\mathrm{final},n} - [\mathrm{rA}]_{\mathrm{sim}}^{\mathrm{final},n})^2$$

where the concentration of rA is measured at 210 min in micromolar scale both for the experiment and simulation (n = experiment number, N = 50). We chose σ = 0.25 μM as a reasonable estimate of the standard deviation of repeated experimental measurements. Let $\vec{\theta}$ be a 11 × 1 vector of logs of model parameters, so that a uniform *a priori* distribution on $\vec{\theta}$ coincides with our expectation that biochemical rate constants and equilibrium constants frequently vary over many orders of magnitude. Then if we assume Gaussian random measurement errors and perform Bayesian inference, the conditional probability of our model (M) with parameter set $\vec{\theta}$ given observed data (D) will be

$$P(\vec{\theta}) = P(M(\vec{\theta})|D) \propto \exp\left(-\frac{E}{T}\right)$$

Our goal then is to generate a thermal ensemble of parameter sets consistent with a Boltzmann distribution with "energy" E and "temperature" $T = 2\sigma^2$ in units where Boltzmann's constant is unity. Twelve sampling trajectories with 440,000 iterations each were generated from random initial conditions followed by a random walk obeying Metropolis criteria, i.e., the acceptance probability for a move in a random direction was $P = 1$ if $\Delta E < 0$ and $P = e^{-\Delta E/T}$ if $\Delta E > 0$. By analyzing the error history (Supplementary Figure S1) and autocorrelation functions (Supplementary Figure S2), the initial 50,000 iteration steps were discarded as not being in equilibrium; thereafter parameter sets were sampled every 50,000 iterations to ensure statistical independence.

The resulting ensemble of 96 independent parameter sets fit the experimental training data and are compatible with experimental noise (Figure 8A). Furthermore, they can be used to make ensemble predictions for novel experimental conditions, for example, "predicting" the time-course experiments of Figure 7B,C (which were not part of the training data). As shown in Figure 8B, the experimentally measured fluorescence signal from A (monitoring switch states) and sampled rA concentrations were close to the ensemble average prediction of the model. Not surprisingly, individual predictions

Box 1. A simple kinetic model for the self-activating switch

To mathematically explore the behavior of the self-activating switch, we developed a model based on the validated hybridization and strand displacement reactions and enzyme reactions. Note that for the self-activating switch, the input RNA signal is identical to the output RNA signal, thus both are labeled rA. We constructed a mathematical model without RNase R, because RNase R was not used for the experimental results presented in Figures 7 and 8. To simplify the model, we also assumed that, although an 'interfering' reaction exists that results in the T·rA complex, the 'recovering' and 'recapturing' reactions are fast enough such that the concentration of the T·rA complex is negligible (see Figure 4). The reactions incorporated in the model are summarized below.

DNA and RNA hybridization reactions

$$\begin{aligned}
\mathrm{T} + \mathrm{A} &\xrightarrow{k_{\mathrm{TA}}} \mathrm{T}\cdot\mathrm{A} && \text{(Activation)}\\
\mathrm{A} + \mathrm{dI} &\xrightarrow{k_{\mathrm{AI}}} \mathrm{A}\cdot\mathrm{dI} && \text{(Annihilation)}\\
\mathrm{rA} + \mathrm{dI} &\xrightarrow{k_{\mathrm{rAI}}} \mathrm{rA}\cdot\mathrm{dI} && \text{(Annihilation)}\\
\mathrm{dI} + \mathrm{T}\cdot\mathrm{A} &\xrightarrow{k_{\mathrm{TAI}}} \mathrm{T} + \mathrm{A}\cdot\mathrm{dI} && \text{(Inhibition)}\\
\mathrm{rA} + \mathrm{A}\cdot\mathrm{dI} &\xrightarrow{k_{\mathrm{AIrA}}} \mathrm{rA}\cdot\mathrm{dI} + \mathrm{A} && \text{(Release)}
\end{aligned}$$

Michaelis–Menten enzyme reactions

$$\begin{aligned}
\mathrm{T}\cdot\mathrm{A} &\xrightarrow{Prod_{\mathrm{ON}}} \mathrm{T}\cdot\mathrm{A} + \mathrm{rA} && \text{(transcription, fast)}\\
\mathrm{T} &\xrightarrow{Prod_{\mathrm{OFF}}} \mathrm{T} + \mathrm{rA} && \text{(transcription, slow)}\\
\mathrm{rA}\cdot\mathrm{dI} &\xrightarrow{Deg_{\mathrm{H}}} \mathrm{dI} && \text{(degradation)}
\end{aligned}$$

We do not consider side-reactions or incomplete transcription and degradation products. We further simplified enzymatic equations from the full Michaelis–Menten equations to an approximation using the pseudo–steady-state approximation for the enzyme-substrate complex, which is reasonably accurate when enzyme concentrations are low compared to substrate concentrations. We express the available enzyme concentrations using the pseudo–steady-state assumption as shown on the right side with the Michaelis constants calculated as usual ($K_{\mathrm{M}} = \frac{k_- + k_{\mathrm{cat}}}{k_+}$).

$$\begin{aligned}
\mathrm{RNAP} + \mathrm{T}\cdot\mathrm{A} &\underset{k_{-,\mathrm{ON}}}{\overset{k_+}{\rightleftharpoons}} \mathrm{RNAP}\cdot\mathrm{T}\cdot\mathrm{A} \xrightarrow{k_{\mathrm{cat,ON}}} \mathrm{RNAP} + \mathrm{T}\cdot\mathrm{A} + \mathrm{rA}\\
\mathrm{RNAP} + \mathrm{T} &\underset{k_{-,\mathrm{OFF}}}{\overset{k_+}{\rightleftharpoons}} \mathrm{RNAP}\cdot\mathrm{T} \xrightarrow{k_{\mathrm{cat,OFF}}} \mathrm{RNAP} + \mathrm{T} + \mathrm{rA}\\
\mathrm{RNaseH} + \mathrm{rA}\cdot\mathrm{dI} &\underset{k_{-,\mathrm{H}}}{\overset{k_{+,\mathrm{H}}}{\rightleftharpoons}} \mathrm{RNaseH}\cdot\mathrm{rA}\cdot\mathrm{dI} \xrightarrow{k_{\mathrm{cat,H}}} \mathrm{RNaseH} + \mathrm{dI}
\end{aligned}$$

$$[\mathrm{RNAP}] = \frac{[\mathrm{RNAP^{tot}}]}{1 + \frac{[\mathrm{T}\cdot\mathrm{A}]}{K_{\mathrm{M,ON}}} + \frac{[\mathrm{T}]}{K_{\mathrm{M,OFF}}}}$$

$$[\mathrm{RNaseH}] = \frac{[\mathrm{RNaseH^{tot}}]}{1 + \frac{[\mathrm{rA}\cdot\mathrm{dI}]}{K_{\mathrm{M,H}}}}$$

The effective rate constants for enzymes are as follows:

$$Prod_{\mathrm{ON}} = \frac{k_{\mathrm{cat,ON}}}{K_{\mathrm{M,ON}}}[\mathrm{RNAP}], \quad Prod_{\mathrm{OFF}} = \frac{k_{\mathrm{cat,OFF}}}{K_{\mathrm{M,OFF}}}[\mathrm{RNAP}], \quad Deg_{\mathrm{H}} = \frac{k_{\mathrm{cat,H}}}{K_{\mathrm{M,H}}}[\mathrm{RNaseH}].$$

Note that this approximation procedure was also used in ref 37, whereas a more accurate but slower numerical method was used in ref 38. Taken together, a set of four ordinary differential equations describes the dynamics of the self-activating switch as follows.

$$\begin{aligned}
\frac{d[\mathrm{T}]}{dt} &= -k_{\mathrm{TA}}[\mathrm{T}][\mathrm{A}] + k_{\mathrm{TAI}}[\mathrm{T}\cdot\mathrm{A}][\mathrm{dI}]\\
\frac{d[\mathrm{A}]}{dt} &= -k_{\mathrm{AI}}[\mathrm{A}][\mathrm{dI}] - k_{\mathrm{TA}}[\mathrm{T}][\mathrm{A}] + k_{\mathrm{AIrA}}[\mathrm{A}\cdot\mathrm{dI}][\mathrm{rA}]\\
\frac{d[\mathrm{dI}]}{dt} &= -k_{\mathrm{AI}}[\mathrm{A}][\mathrm{dI}] - k_{\mathrm{rAI}}[\mathrm{rA}][\mathrm{dI}] - k_{\mathrm{TAI}}[\mathrm{T}\cdot\mathrm{A}][\mathrm{dI}] + Deg_{\mathrm{H}}[\mathrm{rA}\cdot\mathrm{dI}]\\
\frac{d[\mathrm{rA}]}{dt} &= -k_{\mathrm{rAI}}[\mathrm{rA}][\mathrm{dI}] - k_{\mathrm{AIrA}}[\mathrm{A}\cdot\mathrm{dI}][\mathrm{rA}] + Prod_{\mathrm{ON}}[\mathrm{T}\cdot\mathrm{A}] + Prod_{\mathrm{OFF}}[\mathrm{T}]
\end{aligned}$$

The remaining variables, [T·A], [A·dI], and [rA·dI], are calculated from conserved quantities since [$\mathrm{T^{tot}}$], [$\mathrm{A^{tot}}$], and [$\mathrm{dI^{tot}}$] remain constant throughout the reaction — e.g., [rA·dI]=[$\mathrm{dI^{tot}}$]−[$\mathrm{A^{tot}}$]+[$\mathrm{T^{tot}}$]−[T]+[A]−[dI]. The superscript *tot* indicates that all complexes containing that species are considered, i.e., [$\mathrm{T^{tot}}$]=[T]+[T·A], [$\mathrm{A^{tot}}$]=[A]+[T·A]+[A·dI], and [$\mathrm{dI^{tot}}$]=[dI]+[A·dI]+[rA·dI]. Further, note that the dynamic equation for the total concentration of rA molecules can be derived from the differential equations above, reducing to just the enzyme reactions:

$$\frac{d[\mathrm{rA^{tot}}]}{dt} = Prod_{\mathrm{ON}}[\mathrm{T}\cdot\mathrm{A}] + Prod_{\mathrm{OFF}}[\mathrm{T}] - Deg_{\mathrm{H}}[\mathrm{rA}\cdot\mathrm{dI}].$$

Parameters	Selected fit	min	max	Other studies
$K_{\mathrm{M,ON}}$ (nM)	663	1	1000	15–37
$k_{\mathrm{cat,ON}}$ (/s)	0.195	0.001	1	0.73–1.12
$K_{\mathrm{M,OFF}}$ (μM)	1.36	0.1	100	0.1–1.1
$k_{\mathrm{cat,OFF}}$ (/s)	0.0727	0.001	1	0.11–0.18
$K_{\mathrm{M,H}}$ (μM)	14.9	0.1	100	0.016–0.13
$k_{\mathrm{cat,H}}$ (/s)	0.913	0.001	1	0.02–0.6
k_{TA} (/M/s)	4.26×10^4	10^3	10^7	10^5–10^6
k_{AI} (/M/s)	2.61×10^3	10^3	10^7	10^5–10^6
k_{TAI} (/M/s)	5.66×10^3	10^3	10^7	10^5–10^6
k_{rAI} (/M/s)	5.61×10^3	10^3	10^7	10^5–10^6
k_{AIrA} (/M/s)	4.66×10^4	10^3	10^7	10^5–10^6

The mathematical model has 11 parameters ($\vec{\theta}$) with pre-specified minimum and maximum values spanning between 3 and 4 orders of magnitude. For comparison, enzyme constants and hybridization rates from other biochemical studies are listed.[42,43,56,65] The 'selected fit' is a hand-picked parameter set from the Monte Carlo sampling trajectories, one which had a small error on the training data ($E < 1.7\ \mu\mathrm{M}^2$) as well as reasonable agreement with the ensemble average prediction for the test data. Note that this selected fit is not a part of the ensemble of parameter sets ($S = 96$) shown in Figure 8. The parameters of the selected fit are not intended as "realistic" mechanistic parameters, but rather are provided as an example instance of Monte Carlo sampling. (Due to the limits of the training data set, i.e., most DNA and enzyme concentrations being fixed, it is unlikely to achieve realistic estimates for individual parameters. See ref 38 for more realistic parameters.)

showed large variability in the switch response time-courses, possibly because the training data set contained information on rA concentrations at a single time point only. This illustrates that the parameter sensitivity depends on the context of experiment and types of experimental data available, i.e., the final rA measurements were simply not able to constrain

dx.doi.org/10.1021/sb300018h | ACS Synth. Biol. 2012, 1, 299–316

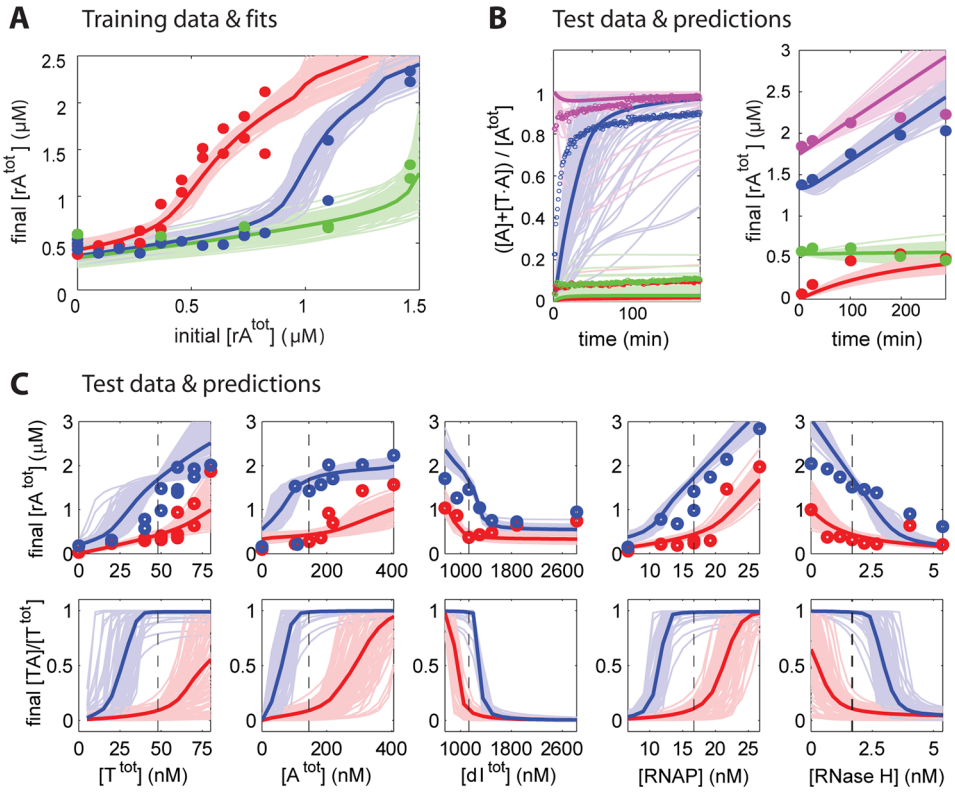


Figure 8. Quantitative measurements and model prediction of the self-activating switch dynamics. A total of 96 sets of ensemble model parameters were selected from fitting a simple kinetic model to the experimental results of panel A. Experimental results in panels B and C are used as test cases of model prediction. Each panel shows experimental measurements (circles), individual predictions from the model (thin lines), and ensemble average model predictions (thick lines). (A) The training data set. The circles are gel measurements of final $[rA^{tot}]$ at 210 min. All reactions contained 48 nM of $[T^{tot}]$, 145 nM of $[A^{tot}]$, 16.7 nM of [RNAP], and 1.68 nM of [RNaseH]. The concentration of dI was varied as follows: $[dI^{tot}]$ = 1.13 μM (red), 1.45 μM (blue), or 1.87 μM (green). (B) Predicted time courses of switch states and rA concentrations. The experimental results are from Figure 7B,C; the experimental conditions are different from those in panel A in that $[dI^{tot}]$ = 1.5 μM. (C) Predicted sensitivity of bistability to experimental variables. The default set of experimental conditions were identical to those colored red in panel A and marked as black dashed lines. (Top) In each panel, a series of experiments varying a single experimental variable was performed, and the final concentration of rA after 210 min was recorded using an initial rA concentration of 0 nM (red) or 730 nM (blue), respectively. (Bottom) Bistable switch responses are predicted by simulation for variations of experimental variables around their default values.

hybridization rate parameters effectively. Note that, for the two ON-state switch conditions, the experimentally measured rA concentrations did not increase indefinitely as the model predicted, but seemed to reach a high steady state; we shall address this discrepancy between the model and experiment later.

To further test the sensitivity of system behavior to experimental variables (DNA and enzyme concentrations) and to validate the predictive power of our mathematical model, we systematically varied these experimental variables and compared with simulation results using the ensemble parameter set (Figure 8C). Bistability versus monostability was qualitatively assessed for each experimental variation by running two experiments, one with an initial rA concentration of 730 nM, and one with 0 nM initially. If, after 210 min, both experiments' final rA concentrations are high, then these experimental conditions are deemed "monostable ON"; if both final rA concentrations are low, we say it is "monostable OFF"; while if the former is high yet the latter is low, we say it is qualitatively "bistable". (Note that for monostable ON runs, the final rA difference was typically near 730 nM because the degradation machinery was saturated.) All three types of outcomes were observed for each experimental variable explored. The experimental results were in reasonably good agreement with the model ensemble average prediction for final rA concentrations despite high variability in individual predictions. Although the switch states were not measured experimentally, the ensemble average model prediction suggested that the self-activating switch would have reached distinct final switch states, indicating bistability, for roughly 2-fold changes of most experimental variables around their default values (Figure 8C).

Given its simplicity, our model predicted the dynamics of the switch surprisingly well. We only included the main hybridization and enzyme reactions that constituted our design intentions, excluding the off-pathway "interfering", "recovering", and "recapturing" reactions as well as any other undesired crosstalk between DNA or RNA species. Notably, the training data (Figure 8A) only gives information about how the final $[rA^{tot}]$ depends on the initial $[rA^{tot}]$ and $[dI^{tot}]$, providing very weak kinetic constraints on the model. Thus, most parameters were under-constrained by design. Nonetheless, qualitative (and sometimes quantitative) predictions were made accurately for a variety of novel experimental conditions (the test data, Figure 8B,C), including the time evolution of the RNA signal and switch state. This suggests that the intended reaction pathways, used to construct the model, are dominant ones in the system dynamics.

The result here highlights an advantage of using mechanistic models over phenomenological models (e.g., Hill functions) or "black box models" (e.g., neural networks): mechanistic models

dx.doi.org/10.1021/sb300018h | ACS Synth. Biol. 2012, 1, 299–316

allow us to incorporate considerable prior knowledge about the system and thus can be used to predict the system behavior under novel conditions. A case in point is the reasonable prediction of sensitivity on the variation of $[A^{tot}]$ (Figure 8C). Phenomenological or black box models without prior mechanistic knowledge on the dependence of the switching threshold on $[A^{tot}]$ would likely perform poorly for this task, given that no such information is available from the training data.

Mechanistic biochemical models are sometimes avoided because, relative to phenomenological models for example, they often have many parameters that are poorly constrained by the experimental data, putting the inferred parameter values and model predictions in question.[61] Indeed, when we chose the very best 26 parameter sets from our simulation runs (rather than the 96 independent samples from the Boltzmann distribution with the temperature matched to the experimental noise), we found that their errors on the training set were nearly indistinguishable, but their parameter values and predictions were nearly as variable as for the Boltzmann ensemble. This gives us insights into the difference between the Bayesian ensemble approach and the more common practice of using a single least-squared-error (LSE) parameter vector: whether due to overfitting, multiple local minima, or imperfect numerical optimization, using a single LSE parameter vector is analogous to choosing just one of the fitting or prediction curves for each experiment in Figure 8, and a random choice is likely to be quite far from the ensemble average. To put it another way, the Bayesian ensemble approach makes mechanistic biochemical models practical despite their poorly constrained parameters.[61] This allows one to take advantage of the structural information implicit in the mechanistic model, even when the individual characterization of each reaction mechanism is coarse (as it was in this study) or even nonexistent.

Nonetheless, some predictions were qualitatively wrong for all parameter sets in the *a posteriori* ensemble. These prediction errors may point to omissions in the model. As an example, for the time evolution of the RNA signal (Figure 8B, right), the model predicted that the upper two reactions would reach a state in which $[rA^{tot}]$ keeps increasing, while in the experiment, these two reactions seemed to converge to a steady state. Whereas exhaustion of the buffer introduced nonidealities in some previous transcription/translation systems,[17] prior studies of genelet systems showed stable behavior for as long as 20 h,[38] making that explanation unlikely here. A possible explanation is that incomplete degradation by RNase H, as observed in previous works,[37,38] yields short fragments of rA that accumulate during the reaction: at high concentrations of these short products, product rebinding to the transcription initiation complex of RNAP may decrease transcription rates.[69] As another example, the model predicts a stronger sensitivity to RNase H concentration than was observed (Figure 8C, rightmost panel). Incomplete degradation fragments may also provide an explanation for this discrepancy. These shorter oligonucleotides could also bind to dI and thus compete for the degradation by RNase H, decreasing its effectiveness. Mathematical modeling in ref 38 indicated that the accumulation of waste products could induce measurable differences in system dynamics within a few hours. Thus, in contrast to phenomenological and black-box models, mechanistic models can be readily interpreted and can directly suggest experiments that would refine or augment the model.

Now, we turn our attention to how effectively the parameters were constrained from the training data set. The log-scale parameter values for all 12 sampling trajectories of 390,000 iterations (after the 50,000 transient iterations were discarded because they did not reach equilibrium during Boltzmann sampling) are plotted as histograms (Figure 9A); the 12 histograms for each sampling trajectory showed similar distributions irrespective of the starting point. Except for a few enzyme parameters, individual parameters for the mechanistic model are poorly constrained. This may not be surprising given that the type of measurement in the training data (final $[rA^{tot}]$ only) is quite limited, but it is somewhat surprising in that, despite the order-of-magnitude variations within the parameter ensemble, the ensemble's fits remain consistent with the experimental training data (within measurement noise) and the predictions also, for the most part, are quite constrained. To identify important parameter combinations that can explain most of the variability in the inferred model parameters, as well as those that must be precisely respected in order to obtain the accurate fits to the data, we performed mode spectrum analysis for ensemble parameter sets as outlined in refs 60 and 61. In this way, ensemble Bayesian analysis goes beyond classical "one-factor-at-a-time" sensitivity analysis.[70]

Specifically, one can construct an empirical covariance matrix Θ from the ensemble of parameters $\{\vec{\theta}_j\}_{j=1}^{S}$ ($S = 96$),

$$\Theta = \langle (\vec{\theta} - \langle\vec{\theta}\rangle)(\vec{\theta} - \langle\vec{\theta}\rangle)^T \rangle$$

where the angle brackets denote ensemble average. Consider $\vec{\theta}$ as Gaussian random vector of size 11×1, then its joint probability density function is given by

$$P(\vec{\theta}) \propto \exp\left(-\frac{1}{2}(\vec{\theta} - \langle\vec{\theta}\rangle)^T \Theta^{-1} (\vec{\theta} - \langle\vec{\theta}\rangle)\right)$$

Recall that $\vec{\theta}$ was sampled according to a Boltzmann distribution with cost function E/T. Assuming that the ensemble average $\langle\vec{\theta}\rangle$ corresponds to the minimum of the cost function E/T which should be approximately quadratic in the vicinity of the minimum, the Hessian H defined as the second derivatives of E/T ($H_{ij} = \partial^2(E/T)/\partial\theta_i\partial\theta_j$) can be used to approximate the shape of the cost surface:

$$P(\vec{\theta}) \propto \exp\left(-\frac{E}{T}\right) \propto \exp\left(-\frac{1}{2}(\vec{\theta} - \langle\vec{\theta}\rangle)^T H (\vec{\theta} - \langle\vec{\theta}\rangle)\right)$$

Comparing terms in $P(\vec{\theta})$, we can see that in this approximation $H = \Theta^{-1}$ (see Supporting Information for details, cf. ref 71). An eigenvalue and eigenvector decomposition of the covariance matrix, called principal component analysis (PCA), tells us the directions along which $\vec{\theta}$ varies most, as each eigenvalue is equal to the variance of $\vec{\theta}$ projected onto the corresponding eigenvector.[72] For parameter inference, we are most interested in the directions of least variability, which correspond to the smallest eigenvalues of Θ, and so it is conventional to instead perform PCA on the inverse of the covariance matrix (H), where the largest eigenvalues now correspond to the "stiff modes" of the parameter ensemble, while the smallest eigenvalues correspond to the "soft modes". (If $\Theta = V\cdot\Lambda\cdot V^{-1}$ where V is a matrix of eigenvectors and Λ is a diagonal matrix containing the eigenvalues, then $H = \Theta^{-1} = V\cdot\Lambda^{-1}\cdot V^{-1}$, and the eigenvalues of H are just the inverses of the eigenvalues of Θ.)

For our analysis, the eigenvalues of the inverse of the empirical covariance matrix and their associated eigenvectors are shown in Figure 9B, listed (from 1 to 11) in order of decreasing eigenvalue, i.e., from stiff to soft. These results show that the eigenvalue spacing is almost uniform in log space; there is no clear cutoff between stiff and soft eigenvalues. When the full set of parameters

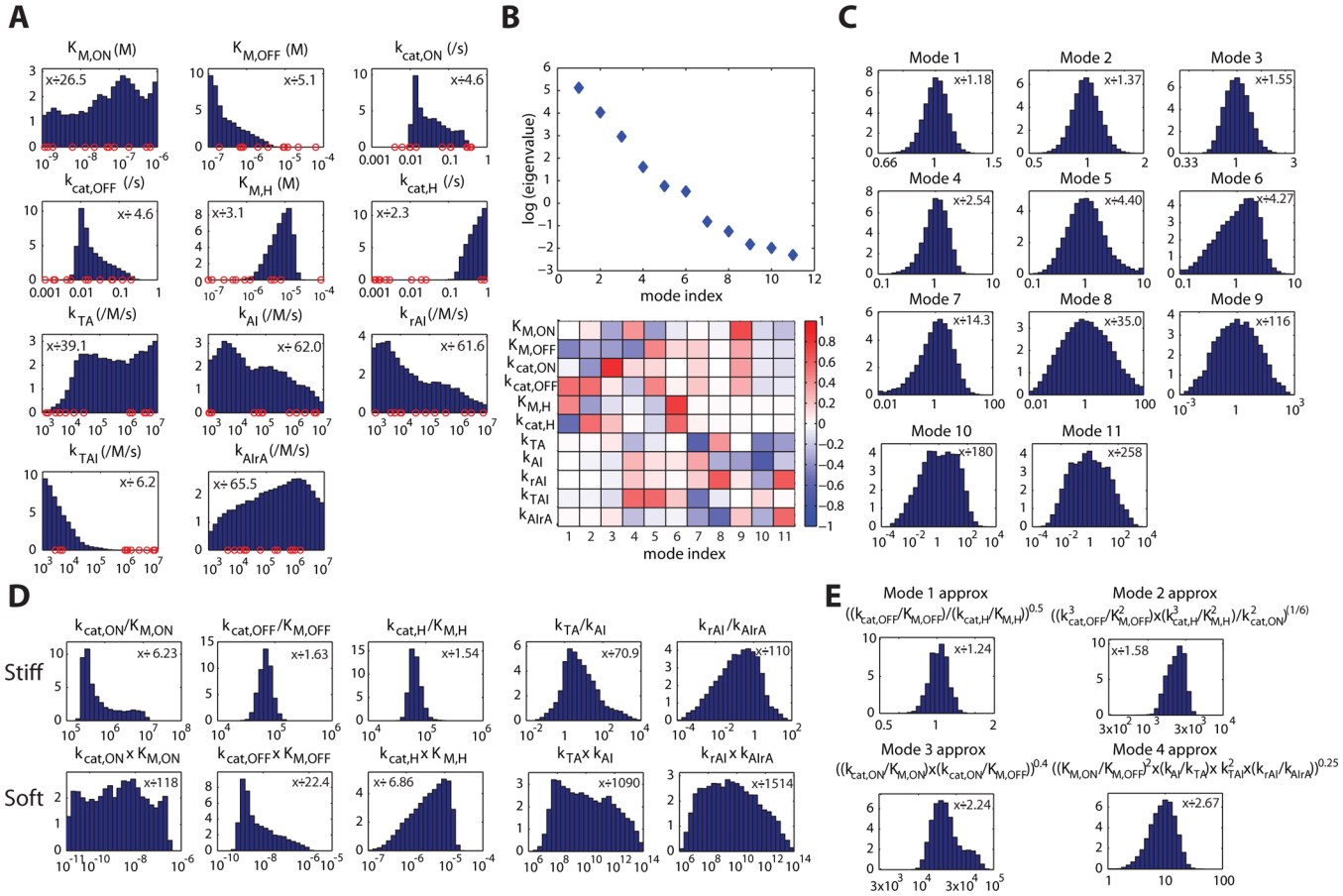


Figure 9. Histograms and mode spectrum analysis of the *a posteriori* parameter set distribution. For histograms, *y*-axes are in units of 100,000 samples. (A) The log-scale parameter values ($\vec{\theta}$) for all 12 trajectories of 390,000 iterations of Monte Carlo sampling, plotted as histograms. The *x*-axes are in log scale with the respective minimum and maximum values corresponding to the prespecified range of each parameter. The 12 initial values for parameters are marked by red circles. The values after '× ÷' symbol indicate the 95% intervals for the histogram ((97.5th percentile/2.5th percentile)$^{1/2}$). (B) Mode spectrum and eigenvectors from PCA of the inverse of the empirical covariance matrix Θ for the set of 96 independent parameter samples. The natural log is used; eigenvector-eigenvalue correspondence is indicated by the mode numbers 1−11: small numbers correspond to large eigenvalues (stiff modes), and large numbers correspond to small eigenvalues (soft modes). (C) Histograms for log-scale parameter values projected onto the corresponding eigenvectors after subtracting the mean. Note that 1's on the *x*-axes correspond to $\langle\vec{\theta}\rangle$; the *x*-axes indicate fold changes in the original parameter space. (D) Small and large variability for two-parameter combinations. Variations of these parameter ratios lie in the direction of stiff modes in contrast to the corresponding parameter multiplications. (E) Histograms for approximate parameter combinations for stiff modes as shown in panel C.

from the sampling trajectories (the same as in Figure 9A) are replotted after subtracting the mean (in log space) and projecting onto each eigenvector, the variances of the resulting distributions increase from the stiffest to the softest modes as expected (Figure 9C), indicating that the selection of 96 independent parameter sets used for PCA was indeed representative of the full set of Boltzmann samples. These "sloppy model" features have been observed in many other high-dimensional multiparameter nonlinear models,[60] as typically encountered in models for biochemical regulatory networks in biological organisms.[61]

By analyzing the mode associated with each eigenvalue, we can obtain valuable information on which parameters are well constrained, which vary together, and which are unconstrained. Although the eigenvector projections are the weighted linear combinations of all 11 parameters, we found that typically only a few parameters significantly contributed for any given eigenvector projection; white-colored entries for eigenvector matrix *V* in Figure 9B indicate negligible contribution. If two log-valued parameters, say, log(X) and log(Y) contribute to and have opposite signs in a stiff eigenvector, the training data effectively constrains log(X/Y), and therefore X/Y but not X or Y individually. If, instead, log(X) and log(Y) have opposite signs in a soft eigenvector, a wide range of X/Y values could fit the training data set equally well.

Consider the stiffest mode (mode 1) where eigenvector elements corresponding to $K_{M,OFF}$, $k_{cat,OFF}$, $K_{M,H}$, and $k_{cat,H}$ contributed significantly, with positive signs for $k_{cat,OFF}$ and $K_{M,H}$ and negative signs for $K_{M,OFF}$ and $k_{cat,H}$. The opposite signs between $K_{M,OFF}$ and $k_{cat,OFF}$ indicate that these two values change in opposite directions in mode 1, which would change the OFF-state switch transcription rate significantly. (Since we are using log-valued parameters, this eigenvector direction amounts to $\log(k_{cat,OFF})$ and $-\log(K_{M,OFF})$ changing together, corresponding to a change in $\log(k_{cat,OFF}/K_{M,OFF})$; more generally, coefficients in the eigenvector would correspond to exponents in the product of (unmodified) model parameters.) Similarly, the opposite signs between $K_{M,H}$ and $k_{cat,H}$ indicate that these two values also change in opposite directions, affecting the degradation rate significantly. Together, the stiffest mode is captured by an increased OFF-state transcription rate

dx.doi.org/10.1021/sb300018h | *ACS Synth. Biol.* 2012, 1, 299−316

with a decreased degradation rate and vice versa, i.e., a "leakiness/degradation balance" (cf. Figure 9E for the approximate parameter combinations deduced from eigenvector elements). The second stiff mode (mode 2) was dominated by contributions from several enzyme parameters: $k_{cat,ON}$, $K_{M,OFF}$, $k_{cat,OFF}$, $K_{M,H}$, and $k_{cat,H}$. Together, the signs of eigenvector elements corresponding to these parameters indicate that this mode is captured by an increase of the ON-state transcription rate with a corresponding decrease of the OFF-state transcription and degradation rates, i.e., a "production/degradation balance". The third stiff mode (mode 3) was dominated by contributions from $K_{M,ON}$, $k_{cat,ON}$, and $K_{M,OFF}$, i.e., the "ON-state transcription speed".

Additional trends can be seen by projecting the eigenvectors onto planes of paired parameters such as, e.g., $k_{cat,ON}$ and $K_{M,ON}$; eigenvectors in quadrants II and IV indicate those reflecting the ratio, rather than product, of the parameters. As seen in Supplementary Figure S3, the stiff modes tended to be sensitive to ratios of enzyme parameters such as, e.g., $k_{cat,ON}/K_{M,ON}$. Consequently, the variability over the Monte Carlo parameter sampling trajectories are indeed small for natural enzyme parameter ratios (Figure 9D). This should not come as a surprise in that the "effective" enzyme rates are captured by k_{cat} and K_M ratios rather than the individual parameters. (However, it is not a trivial observation, as the K_Ms alone appear in the formulas for enzyme concentrations, cf. Box 1.) On the other hand, none of the hybridization rate parameters showed significant contributions for the three stiffest modes. We expect that the relative "stiffness" of the hybridization rate parameters would likely have increased if the kinetic data on switch states, e.g., Figure 7B, were also included in the training set.

Now we examine some of the soft modes revealed in the eigenvector projections. The softest mode (mode 11) is mostly captured by concerted changes of hybridization rate parameters k_{rAI} and k_{AIrA}, i.e., an "annihilation/release balance" (Figure 9B; cf. Supplementary Figure S3). Similarly, the next soft mode (mode 10) is mostly captured by concerted changes of k_{TA} and k_{AI} (with smaller contribution from k_{rAI} and k_{AIrA}), i.e., an "annihilation/activation balance". Consider the concerted changes of k_{TA} and k_{AI}, which would affect the time-scale of hybridization reactions among A and T and dI, but preserve the relative distribution of A among its two binding partners T and dI, and consequently preserve the switch states. On the other hand, their changes in opposite directions would affect switch states such that they are captured by more stiff modes (modes 4, 7, and 8); see Figure 9D and Supplementary Figure S3. Some of the soft modes can be explained by concerted changes in enzyme parameters, e.g., mode 9 is associated with preserving the production rate, $k_{cat,ON}/K_{M,ON}$; this is expected because stiff modes are orthogonal to soft modes. Of note, the highly asymmetric nature of cost functions often present numerical challenges in exploring soft modes as completely as stiff modes; therefore, the stiffest few modes would in general agree better between different sampling runs and approximation schemes.[60]

The ensemble parameter distributions and mode spectrum analysis presented here highlight the advantages of the Bayesian ensemble approach: high-dimensional multiparameter models can be effectively analyzed and used for prediction. This contrasts with choosing a single best-fit parameter, which is unlikely to result in accurate parameter estimation and accurate predictions (even for synthetic and complete data[60]) and which also would not be able to provide confidence intervals on parameter values and predictions. Therefore, the use of the Bayesian ensemble approach together with detailed mechanistic models can allow effective analysis of relationships among parameters with respect to experimental data and offer effective generalization for predicting experiments under novel conditions.

CONCLUSIONS

In this work we have designed and synthesized an activating cascade for *in vitro* transcriptional circuits and demonstrated its use as a single-switch bistable system. When a simple mechanistic model (that reflects only our design intentions) was trained on limited data using a Bayesian ensemble parameter approach, system sensitivities to DNA and enzyme concentrations were qualitatively and sometimes quantitatively predicted using the ensemble parameter set, with confidence intervals for parameter values and predictions.

This study is an example of the reductionist principle at work for a nontrivial programmable biochemical circuit, proceeding from high-level system specification down to low-level component implementation then back up from component properties to system-level behaviors. In comparison to the bistable circuit comprising two mutually inhibiting switches,[37] the self-activating switch shown here is simpler (four strands vs six) but more delicate to design: DNA hybridization reactions can no longer be made arbitrarily strongly favored and irreversible and it is difficult to eliminate spurious reactions such as "interfering", "recovering", and "recapturing". Fortunately, due to the predictable structure and kinetics of nucleic acids, it is straightforward to evaluate component-level interactions computationally and experimentally, allowing for the design of thermodynamic and kinetic constraints at the sequence-level. Fulfilling the promise of predictable system design, when the elementary reactions are combined as a working system, the expected bistable behavior was observed, providing effective use of reductionism in forward engineering. Further, the modularity allows the straightforward rearrangement of switch motifs toward more complex working systems.[38,39]

In this work, we chose to explore the Bayesian ensemble approach over the common practice of choosing a single LSE best-fit parameter set for the mathematical model of self-activating switch. Our relatively simple biochemical network model (simpler than typical mechanistic models of *in vivo* networks) shows features of "sloppy models",[60,61] wherein only a few "stiff" parameter combinations can be well-constrained by experimental data, indicating that the parameter estimation from a single best fit would likely be under-constrained and generalize poorly for novel experimental conditions. Our results support the notion that Bayesian ensemble techniques, by virtue of approximating important marginal distributions of high-dimensional probability distributions generated by complex models, make mechanistic biochemical models useable for prediction and inference.[73] Nonetheless, because the "stiffness" of parameters, the precision with which the combined parameter values can be inferred, depends on the number, precision, and type of experimental data available, careful experimental design and cautious interpretation of modeling and analysis results would be required. Beyond analysis, Bayesian inference could be used in design tools that specify "design objectives" for synthetic biology applications.[63,64]

The analysis of our relatively small system with only 11 parameters took up to 2 weeks on 12 single-core computers working in parallel. The amount of time required to obtain ensemble parameter sets depends on the simulation time for any given parameter set and the equilibration time for Monte Carlo sampling: the former depends on the model complexity as well

as stiffness of numerical solution (e.g., the simulation slowed down by several folds when we widened the parameter range from roughly 2 to roughly 4 orders of magnitude), while the latter depends on the shape of energy landscape. We have made no attempt thus far to optimize for speed. Other works have shown that similar Bayesian ensemble techniques are feasible for more complex models with up to a few hundred parameters.[60,62] Nonetheless, computing power and time could become a bottleneck for applying ensemble Bayesian inference to large biochemical circuits, and more efficient methods should be a subject of future studies.

The design, construction, and analysis of *in vitro* transcriptional circuits can serve as a valuable exercise by making small, incremental steps toward greater complexity, taking advantage of the modular, reconfigurable, and predictable genelet architecture. Our system is simple enough that we know, in principle, all major reactions and components involved; each incremental step, from elementary reactions to the working system, can be characterized to identify the causes of discrepancies between prediction and observation. It is noteworthy that continuously adjustable concentrations of species (rather than their molecular nature) determine most reaction rates within the system, allowing easy "parameter sweeps" to fine-tune system behaviors. Further, due in part to its programmability to achieve complex dynamics, our transcriptional circuits capture examples of common challenges for analyzing complex biochemical networks, e.g., model identifiability, hidden reactions, sloppy parameter values, and the ability to generalize. Therefore, *in vitro* systems that can effectively be used to explore the complexity gap between too-simple biochemical components and too-complex biological organisms could be an effective training ground for studying how system-level behavior arises from the component-level interactions within biochemical systems.

Our modeling technique aims to combine the strengths of systems biology and classical reductionist approaches. The advancement of high-throughput methods provides an opportunity for systems-level analysis to be grounded in molecular-level understanding, resulting in a continuous spectrum of knowledge.[74,75] However, the efficient modeling and analysis of biological systems from the molecular level to the systems level has rarely been realized. Systems biology studies of ill-characterized cellular (sub)systems frequently take the form of a top-down approach to identify correlations between the various components of the systems, and the resulting inductive discoveries rarely lead to molecular mechanistic knowledge.[76] Therefore, systems biology studies, whenever possible, must incorporate the more mechanism-based understanding of bottom-up approaches to constrain regulatory network structures and parameters.[77] On the other hand, the classical reductionist approach can be characterized as the other end of the spectrum, where one takes apart the system under study, carefully determines elementary reaction rates, and tests a mechanistic model with parameters thus obtained; this is possible in *in vitro* settings as demonstrated in ref 31. While the classical reductionist approach is desirable when possible, it may not be practical as the system complexity increases, e.g., in the cellular context. Integrating the two ends of the spectrum, here we used the reductionist approach to construct the model structure starting from individual reactions, while we used the "holistic" Bayesian ensemble approach to constraint parameters. This systems biology approach grounded in reductionism allowed us to make decent predictions for novel experiments without requiring precise measurements of individual parameters.

As the complexity of synthetic systems gradually increases, we will of necessity explore when and how the reductionist principle falls short of expectation. Examining the dynamic features that deviated from Bayesian ensemble predictions of the mechanistic model, a breakdown of the reductionist method that signals a lack in understanding, would likely identify new mechanistic interactions that would presumably be important for other genelet circuits as well. Each new understanding will help push the complexity limit of what we can build and analyze. Ultimately though, robust architectural features will be necessary to successfully scale up system complexity, e.g., modularity, redundancy, and buffering.[78] However, the "sloppy" behavior of multiparameter systems biology models, where wide variations in parameters leaves the system behavior unchanged, begs the question of how one should quantify robustness on solid mathematical grounds.[79] The ability to engineer synthetic biochemical systems with flexible, modular, and programmable architecture will help us explore how robustness at different levels can be organized and coupled across scales in biological designs.

MATERIALS AND METHODS

DNA Oligonucleotides and Enzymes. The sequence of all DNA molecules and expected RNA transcript sequences were chosen to minimize the occurrence of alternative secondary structures, checked by the Vienna group's DNA and RNA folding program.[47] According to the nearest-neighbor model, the energetics for formation of DNA−DNA duplexes or chimeric RNA−DNA duplexes can be described by 16 nearest-neighbor parameters for helix propagation and one parameter for duplex initiation. The predicted free energy changes due exclusively to nearest-neighbor stacking interactions (ignoring duplex initiation energy) were calculated with DNA−DNA parameters[54] for the formation of T·A and A·dI complexes and RNA−DNA parameters[55] for the formation of T·rA and rA·dI complexes. For calculations, the regions that are newly formed upon hybridization are considered as perfectly complementary oligonucleotides of the same length, e.g. ΔG(rCUUAC/dGTAAG) = ΔG(rCU/dAG) + ΔG(rUU/dAA) + ΔG(rUA/dTA) + ΔG(rAC/dGT).

All DNA oligonucleotides were purchased from Integrated DNA Technologies (USA). In addition to unlabeled A and dI, A labeled with Cy5 at the 3′ end and dI labeled with IowaBlack-RQ at the 5′ end were purchased. The T7 RNA polymerase (enzyme mix), transcription buffer, and NTP were purchased as part of the MEGAshortscript T7 Kit (Ambion, Austin, TX, USA; #1354). RNase H (Ambion; #2293) was purchased. RNase R was a gift from Dr. Murray P. Deutscher at the University of Miami School of Medicine. Note that according to the manufacturer's patent (# 5,256,555), the enzyme mix of the MEGAshortscript T7 Kit contains pyrophosphatase to extend the lifetime of the transcription reaction; since pyrophosphatase is involved in regulating the power supply for our transcriptional circuits and is not directly involved in the dynamics, we neglect this enzyme in our models and do not call it an "essential enzyme" for the circuit dynamics. The nominal concentrations of enzyme stocks were estimated as follows: for RNase H, we used the value quoted by the manufacturer; for RNase R, we used the value provided by Dr. Deutscher; for RNAP, we used an estimated value obtained by extinction coefficient measurement. We noticed that different enzyme batches with the same nominal concentrations showed different enzyme activities by a few fold. Therefore, we did not attempt to provide a quantitative comparison among the experimental

data obtained using different enzyme batches: Figure 5 (batch #1), Figure 6 (batch #2), and Figures 7 and 8 (batch #3).

Transcription. DNA templates (T-nt and T-t strands) were annealed with 10% (v/v) 10× transcription buffer from 90 to 20 °C over 1 h at 10 times the final concentration used. To the annealed templates, DNA activator and inhibitor, A and dI, were added from high concentration stocks (50 μM), 7.5 mM each NTP, and 8% (v/v) 10× transcription buffer were added. After adding all ssDNA strands and RNA signals, enzymes (RNAP, RNase H, and RNase R) were added and mixed. Transcription reactions for spectrofluorometer experiments were prepared as a total volume of 60 μL. Transcription reactions for gel studies were prepared as a total volume of 10 μL and were stopped by denaturing dye (80% formamide, 10 mM EDTA, 0.01 g XCFF). For the purification of RNA signal rA for the experiments and for gel controls, the full length template side strand (the complement of T-nt rather than T-t) was used to prepare a fully duplex DNA template for rA. The transcription reaction was prepared as a total volume of 60 μL with 0.2 μM fully duplex DNA templates. The transcription condition was the same as above except that no A or dI was added, 20% (v/v) RNAP was used, and RNases were omitted. After a 6 h incubation at 37 °C, the reaction mixture was treated with 2.5 μL DNase I for 30 min to remove DNA template and stopped by the addition of denaturing dye. The reaction mixture was run on 8% denaturing gel, RNA bands were excised and eluted from gel by the crush-and-soak method, ethanol precipitated, and resuspended in water.

Data Acquisition. For spectrofluorometer experiments, excitation and emission for Cy5-labeled A were at 648 and 665 nm. The fluorescence was recorded every minute using a SPEX Fluorolog-3 (Jobin Yvon, Edison, NJ, USA) and normalized against maximum fluorescence (measured in the presence of Cy5-labeled A prior to the addition of quencher-labeled dI) and minimum fluorescence (measured after the addition of excess quencher-labeled dI) accounting for volume increase due to the addition of rA and enzymes. Denaturing polyacrylamide gels (8% 19:1 acrylamide:bis-acrylamide and 7 M urea in TBE buffer) were allowed to run for 50 min with 10 V/cm at 65 °C in TBE buffer (100 mM Tris, 90 mM boric acid, 1 mM EDTA). The 10-base DNA ladder (Invitrogen, Carlsbad, CA, USA; #10821-015) was used as the size marker. The non-denaturing gels (10% 19:1 acrylamide:bis-acrylamide in TAE buffer) were allowed to run for 100 min with 13 V/cm at 4 °C in TAE buffer containing 12.5 mM Mg^{2+} (40 mM Tris-acetate, 1 mM EDTA, 12.5 mM Mg-acetate, pH 8.3). The gels were stained with SYBR gold (Molecular Probes, Eugene, OR, USA; #S-11494), and the gel data were quantitated using the Molecular imager FX (Biorad, Hercules, CA, USA). The total concentrations of RNA product rA in the denaturing gel were measured with respect to 0.5, 1, and 1.5 μM purified rA in control lanes.

Model Simulation. The kinetic simulations and parameter fittings were implemented in MATLAB. Differential equations were solved using the *ode23s* routine. The ensemble parameter sets used in Figure 8 were obtained using the Monte Carlo Bayesian inference approach. See Supporting Information for details. On average, Monte Carlo sampling of 440,000 instances took about 2 weeks on a single-core computer; we used 12 cores in parallel.

DNA Sequences. T-nt (106 mer), 5′-CATTAGTGTCGTTCGTTCACAGTAATACGACTCACTATAGGGAGAAACAAAGAACGAACGACACTAATGAACTACTACTACACACTAATACTGACAAAGTCAGAAA-3′.

T-t (79 mer), 5′-TTTCTGACTTTGTCAGTATTAGTGTGTAGTAGTAGTTCATTAGTGTCGTTCGTTCTTTGTTTCTCCCTATAGTGAGTCG-3′.

A (36 mer), 5′-TATTACTGTGAACGAACGACACTAATGAACTACTAC-3′.

dI (38 mer), 5′-GTGTGTAGTAGTAGTTCATTAGTGTCGTTCGTTCACAG-3′.

rA (67 mer), 5′-GGGAGAAACAAAGAACGAACGACACUAAUGAACUACUACUACACACUAAUACUGACAAAGUCAGAAA-3′.

Note that the RNA activator rA sequence is identical to I1 in ref 37, and therefore, the T-t sequence is also identical to T12-t in ref 37. The self-activating switch reported here is used as part of the Design II oscillator in ref 38, and the repeater motif is also used, configured differently, in the Design I oscillator of that same work.

■ ASSOCIATED CONTENT

Supporting Information

Detailed discussion of Monte Carlo Bayesian inference methods, Figures S1−3, and references. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: winfree@caltech.edu.

Present Address

[¶]Department of Bioengineering, Stanford University, Stanford, California

Author Contributions

[⊥]These authors contributed equally to this work.

P.S., J.K., and E.W. designed research; P.S. performed experiments; P.S., J.K., and E.W. analyzed data; P.S., J.K., and E.W. wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Science Foundation awards 0103002, 0608889, and 0832824 (The Molecular Programming Project); the Human Frontiers Science Program award RGY0074/2006-C; and the Caltech Center for Biological Circuit Design. A preliminary version of this work appeared in ref 80.

■ REFERENCES

(1) Andrianantoandro, E., Basu, S., Karig, D. K., and Weiss, R. (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol. 2*, No. 2006.0028.

(2) Benner, S. A., and Sismour, A. M. (2005) Synthetic biology. *Nat. Rev. Genet. 6*, 533−543.

(3) Khalil, A. S., and Collins, J. J. (2010) Synthetic biology: applications come of age. *Nat. Rev. Genet. 11*, 367−379.

(4) Endy, D. (2005) Foundations for engineering biology. *Nature 438*, 449−453.

(5) Drexler, K. E. (1981) Molecular engineering: an approach to the development of general capabilities for molecular manipulation. *Proc. Natl. Acad. Sci. U.S.A. 78*, 5275−5278.

(6) Merkle, R. C. (1999) Biotechnology as a route to nanotechnology. *Trends Biotechnol. 17*, 271−274.

(7) Seeman, N. C. (2010) Structural DNA nanotechnology: growing along with Nano Letters. *Nano Lett. 10*, 1971−1978.

(8) Regenmortel, M. V. (2004) Reductionism and complexity in molecular biology. *EMBO Rep. 5*, 1016−1020.
(9) Gatherer, D. (2010) So what do we really mean when we say that systems biology is holistic? *BMC Syst. Biol. 4*, 22.
(10) Elowitz, M., and Lim, W. A. (2010) Build life to understand it. *Nature 468*, 889−890.
(11) Isaacs, F. J., Dwyer, D. J., and Collins, J. J. (2006) RNA synthetic biology. *Nat. Biotechnol. 24*, 545−554.
(12) Bashor, C. J., Horwitz, A. A., Peisajovich, S. G., and Lim, W. A. (2010) Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu. Rev. Biophys. 39*, 515−537.
(13) Kwok, R. (2010) Five hard truths for synthetic biology. *Nature 463*, 288−290.
(14) Hodgman, C. E., and Jewett, M. C. (2012) Cell-free synthetic biology: thinking outside the cell. *Metab. Eng. 14*, 261−269.
(15) Rondelez, Y. (2011) Breaking down complexity. *Physics 4*, 8.
(16) Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol. 19*, 751−755.
(17) Noireaux, V., Bar-Ziv, R., and Libchaber, A. (2003) Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. U.S.A. 100*, 12672−12677.
(18) Noireaux, V., Bar-Ziv, R., Godefroy, J., Salman, H., and Libchaber, A. (2005) Toward an artificial cell based on gene expression in vesicles. *Phys. Biol. 2* (3), 1−8.
(19) Shimizu, Y., Kanamori, T., and Ueda, T. (2005) Protein synthesis by pure translation systems. *Methods 36*, 299−304.
(20) Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., and Swartz, J. R. (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol. 4*, 220.
(21) Noireaux, V., Maeda, Y., and Libchaber, A. (2011) Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc. Natl. Acad. Sci. U.S.A. 108*, 3473−3480.
(22) Karzbrun, E., Shin, J., Bar-Ziv, R., and Noireaux, V. (2011) Coarse-grained dynamics of protein synthesis in a cell-free system. *Phys. Rev. Lett. 106*, 048104.
(23) Karig, D. K., Iyer, S., Simpson, M. L., and Doktycz, M. J. (2012) Expression optimization and synthetic gene networks in cell-free systems. *Nucleic Acids Res. 40*, 3763−3774.
(24) Shin, J., and Noireaux, V. (2012) An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synth. Biol. 1*, 29−41.
(25) Wlotzka, B., and McCaskill, J. S. (1997) A molecular predator and its prey: coupled isothermal amplification of nucleic acids. *Chem. Biol. 4*, 25−33.
(26) Benenson, Y., Paz-Elizur, T., Adar, R., Keinan, E., Livneh, Z., and Shapiro, E. (2001) Programmable and autonomous computing machine made of biomolecules. *Nature 414*, 430−434.
(27) Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., and Shapiro, E. (2004) An autonomous molecular computer for logical control of gene expression. *Nature 429*, 423−429.
(28) Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., Oyama, T., and Kondo, T. (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. *Science 308*, 414−415.
(29) Simpson, Z. B., Tsai, T. L., Nguyen, N., Chen, X., and Ellington, A. D. (2009) Modelling amorphous computations with transcription networks. *J. R. Soc. Interface 6*, S523−S533.
(30) Ayukawa, S., Takinoue, M., and Kiga, D. (2011) RTRACS: a modularized RNA-dependent RNA transcription system with high programmability. *Acc. Chem. Res. 44*, 1369−1379.
(31) Montagne, K., Plasson, R., Sakai, Y., Fujii, T., and Rondelez, Y. (2011) Programming an *in vitro* DNA oscillator using a molecular networking strategy. *Mol. Syst. Biol. 7*, 466.
(32) Seelig, G., Soloveichik, D., Zhang, D. Y., and Winfree, E. (2006) Enzyme-free nucleic acid logic circuits. *Science 314*, 1585−1588.
(33) Zhang, D. Y., Turberfield, A. J., Yurke, B., and Winfree, E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA. *Science 318*, 1121−1125.
(34) Soloveichik, D., Seelig, G., and Winfree, E. (2010) DNA as a universal substrate for chemical kinetics. *Proc. Natl. Acad. Sci. U.S.A. 107*, 5393−5398.
(35) Qian, L., Winfree, E., and Bruck, J. (2011) Neural network computation with DNA strand displacement cascades. *Nature 475*, 368−372.
(36) Kim, J., Hopfield, J. J., and Winfree, E. (2004) Neural network computation by *in vitro* transcriptional circuits. *Advances in Neural Information Processing Systems (NIPS)*, Vol. *17*, pp 681−688, MIT Press, Cambridge, MA.
(37) Kim, J., White, K. S., and Winfree, E. (2006) Construction of an *in vitro* bistable circuit from synthetic transcriptional switches. *Mol. Syst. Biol. 2*, 68.
(38) Kim, J., and Winfree, E. (2011) Synthetic *in vitro* transcriptional oscillators. *Mol. Syst. Biol. 7*, 465.
(39) Franco, E., Friedrichs, E., Kim, J., Jungmann, R., Murray, R., Winfree, E., and Simmel, F. C. (2011) Timing molecular motion and production with a synthetic transcriptional clock. *Proc. Natl. Acad. Sci. U.S.A. 108*, E784−E793.
(40) Hooshangi, S., Thiberge, S., and Weiss, R. (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. U.S.A. 102*, 3581−3586.
(41) Yurke, B., Turberfield, A., Mills, A., Jr, Simmel, F., and Neumann, J. (2000) A DNA-fuelled molecular machine made of DNA. *Nature 406*, 605−608.
(42) Martin, C. T., and Coleman, J. E. (1987) Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry 26*, 2690−2696.
(43) Yurke, B., and Mills, A. P., Jr. (2003) Using DNA to power nanostructures. *Genetic Program. Evolvable Machines 4*, 111−122.
(44) Zhang, D. Y., and Winfree, E. (2009) Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc. 131*, 17303−17314.
(45) Lima, W. F., and Crooke, S. T. (1997) Cleavage of single strand RNA adjacent to RNA-DNA duplex regions by *Escherichia coli* RNase H1. *J. Biol. Chem. 272*, 27513−27516.
(46) Cheng, Z.-F., and Deutscher, M. P. (2002) Purification and characterization of the *Escherichia coli* exoribonuclease RNase R. *J. Biol. Chem. 277*, 21624−21629.
(47) Hofacker, I. L., Fontana, W., Stadler, P., Bonhoeffer, L., Tacker, M., and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem. 125*, 167−188.
(48) Seeman, N. C. (1982) Nucleic-acid junctions and lattices. *J. Theor. Biol. 99*, 237−247.
(49) Seeman, N. C. (1990) De novo design of sequences for nucleic acid structural engineering. *J. Biomol. Struct. Dyn. 8*, 573−581.
(50) Triana-Alonso, F. J., Dabrowski, M., Wadzack, J., and Nierhaus, K. H. (1995) Self-coded 3′-extension of run-off transcripts produces abberant products during *in vitro* transcription with T7 RNA polymerase. *J. Biol. Chem. 270*, 6298−6307.
(51) McGinness, K. E., and Joyce, G. F. (2002) Substitution of ribonucleotides in the T7 RNA polymerase promoter element. *J. Biol. Chem. 277*, 2987−2991.
(52) Milligan, J. F., and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol. 180*, 51−62.
(53) Martin, C. T., Muller, D. K., and Coleman, J. E. (1988) Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry 27*, 3966−3974.
(54) SantaLucia, J., Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.
(55) Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry 34*, 11211−11216.

(56) Gao, Y., Wolf, L. K., and Georgiadis, R. M. (2006) Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Res. 34*, 3370−3377.
(57) Bloomfield, V. A., Crothers, D. M., Tinoco, I., Hearst, J. E., Wemmer, D. E., Killman, P. A., Turner, D. H. (1999) *Nucleic Acids: Structures, Properties, and Functions*, University Science Books, Sausalito, CA.
(58) Jia, Y., and Patel, S. S. (1997) Kinetic mechanism of transcription initiation by bacteriophage T7 RNA polymerase. *Biochemistry 36*, 4223−4232.
(59) Marras, S. A. E., Kramer, F. R., and Tyagi, S. (2002) Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res. 30*, e122.
(60) Brown, K. S., and Sethna, J. P. (2003) Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E 68*, 021904.
(61) Gutenkunst, R., Waterfall, J., Casey, F., Brown, K., Myers, C., and Sethna, J. (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol. 3*, 1871−1878.
(62) Chen, W. W., Schoeberl, B., Jasper, P. J., Niepel, M., Nielsen, U. B., Lauffenburger, D. A., and Sorger, P. K. (2009) Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol. Syst. Biol. 5*, 239.
(63) Barnes, C. P., Silk, D., and Stumpf, M. P. H. (2011) Bayesian design strategies for synthetic biology. *Interface Focus 1*, 895−908.
(64) Barnes, C. P., Silk, D., Sheng, X., and Stumpf, M. P. H. (2011) Bayesian design of synthetic biological systems. *Proc. Natl. Acad. Sci. U.S.A. 108*, 15190−15195.
(65) Rizzo, J., Gifford, L. K., Zhang, X., Gewirtz, A. M., and Lu, P. (2002) Chimeric RNA-DNA molecular beacon assay for ribonuclease H activity. *Mol. Cell. Probes 16*, 277−283.
(66) Kim, J. (2009) Synthetic networks. *Automation in Proteomics and Genomics*, pp 251−271, John Wiley & Sons, Ltd., New York.
(67) Rondelez, Y. (2012) Competition for catalytic resources alters biological network dynamics. *Phys. Rev. Lett. 108*, 018102.
(68) Hengl, S., Kreutz, C., Timmer, J., and Maiwald, T. (2007) Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics 23*, 2612−2618.
(69) Kuzmine, I., and Martin, C. T. (2001) Pre-steady-state kinetics of initiation of transcription by T7 RNA polymerase: a new kinetic model. *J. Mol. Biol. 305*, 559−566.
(70) Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F. (2006) Sensitivity analysis practices: strategies for model-based inference. *Reliability Engineering and System Safety 91*, 1109−1125.
(71) Yuen, K.-V. (2010) *Bayesian Methods for Structural Dynamics and Civil Engineering*. John Wiley & Sons, Singapore.
(72) Jolliffe, I. T. (2002) *Principal Component Analysis.* 2nd ed., Springer, New York.
(73) Balding, D. (2011) Inference in complex systems. *Interface Focus 1*, 805−806.
(74) Kitano, H. (2002) Systems biology: a brief overview. *Science 295*, 1662−1664.
(75) Sauer, U., Heinemann, M., and Zamboni, N. (2007) Getting closer to the whole picture. *Science 316*, 550−551.
(76) Bruggeman, F. J., and Westerhoff, H. V. (2007) The nature of systems biology. *Trends Microbiol. 15*, 45−50.
(77) Friedman, N. (2004) Inferring cellular networks using probabilistic graphical models. *Science 303*, 799−805.
(78) Kitano, H. (2004) Biological robustness. *Nat. Rev. Genet. 5*, 826−837.
(79) Daniels, B., Chen, Y.-J., Sethna, J., Gutenkunst, R., and Myers, C. (2008) Sloppiness, robustness, and evolvability in systems biology. *Curr. Opin. Biotechnol. 19*, 389−395.
(80) Subsoontorn, P., Kim, J., and Winfree, E. (2011) Bistability of an *in vitro* synthetic autoregulatory switch. http://arxiv.org/abs/1101.0723.