ACS SyntheticBiology

Letter

pubs.acs.org/synthbio

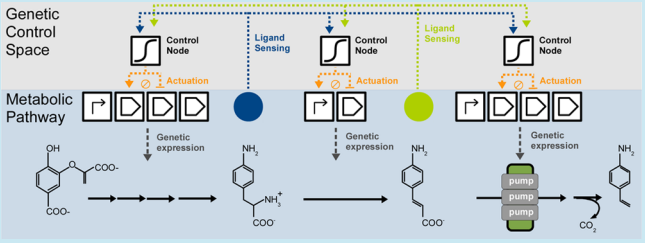
# Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways

Jason T. Stevens and James M. Carothers*

Departments of Chemical Engineering and Bioengineering, Molecular Engineering & Sciences Institute, and Center for Synthetic Biology, University of Washington, Seattle, Washington 98195, United States

Supporting Information

**ABSTRACT:** Engineered metabolic pathways can be augmented with dynamic regulatory controllers to increase production titers by minimizing toxicity and helping cells maintain homeostasis. We investigated the potential for dynamic RNA-based genetic control systems to increase production through simulation analysis of an engineered *p*-aminostyrene (*p*-AS) pathway in *E. coli*. To map the entire design space, we formulated 729 unique mechanistic models corresponding to all of the possible control topologies and mechanistic implementations in the system under study. Two thousand sampled simulations were performed for each of the 729 system designs to relate the potential effects of dynamic control to increases in *p*-AS production (total of $3 \times 10^6$ simulations). Our analysis indicates that dynamic control strategies employing aptazyme-regulated expression devices (aREDs) can yield >10-fold improvements over static control. We uncovered generalizable trends in successful control architectures and found that highly performing RNA-based control systems are experimentally tractable. Analyzing the metabolic control state space to predict optimal genetic control strategies promises to enhance the design of metabolic pathways.




**KEYWORDS:** *synthetic biology, metabolic engineering, dynamic control, RNA device, simulation*

Metabolic pathways found in natural systems have evolved complex control circuitry that responds dynamically to chemical and physical inputs to actuate genetic regulatory functions, minimize the buildup of toxic intermediates, and maintain homeostasis.[1] Many pathways that are targets for metabolic engineering contain cytotoxic intermediates; metabolites that are inadvertently lost from the cell; or enzymes or proteins whose overexpression depletes metabolic resources, generates toxic compounds, or otherwise has deleterious effects on cellular physiology.[2] Constructing chemically responsive genetic systems to solve control problems in engineered metabolic pathways could therefore lead to optimized production titers[3] of industrially and medically relevant compounds in microbial hosts.

In principle, there are a number of regulatory mechanisms that could be applied to solve pathway control problems and enable high levels of production. In seminal work, Farmer and Liao engineered a dynamic controller comprising a promoter and its associated transcriptional activator, which induces transcription in conditions of excess flux, reducing toxic metabolic intermediate (acetate) accumulation and improving final titers of lycopene.[4] More recently, a fatty acid/acyl-CoA-inducible transcriptional repressor (TR), FadR, was used to program enzyme expression levels in response to high levels of fatty acyl-CoA, minimizing the buildup of toxic intermediates and increasing fatty acid ethyl ester (FAEE) biodiesel production.[5] In the latter case, simulation analysis identified a broad range of genetic control parameter values under which the dynamically regulated system was expected to generate higher levels of FAEE than a comparable system of static controllers. Elsewhere, modeling has shown that dynamic control of efflux pump expression[6] can expand the functional design space by removing toxic intermediates or products from cells more effectively than pump expression under static control.[7]

An RNA-based approach for constructing sensors (e.g., ligand-responsive aptamers) and actuators (e.g., catalytic ribozymes) to control mRNA transcript stability may obviate the need to rely on preexisting genetic components and increase the speed and efficacy with which dynamic control systems can be constructed.[8] *In vitro* selection[9] can be used to select aptamers that bind in physiological conditions to a variety of target ligands. Using these aptamers, catalytic ribozymes with activities that are responsive to desired metabolites (aptazymes) can be generated and further assembled into dynamic regulators (sensor-actuators) of gene expression in bacteria, yeast, and mammalian cells.[10−12] Aptazymes and other aptamer-based RNA components have been used as pathway biosensors to implement higher-order functions,[13] for *in vivo* enzyme evolution,[12] and for screening metabolite-producing strains.[14] Building on this work, model-

**Special Issue:** Circuits in Metabolic Engineering

**Received:** December 5, 2013
**Published:** October 14, 2014

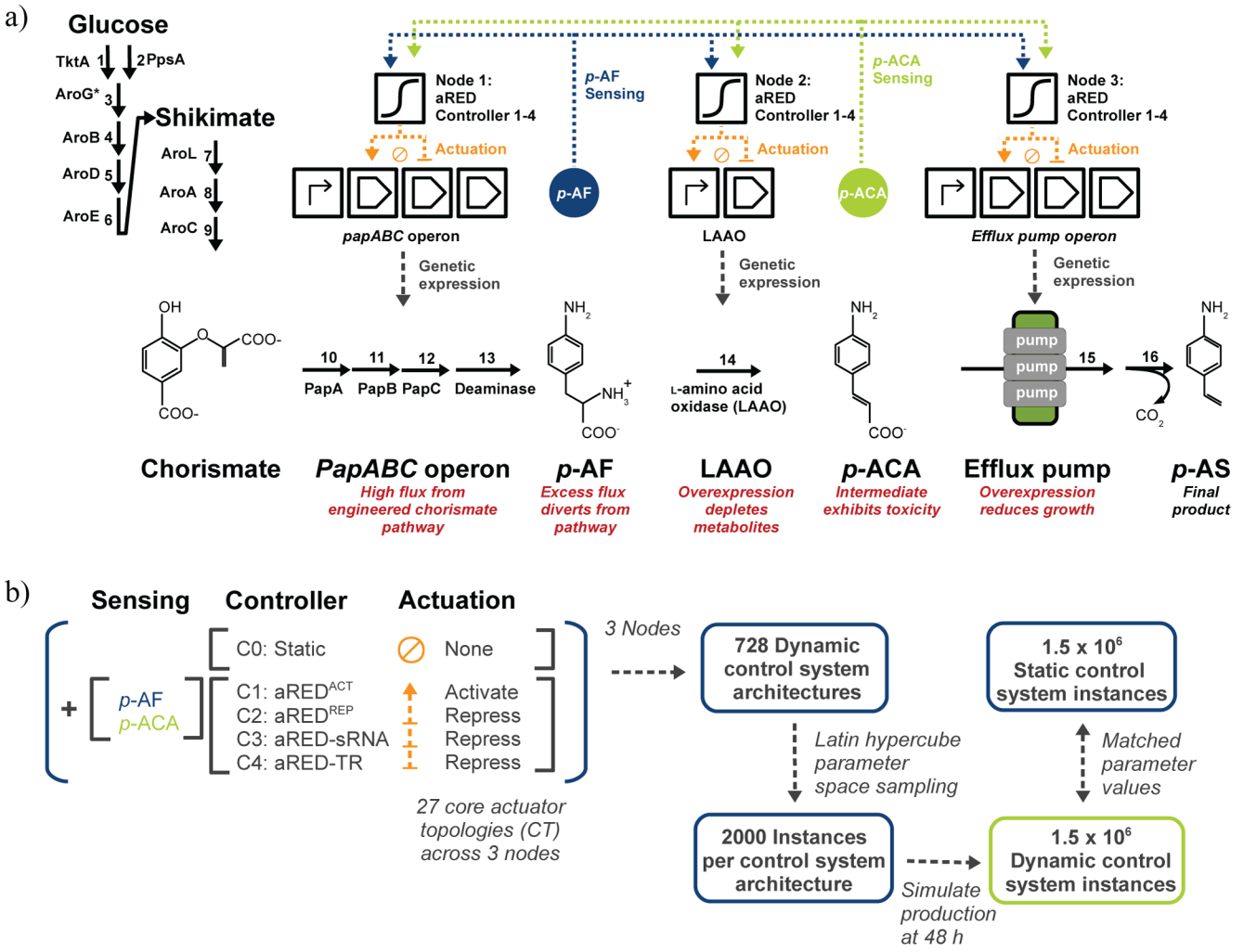



**Figure 1.** Engineered *p*-aminostyrene (*p*-AS) pathway and control system design space. (A) The proposed pathway includes optimized modules for converting glucose to chorismate (steps 1−9),[35] along with the PapABC operon from *Streptomyces venezuelae* and an *E. coli* deaminase to convert chorismate to *p*-aminophenyalanine (*p*-AF, steps 10−13).[36] *p*-AF is converted to *p*-aminocinnamic acid (*p*-ACA) via an L-amino acid oxidase (LAAO, step 14), and *p*-ACA export occurs via an efflux pump (step 15), after which it can be converted to *p*-AS (step 16). Blue and green dashed lines indicate sensing functions, responsive to metabolites, and orange dashed lines indicate actuation functions using aptazyme-regulated expression devices[10] (aREDs) as genetic controllers to integrate the sensing information. (B) There are a total of 729 ways to control the system with four actuators and two sensing molecules (*p*-AF and *p*-ACA), which comprises the control system design space. To map the effect of control on *p*-AS production, unique models were created for all 728 dynamic architectures (the 729th is all static control) and the outputs from each were simulated 2000 times with Latin hypercube-sampled parameter instances, along with 2000 parameter-matched static control reference instances.

driven processes were developed for engineering aptazyme-regulated expression devices (aREDs),[10] to produce quantitatively predictable genetic outputs in *E. coli*. In these devices, *cis*-aptazyme cleavage within the 5′ UTR results in a programmable increase in mRNA half-life in response to a target ligand (Supporting Information Figure 1). Because aREDs can be readily built from component parts separately generated and characterized, it may be possible to engineer and then combine multiple genetic devices to give predictable systems-level functions. If so, fundamental questions about the space of genetic control system architecture designs and the corresponding functions may then become accessible to experiment, including (1) the effect of increasing the complexity of control systems, (2) the degree to which the desired architectures can actually be implemented using RNA-based genetic controls, and (3) whether formal processes can be developed for identifying design specifications.

To begin to address these long-term questions and better understand the space of potential designs for RNA-based genetic control systems, we are engineering an RNA-regulated *p*-aminostyrene (*p*-AS) production pathway in *E. coli* (Figure 1a). *p*-AS is an industrially relevant vinyl aromatic monomer with properties amenable to advanced applications in photonics, photolithography, and biomedicine.[15] Cytotoxity of intermediates in production pathways have made substituted styrenes, like *p*-AS, challenging to produce in microbes.[16] Production of *p*-AS in *E. coli* therefore offers an excellent test case for demonstrating the utility of genetic control systems. The pathway enzymes and metabolic products are well-defined, and it provides control problems to address to optimize titers (Figure 1, in blue). These problems include an intermediate, *p*-AF, leaking from cells;[10] another intermediate, *p*-ACA, exhibiting cytotoxicity (Supporting Information Figure 2a); L-amino acid oxidases (LAAOs) potentially depleting the aromatic amino acid pool and creating hydrogen peroxide as


dx.doi.org/10.1021/sb400201u | *ACS Synth. Biol.* 2015, 4, 107−115

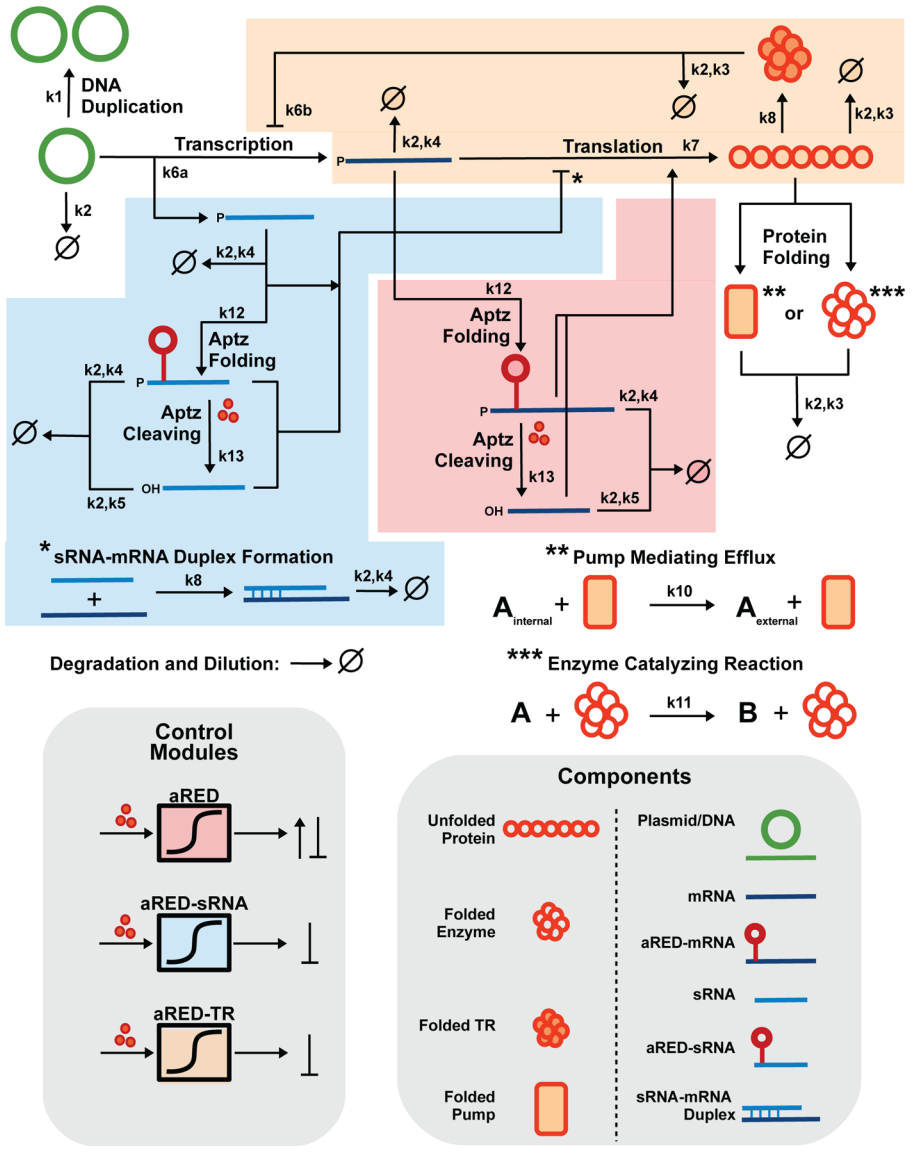



**Figure 2.** Mechanistic model and genetic control mechanisms. Reactions in the red module correspond to aptazyme-regulated expression device (aRED) function. aREDs program genetic expression through ligand-responsive aptazyme cleavage in the 5′ UTR that increases RNA half-life by selectively removing the binding target of RppH-mediated degradation[37] (Supporting Information Figure 1). Reactions in the orange module correspond to transcriptional repressors (TR) where binding operator regions in target promoter sequences downregulates transcription initiation.[38] Reactions in the blue module describe a small RNA (sRNA) mechanism that represses translation by binding a complementary sequence of a target mRNA near the ribosome binding site or start codon and inhibiting ribosomal interactions.[39] The blue or orange modules can be combined with a red module (controller C1, aRED$^{ACT}$, or C2, aRED$^{REP}$) to create an aRED$^{ACT}$-sRNA controller (C3) or an aRED$^{ACT}$-TR controller (C4). Small orange circles represent ligand; A and B represent chemicals. Rate constants are drawn beside arrows. Supporting Information Table 1 gives full descriptions, mathematical expressions, and ranges for the sampled rate constants.

a byproduct; and efflux pump overexpression (for *p*-ACA removal) also generating toxicity (Supporting Information Figure 2b).

Although flux balance analysis[17] and metabolic control analysis[18] can be used to study metabolic pathways and control systems, results obtained with those methods can be difficult to formulate in terms of kinetic parameters for the genetic components needed to build the system. Elsewhere, the importance of developing new formalisms to relate network topology and kinetic parameters in engineered circuits has been noted.[19] To maximize engineering tractability, the challenge is to create approaches for mapping potential design space inputs to production outputs in terms of measurable and tunable kinetic parameters. These parameters can then be associated with specific genetic components required for building metabolic control systems.

Our approach is to overlay a dynamic control layer onto an engineered metabolic pathway and sample kinetic parameters to identify functional control architectures that improve production. Specifically, we automated the generation of coarse-grained mechanistic models for 728 unique control architectures. To sample the underlying design space and account for parameter uncertainty within each of the 728 control architectures, we computed the *p*-AS production yields expected with dynamic control for 2000 unique parameter instances, for a total of $1.5 \times 10^6$ solutions. The impact of dynamic control on *p*-AS production was evaluated by comparison to parameter-matched systems comprising only


dx.doi.org/10.1021/sb400201u | ACS Synth. Biol. 2015, 4, 107−115

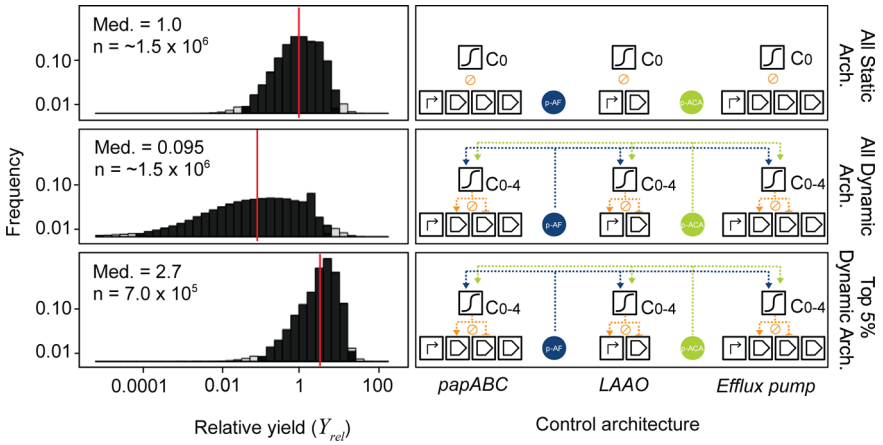



**Figure 3.** Relative production yields across the entire design space. $Log_{10}$ relative yields ($Y_{rel}$) are plotted for all static control parameter instances (top), all dynamic control parameter instances (middle), and all instances for architectures with median $Y_{rel}$ in the top 5% (i.e., $Y_{rel} > 2.11$). The vertical red line denotes median $Y_{rel}$ within a set. Black shading indicates the $Y_{rel}$ 95% confidence interval. As depicted at right, static control systems have no sensing and no actuation, while there are 728 possible combinations of sensing and actuation functions defining the set of dynamic control system architectures under study.

static control functions (i.e., those that could be obtained with promoter,[20] RBS,[21] or ribozyme-regulated expression device[10] (rRED) tuning). We discovered control architectures that dramatically increase the likelihood of generating high *p*-AS production and show that the choice of mechanistic implementation may be important for effectively coordinating the stoichiometries of intermediates, enzymes, and efflux pumps. Moreover, we find that large improvements in *p*-AS yield are within reach of currently achievable controller component specifications and expect that the approach described here will be useful for driving the design of functional metabolic production systems.

We first formulated a genetic control framework for the *p*-AS pathway (Figure 1) consisting of three control nodes, where a node is defined as a possible target of positive or negative dynamic regulation. These three control nodes were chosen to be genetic expression of (1) the Pap operon enzymes (PapA, PapB, and PapC) that convert chorismate to *p*-AF, controlling pathway flux; (2) the LAAO that converts *p*-AF to *p*-ACA, producing a toxic intermediate; and (3) the efflux pump that exports *p*-ACA from the cell. The total number of potential ways of regulating these three nodes, that is, positive, negative, or no dynamic regulation, comprise the control state space. (Note that there are 27 combinations, or core topologies, with the 27th being no dynamic regulation at any node.) Two different ligands were examined as sensing metabolites, and four different genetic regulatory mechanisms were examined as potential means of implementing actuation, at each node (see Figure 2 for mechanistic details). Altogether, the combinations of sensors and actuators lead to 729 distinct architectures for creating a control system (Figure 1b).

While aREDs can be fashioned into positive or negative actuators, we reasoned that coupling dynamic aREDs with transcriptional repressors (TRs) or translation-repressing sRNAs could result in composite devices with enhanced functions compared to aREDs alone. Composite devices have not been built experimentally; however, the individual components have been used with success,[10,22,23] and it is expected that composite devices are experimentally tractable. We thus investigated one positive actuator and three negative actuators, two of which are composite devices. As positive actuators, aREDs were employed where ligand-binding increases genetic output by extending mRNA half-life, increasing target mRNA levels relative to the basal level ($aRED^{ACT}$, controller C1; Supporting Information Figure 1). Controller C2 generates negative actuation because ligand-binding represses aptazyme cleavage activity that would otherwise increase mRNA half-life ($aRED^{REP}$; Figure 2, in red; Supporting Information Figure 1). Controller C3 is a composite device for negative actuation where ligand-binding activates aptazyme cleavage, which programs an increase in the half-life of an sRNA engineered to inhibit translation from a target mRNA ($aRED^{ACT}$-sRNA; Figure 2, in blue). Controller C4 is a composite device for negative actuation where ligand-binding activates aptazyme cleavage, increasing the mRNA half-life of a transcriptional repressor ($aRED^{ACT}$-TR; Figure 2, in orange).

We determined architecture performance by creating 728 unique coarse-grained biochemical models (Figure 2), corresponding to each possible dynamic control architecture (see Methods). Each of these 728 models was evaluated 2000 times with Latin-hypercube-sampled[24] parameter values constrained to biochemically plausible ranges (Supporting Information Table 1). Each of these 2000 parameter sets is referred to as an instance of that architecture. Models were formulated to emphasize measurable and tunable kinetic design parameters. Sampling parameter values thus allows exploration of the design space, in the case of tunable parameters, and additionally accounts for parameter uncertainty within the cell in the case of nontunable parameters. For each architecture instance, i.e., parameter set, the yield on glucose was computed (see Methods). Yields were then normalized to the median yield for static controllers to arrive at relative yield ($Y_{rel}$). Yield comparisons are thus relative to static control, which contains no dynamic regulation, though parameters still vary by sampling.

To broadly ascertain how dynamic control can influence production, distributions of relative yields for all of the parameter sets were analyzed. The distribution of $Y_{rel}$ for the full set of dynamic control architecture instances shows a wide range of production levels relative to static control levels, a large portion of which is worse than static control (Figure 3). However, 137 out of the 728 (~19%) architectures have median $Y_{rel}$ greater than one, and expected *p*-AS production


dx.doi.org/10.1021/sb400201u | *ACS Synth. Biol.* 2015, 4, 107−115

from the top 5% of architectures is almost three times higher than that of comparable systems with only static control. While a subset of the static control parameter sets can produce results comparable to the best dynamic controllers, as evidenced by comparing the largest $Y_{rel}$ values for dynamic and static control (Supporting Information Figure 3), the distributions of highly performing controllers are shifted toward higher $Y_{rel}$. Taken together, these results show that dynamic control can greatly increase the likelihood of obtaining high *p*-AS production yields across the space of system architectures that we studied.

We then investigated whether dynamic control system performance can be understood at the level of core topology or the set of actuator functions combined across the three nodes. All 728 dynamic control architectures were grouped according to the 26 distinct core topologies (Supporting Information Figure 4), and we then evaluated the performance of the control architectures as a set. For example, Core Topology 1 (CT 1) consists of the 12 architectures corresponding to all the possible combinations of sensor targets and controllers that can lead to positive actuation at Node 1, negative actuation at Node 2, and no actuation at Node 3. Predicted *p*-AS production varies dramatically between core topologies, as indicated by median $Y_{rel}$ values that span more than 2 orders of magnitude (Figure 4). Notably, several of

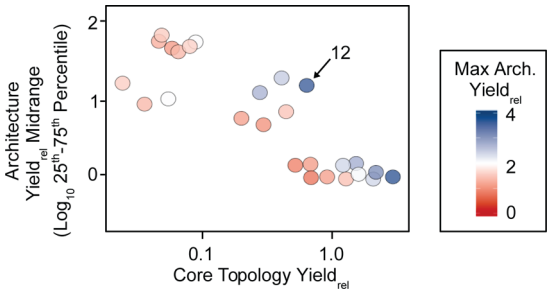



**Figure 4.** Performance of core topologies. Each circle corresponds to all of the architectures within a given core topology. Placement on the horizontal axis indicates the median relative yield ($Y_{rel}$) of all instances of the core topology. Placement on the vertical axis indicates the $\log_{10}$-difference of the 25th to 75th percentile range of architecture $Y_{rel}$ (i.e., the interquartile range). Color denotes the 95th percentile $Y_{rel}$ for architectures within the core topology. Core Topology 12 is identified by call-out (see text).

the core topologies have both high $Y_{rel}$ and only small variations in the midrange, or first to third quartile, of the $Y_{rel}$ values, indicating that most of the architectures in those core topologies enable high levels of production (Figure 4, blue points at bottom right). Combined, these results are consistent with the idea that there are regions of the dynamic control design space defined by the core topologies that reliably outperform systems with only static controllers.

Variation in performance among the architectures within a core topology can be almost as large as the variation between core topologies, where nearly 100-fold differences in $Y_{rel}$ at the midrange are observed in some cases (Figure 4, upper points). For example, Architecture 270 (Arch 270) within CT 12 (Figure 4, labeled) has the second-highest median $Y_{rel}$ (3.6) out of 728 architectures examined, even though the median relative yield for CT 12 itself is only 0.7. Thus, although classifying control system architectures according to core topology can help identify high- and low-performing system designs, core topology alone does not fully determine production yields.

We can begin to understand how variation in expected yields within a core topology arises by examining architecture-level details. For instance, there are 24 different architectures in CT 12, all with aRED-mediated positive actuation at Nodes 1 and 3, and negative actuation at Node 2. However, $Y_{rel}$ for the best CT 12 architecture (270) is 20-fold higher than for the worst architecture (266) (Figure 5). Arch 270 and Arch 266 employ different Node 2 actuation mechanisms and have very different sets of sensing metabolite−actuator linkages. As a result, significant differences in *p*-AS production across the sampled parameter spaces are observed (Mann−Whitney, $W = 3.7 \times 10^5$, $P < 2.2 \times 10^{-16}$). One-way ANOVA shows that the type of controller employed for negative actuation at Node 2 has a significant impact on *p*-AS production within CT 12 ($F = 435.4$, $P < 2.2 \times 10^{-16}$). For CT 12 architectures with C2 controllers (aRED$^{REP}$), more than 22% (3611 out of 16 000) of the simulated parameter instances gave high-performing production yields better than 95% of the static control system instances. In contrast, for CT 12 architectures with controllers C3 (aRED$^{ACT}$-sRNA) and C4 (aRED$^{ACT}$-TR) at Node 2, only 4.1% and 0.42% of the parameter instances, respectively, resulted in predicted *p*-AS levels at the 95th percentile of static control system production.

Interestingly, within the CT 12 architectures, Node 2 controllers with high levels of repression appear to have deleterious effects on production (Supporting Information Figure 5). For C3 controllers, there is a negative correlation (Spearman rank correlation coefficient, $R_s = -0.38$, $P < 2.2 \times 10^{-16}$) between *p*-AS production and the sRNA-target mRNA association rate, the latter being a design variable directly increasing translation repression in the model. Similarly, for C4 controllers, tighter TR-DNA binding results in less *p*-AS production ($R_s = -0.71$, $P < 2.2 \times 10^{-16}$) across a range of TR $k_d$s similar to those reported for the well-known Lac repressor (LacI-DNA $k_d \approx 10^{-7}$ to $10^{-9}$ M[25]). Counterintuitively, these data show that actuators with strong repression characteristics are not always well-suited to the problem of implementing negative feedback in this system. Instead, the choice of controller may be important for creating fine-grained modulation that can mitigate regulatory problems while maintaining high levels of flux through the pathway.

To confirm that the observed differences in predicted *p*-AS production are caused by the dynamic control layer, we simulated production for the 24 CT 12 architectures with deactivated aREDs. The design variable for aptazyme folding, needed for ligand-responsive cleavage, was set to zero, and $Y_{rel}$ was recomputed for all the CT 12 architectures with the same 2000 parameter instances as before (Supporting Information Figure 6). Over the entire design space, $Y_{rel}$ values obtained with deactivated aptazymes are lower than for actuators with functional controllers, indicating that dynamic, ligand-responsive control is needed for the highest performance (one-way ANOVA, $F = 1914$, $P < 2.2 \times 10^{-16}$). Similarly, for the best-performing CT 12 architecture (Arch 270), the predicted *p*-AS production with functional aptazymes is significantly better than when the controllers are disabled (Mann−Whitney, $W = 2.9 \times 10^6$, $P < 2.2 \times 10^{-16}$). These results suggest that aRED-mediated dynamic control is responsible for the improvements in production obtained relative to systems comprising only static controllers.

Because aRED function is central to the connection between dynamic sensing and actuation, we examined the feasibility of engineering aptazymes that can meet component specifications


dx.doi.org/10.1021/sb400201u | *ACS Synth. Biol.* 2015, 4, 107−115

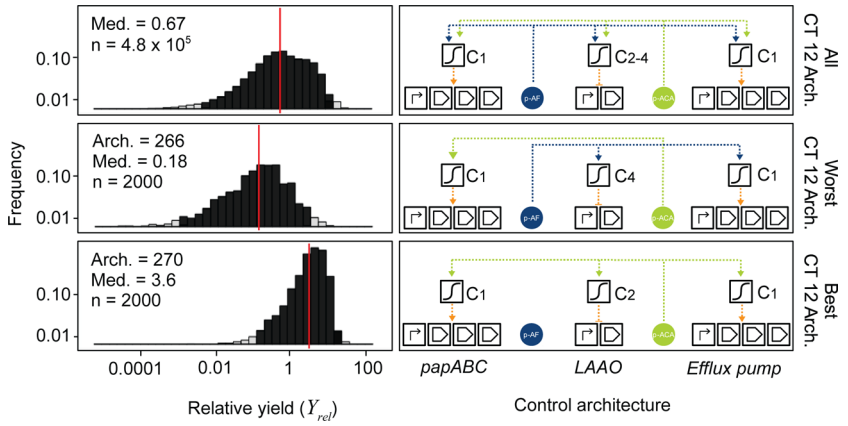



**Figure 5.** Relative production yields for Core Topology 12 (CT 12). CT 12 comprises the entire set of architectures that effect positive actuation on Nodes 1 (PapABC operon) and 3 (efflux pump), and negative actuation on Node 2 (LAAO). Shown are $\log_{10}$ relative yields ($Y_{rel}$) alongside a schematic for all CT 12 instances (top), all instances for the CT 12 architecture (266) with the lowest median $Y_{rel}$ (middle), and all instances for the CT 12 architecture (270) with the highest median $Y_{rel}$ (bottom). The vertical red line denotes median $Y_{rel}$ within a set. Black shading indicates the $Y_{rel}$ 95% confidence interval.

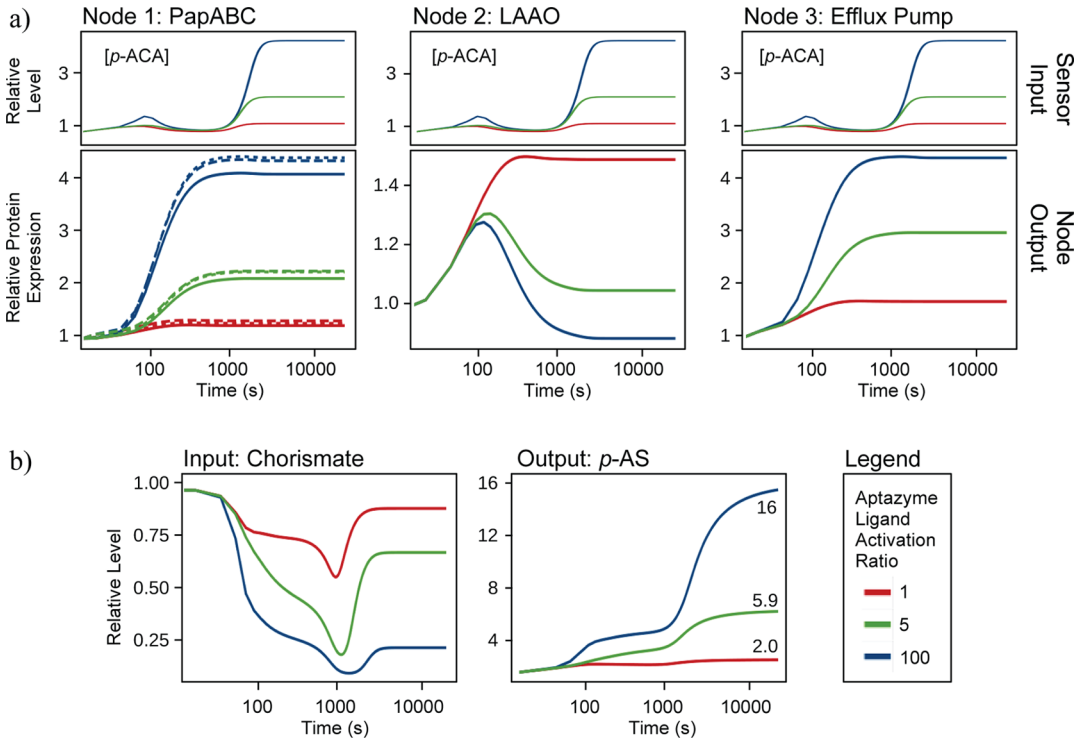



**Figure 6.** Dependence of relationship between sensor input and node output on aptazyme ligand-activation ratio. (A) The predicted level of sensor input (*p*-ACA) and outputs (i.e., folded proteins) for each of the three nodes in the CT 12 Arch 270 system are plotted, as a function of time, relative to parameter-matched systems of static controllers. (B) The relative levels of initial input (chorismate) and final output (*p*-AS) are plotted as a function of time. Colors indicate the modeled aptazyme ligand-activation ratio or the cleavage rate in the presence versus the absence of ligand ($k_{obs+}/k_{obs-}$) specified for the aRED controllers in the system.

needed for high-performing systems. As previously described,[10] aptazyme ligand-activation ratio (LAR), or cleavage rate in the presence versus absence of ligand (i.e., $k_{obs+}/k_{obs-}$), has a significant impact on aRED genetic output. In fact, differences in LAR and transcript half-life have been shown to account for as much as 85% of the variation in relative ligand-responsive expression.[10] *p*-AF-responsive aptazymes with LARs of ∼5−6 have been successfully engineered,[10] and aptazymes responsive to other molecules with LARs > 100 have been characterized.[26] New methods that combine kinetic RNA folding design, *in vitro* selection,[27,28] and *in vivo* screening[29] should enable the rapid generation of *p*-ACA aptazymes with similar LARs.

Notably, for aRED controllers with aptazyme LAR = 5 (see Methods), the predicted *p*-AS production from the best-performing CT 12 architecture (270) is almost 6 times higher than for static controllers and 3 times higher than for controllers without ligand-responsive actuation (i.e., LAR = 1). When LAR = 100, predicted *p*-AS yields are 16 times greater than for systems with static control (Figure 6; Supporting Information Figure 7). As expected,[30] positive feedback results in faster rise times to steady-state for genes


dx.doi.org/10.1021/sb400201u | *ACS Synth. Biol.* 2015, 4, 107−115

expressed from Nodes 1 (pap operon) and 3 (efflux pump operon), with the slopes of the rise increasing 4−8 times as the LAR increases from 1 to 5 to 100. Negative actuation at Node 2 results in the downregulation of LAAO expression as a function of time, with the degree of repression increasing along with aptazyme LAR. Overall, we can see that the aRED-mediated controllers program outputs from each node in response to changing sensor inputs (*p*-ACA concentrations), with the net effect of lowering steady-state chorismate concentrations and increasing *p*-AS production. The fact that even aptazymes with modest LARs can enable substantial increases in *p*-AS production highlights the potential for engineering aRED-based systems to dramatically improve metabolic pathway function.

In brief summary, we combined coarse-grained mechanistic modeling with statistical analysis to map the combinatorial design space of RNA-based genetic control system inputs to the space of production outputs in an engineered *p*-AS production pathway. We found that core topology, the set of genetic actuator functions, alone may not be enough to specify a successful dynamic control system; architecture-level details can determine if a core topology performs well. Furthermore, for a small subset of all possible core topologies, there are broad ranges of design parameters and underlying regulatory mechanisms, including composite aRED-based devices, that could be employed to solve pathway control problems and enable high levels of production. Interestingly, our results suggest that, for the systems under study, stronger negative regulation is not always better, and controllers capable of finer modulations of gene expression may be important for designing and constructing metabolic production systems such as the one in this study. More generally, our results show that our approach can be used to develop dynamic control systems that couple the expression of production enzymes, toxic intermediates, and efflux pumps to keep toxicities at manageable levels and optimize production yields. Building from the computational approach in this study, a larger body of work in metabolic control system design and engineering should help further elucidate general strategies for increasing production.

While the present genetic control design space analysis was conducted on a pathway to produce substituted styrenes, the sampling-based simulation approach can be applied to any metabolic control problem. The modular nature of the modeling approach used in this study is easily extensible for incorporating additional regulatory controllers. Alternative actuation implementation mechanisms, such as riboswitches[13] or CRISPR-Cas9-based systems,[31] could be easily analyzed to further explore how actuator characteristics can impact pathway performance. For a new mechanism to fit within this approach, the underlying biochemical mechanisms should be understood in enough detail to formulate models in terms of measurable and tunable kinetic design parameters. Taking CRISPR-Cas9-based systems as an example, the mechanism of action is well understood, but at a minimum, coarse-grained kinetic details regarding Cas9-mediated-DNA binding and possibly Cas9-RNA binding will be needed to integrate such a system into this framework.

The approach described here can be scaled to study more complex pathway control topologies found in nature, such as those involving an explicit temporal component. For such systems, analytical techniques to perform statistical analysis on time-resolved data should lend insight into the role that temporal evolution can play in control systems. It will be interesting to gain further understanding of how naturally occurring control architectures can be used for metabolic engineering and perhaps when they fall short and new strategies are needed. Organisms have evolved metabolic control under pressures for survival, and the evolved strategies may not be appropriate for engineering synthetic circuits for high levels of production.

Practically speaking, computational genetic control system design and screening can be a time-, energy-, and cost-efficient way to drive experimental efforts.[32] Using the design specifications arising from this work as a starting point, we can now begin to experimentally address fundamental questions about the range of dynamic, RNA-based systems that can be constructed to solve control problems in engineered pathways.

## METHODS

**Modeling.** *p*-AS production systems were modeled as sets of deterministic ordinary differential equations (ODEs). All 728 architectures were automatically generated as PySCeS,[33] v. 0.9.0, model files using custom Python scripts to write models in the PySCeS modeling language. Models were composed of mechanistic differential equations described in Figure 2 and Supporting Information Table 1. Incorporating the control problems into the model was achieved by adding a first-order loss reaction for *p*-AF and introducing a toxicity term into the model that placed a general fitness disadvantage on cells by slowing all nondegradation and dilution reaction rates.[5] The value of this toxicity term (see Supporting Information Table 1) was dynamically dependent on the concentration of *p*-ACA, LAAO, and efflux pump. Toxicity coefficient ranges were modeled so that *p*-ACA slows growth by half at ∼10 mM (Supporting Information Figure 2) and so that the efflux pump and LAAO slow growth when present in the thousands of molecules range.

Two full sets (one static and one dynamic) of model reactions for the best Core Topology 12 architecture are in the Supporting Information, along with four parameter sets for instances with relative yield performance ranked first, 50th, 1000th, and 1950th. A data chart with final relative yield values and statistics is also provided.

For time-course simulations (Figure 6 and Supporting Information Figure 7), data were collected from simulations of the architecture (270) of Core Topology 12 with the highest median $Y_{rel}$, using the parameter set from the best instance (highest $Y_{rel}$) of that architecture (see Figure 5). All parameters except the three aRED LARs were kept constant, and the aptazyme ligand-activation ratios (LARs) were all set to 1, 5, and 100 and the models re-evaluated.

**Simulation.** Models were simulated on the University of Washington Hyak Computing Cluster using custom software for setup and analysis. Models were simulated to 48 h, the length of a typical batch reactor run, using the LSODA ODE[34] solver packaged with PySCeS, v. 0.9.0.

**Yield Calculations.** Yield ($Y$) on glucose for an architecture instance was calculated by summing internal (int) and external (ext) *p*-ACA molecule counts ($N_{p\text{-ACA,int}}$ and $N_{p\text{-ACA,ext}}$) and dividing that by the total number of chorismate molecules ($N_{chor}$) produced: $Y = (N_{p\text{-ACA,int}} + N_{p\text{-ACA,ext}})/N_{chor}$. Relative yield ($Y_{rel}$), for static or dynamic cases, is $Y$ divided by the median relative yield value of all static controllers: $Y_{rel} = Y/\text{median}(Y_{static})$. Note that, for the purposes of this study, *p*-AS is used interchangeably with *p*-ACA, as the conversion is extracellular.

**Parameter Sampling.** Parameters were sampled using the Latin hypercube sampling[24] method from uniform distributions centered around nominal values gathered from the literature. Latin hypercube sampling aims to approximate an output distribution with fewer inputs by dividing up input space and sampling within the divided regions. This makes it less likely that large regions of input space will go unsampled. Ranges for the sampling and sources for nominal values are given in Supporting Information Table 1. The sampled parameter sets were generated using the lhs method in the lhs package for R.

**Statistical Analysis.** Mann−Whitney testing was performed in R using the wilcox.test method. The Mann−Whitney test is a nonparametric statistical test with the null hypothesis that the two populations being tested are from the same distribution. All tests had a one-sided alternative hypothesis of one population being greater than the other. Two-sided Spearman rank correlations were performed in R using the cor.test method. Spearman rank correlation is a nonparametric statistical test that assesses the monotonicity of the relationship between two variables. One-way ANOVA was performed in R using the lm and anova methods. ANOVA is a statistical test that compares the differences between group means to determine the likelihood of data originating from the same distribution. The null hypothesis is that the data are from the same distribution.

**Experimental Measurements of Toxicity.** An overnight culture of *E. coli* MG1655 was inoculated 1:100 into fresh MOPS EZ (Teknova) media and allowed to reach an OD600 of 0.1. Two hundred microliters of culture was then placed into wells of a black 96-well plate with a clear bottom (Corning), and amounts of inducer or *p*-ACA (in DMSO) were added to cultures. Final DMSO concentration in media (v/v) was kept at or below 2% to avoid DMSO toxicity. The plate was then placed into a Tecan M1000 plate reader with orbital shaking at 37 C for 12 h. OD600 measurements were taken every 10 min.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

Supporting table, figures, references, model files, and simulation data. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: jcaroth@uw.edu. Phone: 206-221-4902.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank W. M. Voje, Jr., R. C. Correa, D. Sparkman-Yager, A. L. Smith, and J. B. Matsen for helpful discussions and comments. This work was supported by the Molecular Engineering & Sciences Institute and the University of Washington. J.M.C. is a fellow of the Alfred P. Sloan Foundation. J.T.S. is supported in part by a National Science Foundation Graduate Research Fellowship.

## ■ REFERENCES

(1) Martínez-Antonio, A., Janga, S. C., Salgado, H., and Collado-Vides, J. (2006) Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. *Trends Microbiol. 14*, 22−7.

(2) Carothers, J. M., Goler, J. A., and Keasling, J. D. (2009) Chemical synthesis using synthetic biology. *Curr. Opin. Biotechnol. 20*, 498−503.

(3) Dunlop, M. J. (2011) Engineering microbes for tolerance to next-generation biofuels. *Biotechnol. Biofuels 4*, 32.

(4) Farmer, W. R., and Liao, J. C. (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol. 18*, 533−7.

(5) Zhang, F., Carothers, J. M., and Keasling, J. D. (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol. 30*, 354−9.

(6) Dunlop, M. J., Dossani, Z. Y., Szmidt, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., Hadi, M. Z., and Mukhopadhyay, A. (2011) Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol. 7*, 487.

(7) Harrison, M. E., and Dunlop, M. J. (2012) Synthetic feedback loop model for increasing microbial biofuel production using a biosensor. *Front. Microbiol. 3*, 360.

(8) Stevens, J. T.; Carothers, J. M. Programming gene expression by engineering transcript stability control and processing in bacteria. *Synth. Biol.* 2014, in press.

(9) Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature 346*, 818−22.

(10) Carothers, J. M., Goler, J. A., Juminaga, D., and Keasling, J. D. (2011) Model-driven engineering of RNA devices to quantitatively program gene expression. *Science 334*, 1716−9.

(11) Chen, Y. Y., Jensen, M. C., and Smolke, C. D. (2010) Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc. Natl. Acad. Sci. U.S.A. 107*, 8531−6.

(12) Michener, J. K., and Smolke, C. D. (2012) High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab. Eng. 14*, 306−16.

(13) Win, M. N., and Smolke, C. D. (2008) Higher-order cellular information processing with synthetic RNA devices. *Science 322*, 456−60.

(14) Yang, J., Seo, S. W., Jang, S., Shin, S.-I., Lim, C. H., Roh, T.-Y., and Jung, G. Y. (2013) Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. *Nat. Commun. 4*, 1413.

(15) Goikhman, M. Y., Yevlampieva, N. P., Kamanina, N. V., Podeshvo, I. V., Gofman, I. V., Mil'tsov, S. A., Khurchak, A. P., and Yakimanskii, A. V. (2011) New polyamides with main-chain cyanine chromophores. *Polym. Sci., Ser. A 53*, 457−468.

(16) Qi, W. W., Vannelli, T., Breinig, S., Ben-Bassat, A., Gatenby, A. A., Haynie, S. L., and Sariaslani, F. S. (2007) Functional expression of prokaryotic and eukaryotic genes in *Escherichia coli* for conversion of glucose to p-hydroxystyrene. *Metab. Eng. 9*, 268−76.

(17) Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010) What is flux balance analysis? *Nat. Biotechnol. 28*, 245−8.

(18) Schuster, S. (1999) Use and limitations of modular metabolic control analysis in medicine and biotechnology. *Metab. Eng. 1*, 232−42.

(19) Chiang, A. W. T., and Hwang, M.-J. (2013) A computational pipeline for identifying kinetic motifs to aid in the design and improvement of synthetic gene circuits. *BMC Bioinf. 14* (Suppl 1), S5.

(20) Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res. 39*, 1131−41.

(21) Seo, S. W., Yang, J.-S., Cho, H.-S., Yang, J., Kim, S. C., Park, J. M., Kim, S., and Jung, G. Y. (2014) Predictive combinatorial design of mRNA translation initiation regions for systematic optimization of gene expression levels. *Sci. Rep. 4*, 4515.

(22) Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., and Lee, S. Y. (2013) Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat. Biotechnol. 31*, 170−4.

(23) Elowitz, M. B., and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature 403*, 335−8.

(24) Mckay, M. D., and Beckman, R. J. (1979) Three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 21*, 239−245.

(25) Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005) Transcriptional

regulation by the numbers: applications. *Curr. Opin. Genet. Dev. 15*, 125−35.
(26) Link, K. H., Guo, L., Ames, T. D., Yen, L., Mulligan, R. C., and Breaker, R. R. (2007) Engineering high-speed allosteric hammerhead ribozymes. *Biol. Chem. 388*, 779−86.
(27) Thimmaiah, T., Voje, W. E., and Carothers, J. M. (2014) Computational Design of RNA Parts, Devices, and Transcripts with Kinetic Folding Algorithms Implemented on Multiprocessor Clusters, in *Computational Methods in Synthetic Biology*, Springer, New York.
(28) Sparkman-Yager, D.; Correa-Rojas, R. A.; Carothers, J. M. Kinetic folding design of aptazyme-regulated expression devices (aREDs) as riboswitches for metabolic engineering. *Methods Enzymol.* 2014, in press.
(29) Goler, J. A., Carothers, J. M., and Keasling, J. D. (2014) Dual-selection for evolution of in vivo functional aptazymes as riboswitch parts. *Methods Mol. Biol. 1111*, 221−35.
(30) Alon, U. (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet. 8*, 450−61.
(31) Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell 152*, 1173−83.
(32) Carothers, J. M. (2013) Design-driven, multi-use research agendas to enable applied synthetic biology for global health. *Syst. Synth. Biol. 7*, 79−86.
(33) Olivier, B. G., Rohwer, J. M., and Hofmeyr, J.-H. S. (2005) Modelling cellular systems with PySCeS. *Bioinformatics 21*, 560−1.
(34) Hindmarsh, A. (1983) ODEPACK, A Systematized Collection of ODE Solvers, in *IMACS Transactions on Scientific Computation* (Stepleman, R. S., Ed.) Vol. *1*, pp 55−64, North-Holland, Amsterdam, The Netherlands.
(35) Juminaga, D., Baidoo, E. E. K., Redding-Johanson, A. M., Batth, T. S., Burd, H., Mukhopadhyay, A., Petzold, C. J., and Keasling, J. D. (2012) Modular engineering of L-tyrosine production in *Escherichia coli*. *Appl. Environ. Microbiol. 78*, 89−98.
(36) Mehl, R., Anderson, J. C., Santoro, S. W., Wang, L., Martin, A. B., King, D. S., Horn, D. M., and Schultz, P. G. (2003) Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc. 125*, 935−9.
(37) Celesnik, H., Deana, A., and Belasco, J. G. (2007) Initiation of RNA decay in *Escherichia coli* by 5′ pyrophosphate removal. *Mol. Cell 27*, 79−90.
(38) Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005) Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev. 15*, 116−24.
(39) De Lay, N., Schu, D. J., and Gottesman, S. (2013) Bacterial small RNA-based negative regulation: Hfq and its accomplices. *J. Biol. Chem. 288*, 7996−8003.