Combinatorial promoter design for engineering noisy gene expression

Kevin F. Murphy[†‡], Gábor Balázsi[†§], and James J. Collins[†¶]

[†]Department of Biomedical Engineering, Center for BioDynamics and Center for Advanced Biotechnology, and [‡]Department of Biology, Boston University, Boston, MA 02215

Understanding the behavior of basic biomolecular components as parts of larger systems is one of the goals of the developing field of synthetic biology. A multidisciplinary approach, involving mathematical and computational modeling in parallel with experimentation, is often crucial for gaining such insights and improving the efficiency of artificial gene network design. Here we used such an approach and developed a combinatorial promoter design strategy to characterize how the position and multiplicity of $tetO_2$ operator sites within the $GAL1$ promoter affect gene expression levels and gene expression noise in *Saccharomyces cerevisiae*. We observed stronger transcriptional repression and higher gene expression noise as a single operator site was moved closer to the TATA box, whereas for multiple operator-containing promoters, we found that the position and number of operator sites together determined the dose–response curve and gene expression noise. We developed a generic computational model that captured the experimentally observed differences for each of the promoters, and more detailed models to successively predict the behavior of multiple operator-containing promoters from single operator-containing promoters. Our results suggest that the independent binding of single repressors is not sufficient to explain the more complex behavior of the multiple operator-containing promoters. Taken together, our findings highlight the importance of joint experimental–computational efforts and some of the challenges of using a bottom-up approach based on well characterized, isolated biomolecular components for predicting the behavior of complex, synthetic gene networks, e.g., the whole can be different from the sum of its parts.

combinatorial design | mathematical modeling | promoter engineering | stochastic gene expression | synthetic biology

Designing and constructing novel biomolecular systems is a fundamental goal of synthetic biology (1–21), which is often challenging due to the inherent complexity of biological systems. In contrast to electronics, where most components are relatively simple and well characterized, allowing for reliable circuit design through integration, only a limited number of biological "parts" are known in sufficient detail to allow for predictable behavior. Even well studied, apparently simple biological systems can exhibit surprisingly complex, context-dependent behavior when they interact with each other. Therefore, it is crucial to characterize the behavior of proteins, genes, promoters, and operator sites not simply as isolated components, but also when they are brought together as parts of a larger system.

Many promoters contain regulatory elements for multiple transcription factors, and are responsible for biological computation and signal integration through gene regulation (22–29). However, the combination of regulatory sites in a promoter region can result in behavior that is not predictable from studying the individual sites alone (27, 30). Therefore, to more accurately use these natural regulatory components in synthetic networks, it is crucial to understand how the combination and multiplicity of regulatory sites affect gene expression.

Gene expression noise, which can be promoter-dependent (8, 31–35), can have important effects on survival, differentiation, and information processing (36–44). As a consequence, it is important for synthetic biologists to study the effect of stochastic gene expression in engineered gene networks. Several studies have focused on noise propagation (8, 17, 19) and the effect of feedback on noise in gene circuits (3, 4, 10, 45, 46). Others have shown that gene expression noise is influenced by diverse biological factors and processes, including transitions between active and inactive promoter states (8, 31, 47), transcription and translation (5, 6, 8, 31, 32, 48–50), cell division (20, 32), and general regulatory events such as environment-induced signaling and chromatin remodeling (6, 31, 51). Individual promoter components such as the TATA box have also been examined in terms of their influence on gene expression noise (8, 31, 43). However, the way in which the number and configuration of operators within a single promoter affect gene expression noise is still unknown.

Here, we study the effect of operator positions within the $GAL1$ promoter from *Saccharomyces cerevisiae*, by building a combinatorial set of seven synthetic promoters containing one, two, and three $tetO_2$ operator sites. We develop a generic computational model to describe how dose–response curves and gene expression noise depend on the location of the operator within the promoter, and discuss how the description of single operator-containing promoters can be used to characterize the double and triple operator-containing promoters. Our results suggest that the independent binding of individual repressors is not sufficient to explain the more complex behavior of the multiple operator-containing promoters. The predictability of multiple operator-containing promoters decreases with the number of inserted operator sites, which suggests that bottom-up approaches, based on well characterized, isolated components, may not always be useful for predicting the behavior of complex, synthetic gene networks.

Results

Combinatorial Promoter Design. We built a combinatorial set of seven synthetic promoters to investigate the effects of various operator site combinations on different gene expression output variables. A yeast integrative plasmid (8) served as the template for engineering the complete set of TetR-repressible $GAL1$ promoters used in this study (Fig. 1*A*). After chromosomal integration, TetR is constitutively expressed from the synthetic P_{GAL10} when grown in the presence of galactose, and represses
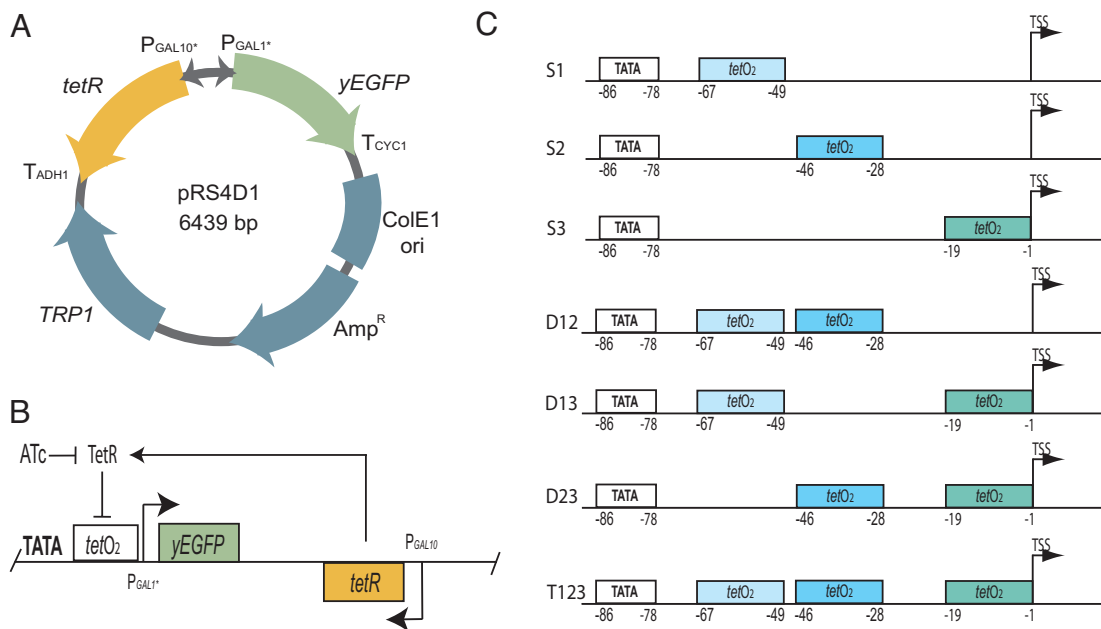
Fig. 1. Diagram of synthetic constructs. (*A*) Yeast integrative plasmid pRS4D1 contains the bacterial *Col*E1 origin of replication and ampicilin resistance gene as indicated. The *TRP1* gene allows for selection in yeast. The *tetR* gene is under the control of the P_{GAL10} promoter, whereas *yEGFP* reporter gene expression is under the control of P_{GAL1*}. Transcriptional terminators (T_{CYC1} and T_{ADH1}) are also indicated. (*B*) Schematic depicting integrated P_{GAL1*} transcriptional control. The *tetR* gene is transcribed constitutively from P_{GAL10} in galactose-containing media. The TetR repressor protein binds inserted *tet*O_2 operator(s) downstream of the P_{GAL1*} TATA box and inhibits transcription of *yEGFP*. Addition of anhydrotetracycline inhibits TetR binding of operator(s), allowing transcription from P_{GAL1*}. (*C*) Diagram of P_{GAL1*} promoter constructs containing all seven *tet*O_2 operator combinations. The TATA box and *tet*O_2 operator locations are indicated by base position number relative to transcription start site (TSS). The name of each promoter is indicated to its left in the diagram. Here, single, double, and triple operator-containing promoters are designated by the letters S, D, and T, respectively. The numbers 1, 2, and 3 following these letters indicate the inclusion of the corresponding operator site.

expression of yeast-enhanced green fluorescent protein (*yEGFP*) from P_{GAL1*} through its binding of *tet*O_2 operator sites inserted downstream from the P_{GAL1*} TATA box (Fig. 1*B*). This repression can be relieved by the addition of anhydrotetracycline (ATc) to the growth medium. Upon the binding of two ATc molecules, each TetR dimer undergoes a conformational change (52), which prevents free dimers from binding operator sites and causes the release of operator-bound dimers. Subsequent induction of P_{GAL1*} can then be measured by *yEGFP* reporter expression.

As a basis for combinatorial promoter design, we first constructed a set of three promoters (S1, S2, and S3), each containing a single operator inserted at a different position between the TATA box and transcription start site (Fig. 1*C*). Next, we designed and constructed a set of double operator-containing promoters (D12, D13, and D23), combining the operator of S1 with that of S2, S1 with S3, and S2 with S3, respectively (Fig. 1*C*). Finally, we designed and constructed a triple operator-containing promoter (T123), combining all operators of S1, S2, and S3. The letter in each promoter name indicates the number of operator sites (S, single; D, double; T, triple), whereas the numbers 1, 2, and 3 indicate their positions as in promoters S1, S2, and S3, respectively. The numbers also reflect the distance between the operator site and TATA box, the operator in S1 being the closest and the operator in S3 being the farthest from the TATA box (Fig. 1*C*). We used the wild-type (WT) *GAL1* promoter as a control.

Repressor Binding Site Location Affects the Dose–Response Curves. We first compared the basal expression levels (P_{min}) of the single operator-containing promoters S1, S2, and S3. We noted an increase in basal expression levels ($P_{min} = 21.2 \pm 0.5$, 50.8 ± 2.0, and 637.6 ± 22.7, respectively, Fig. 2 *B–D*) as the operator site

was moved farther downstream from the TATA box toward the transcription start site. We observed a similar dependence on operator site location for the double operator-containing promoters D12, D13, and D23 ($P_{min} = 6.3 \pm 0.01$, 18.2 ± 1.3, and 76.8 ± 2.4, respectively, Fig. 2 *E–G*). Along these lines, the triple operator-containing promoter T123 exhibited the lowest basal expression level ($P_{min} = 3.6 \pm 0.1$, Fig. 2*H*). These data indicate that increasing the number of operator sites and/or their proximity to the TATA box results in more efficient repression of the *GAL1* promoter.

We observed gene expression differences between the single operator-containing promoters S1, S2, and S3 ($P_{max} = 1,462 \pm 22$, 855 ± 9 and $1,694 \pm 33$, respectively, Fig. 2 *B–D*), as well as between the multiple operator-containing promoters D12, D13, D23, and T123 ($P_{max} = 1,039 \pm 64$, $1,565 \pm 124$, $1,235 \pm 68$, and $1,357 \pm 29$, respectively, Fig. 2 *E–H*) at full induction. To determine whether these differences were due to the replacement of native *GAL1* promoter sequences with the *tet*O_2 operators, we replaced the operator site in the S2 promoter (having the largest decrease in expression) with two random sequences. This caused even larger decreases in gene expression [see Fig. S7 in supporting information (SI) *Appendix*], indicating that the *GAL1* promoter sequence in this region plays a role in promoter activity. These results are consistent with earlier observations that *tet*O_2 operators can affect promoter activity in a position-dependent manner (53, 54). Additional controls involving a premature stop codon in the *tetR* coding sequence ruled out the possibility of the TetR protein somehow affecting the maximum expression levels. We used the expression levels of all seven promoters at full induction, together with the wild-type promoter, to model and study computationally the effect of the operator sites at the three different positions on preinitiation events and gene expression; see *SI Appendix* for a discussion of these analyses and results.
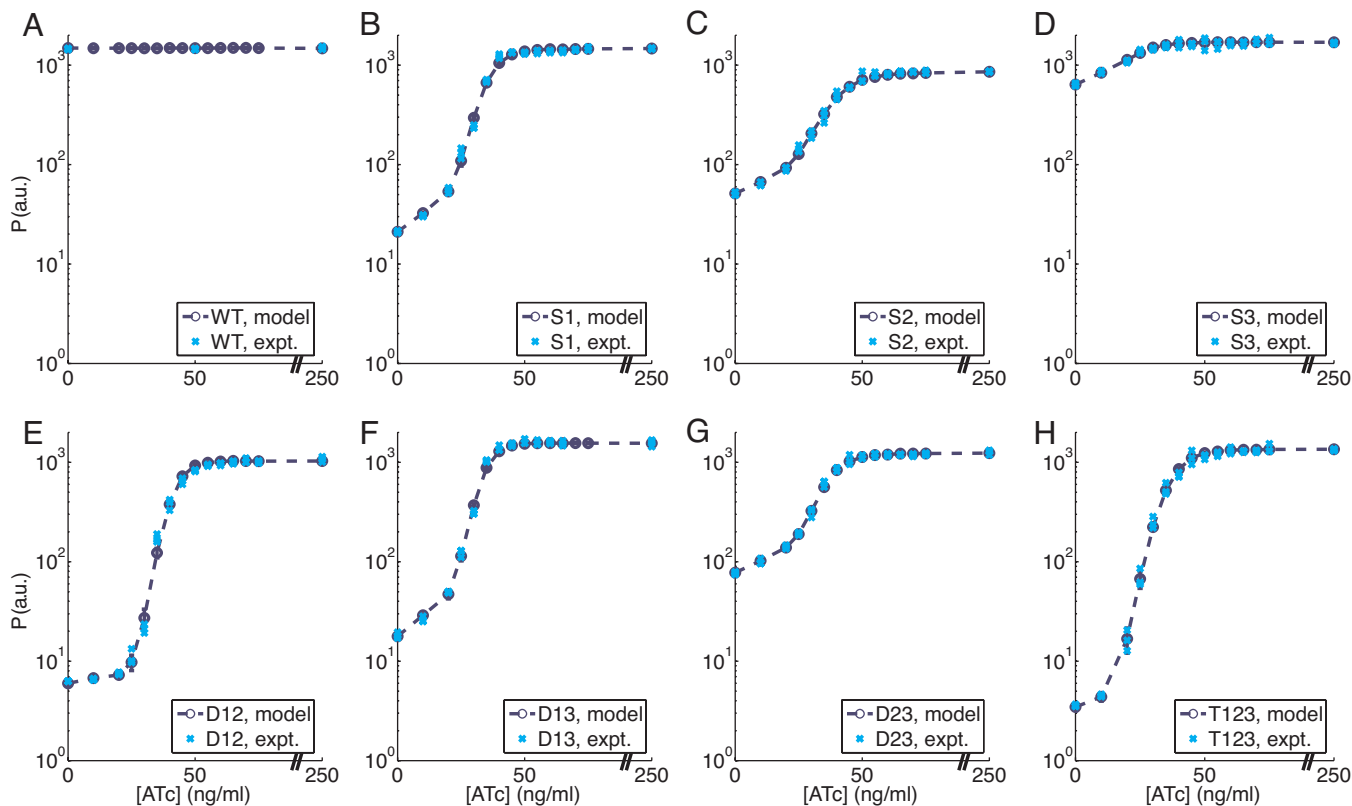
Fig. 2. Gene expression from the set of P$_{GAL1*}$ promoters. Experimental (light blue crosses) and simulated (dark blue circles) dose–response curves of the wild-type promoter (WT), single operator-containing promoters (S1, S2, and S3), double operator-containing promoters (D12, D13, and D23), and the triple operator-containing promoter (T123) are shown. The error bars indicate standard deviations from 10 different stochastic simulations.

Dose–response curves are widely used to characterize the input-output characteristics of biological systems, and are often approximated by the empirical Hill function

$$P(I) = P_{min} + (P_{max} - P_{min}) \frac{I^h}{H^h + I^h}. \quad \text{[1]}$$

The parameters P_{min} (basal response) and P_{max} (response at full induction) are usually determined by direct measurement, whereas H (the induction threshold) and h (the steepness of response or Hill coefficient) are estimated by fitting the Hill function to the experimental data (17, 55). We attempted to characterize the steady-state response of our seven engineered promoters by this methodology, but found that empirical Hill functions are insufficient to describe our experimental dose–response curves, which seem to be less steep at low levels of induction compared with high levels (see Fig. 2 and *SI Appendix*).

Therefore, to model gene expression from the seven promoters, we developed a chemical reaction scheme that included transitions between three promoter states [repressed (R), neutral (N), and active (A)], as well as mRNA (M) and protein (P) synthesis and degradation (Fig. 3). The promoter states were defined based on TetR and TATA-binding protein (TBP) occupancy, corresponding to one or more TetR dimers bound (R), neither TetR nor TBP bound (N), and TBP bound (A). We included TBP occupancy to simulate transcriptional reinitiation, which involves successive rounds of mRNA production upon a stably bound, TBP-anchored intermediate preinitiation complex (56, 57). In our model, mRNA can be synthesized either from promoter state A through transcription, or from promoter state R through promoter leakage (Fig. 3). Based on this chemical reaction scheme, we calculated a theoretical dose–response

function, and characterized the data by three parameters (v, n, and L) in addition to P_{min} and P_{max}

$$P(I) = \frac{P^{max}(vI)^n(I^2 + I + 1) + P^{min}(1 + LI)}{(vI)^n(I^2 + I + 1) + 1}. \quad \text{[2]}$$

The parameter L accounts for the steepness of the dose–response curve at low induction, whereas v and n determine the induction threshold and steepness of the dose–response curve at high induction. The function (Eq. **2**) is more suitable to fit both the single and multiple operator-containing promoters than the empirical Hill function, because it accounts for inducer-dependent promoter leakage from the repressed state, which
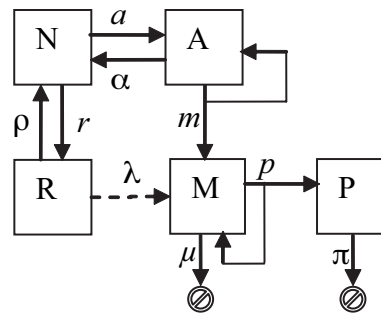

Fig. 3. Reaction scheme used for modeling the set of P$_{GAL1*}$ promoters. The letters R, N, and A indicate the repressed (TetR bound), neutral (neither TetR nor TBP bound), and active (TBP bound) promoter states, respectively, based on TetR/TBP binding. The letters M and P indicate mRNA and protein, respectively.
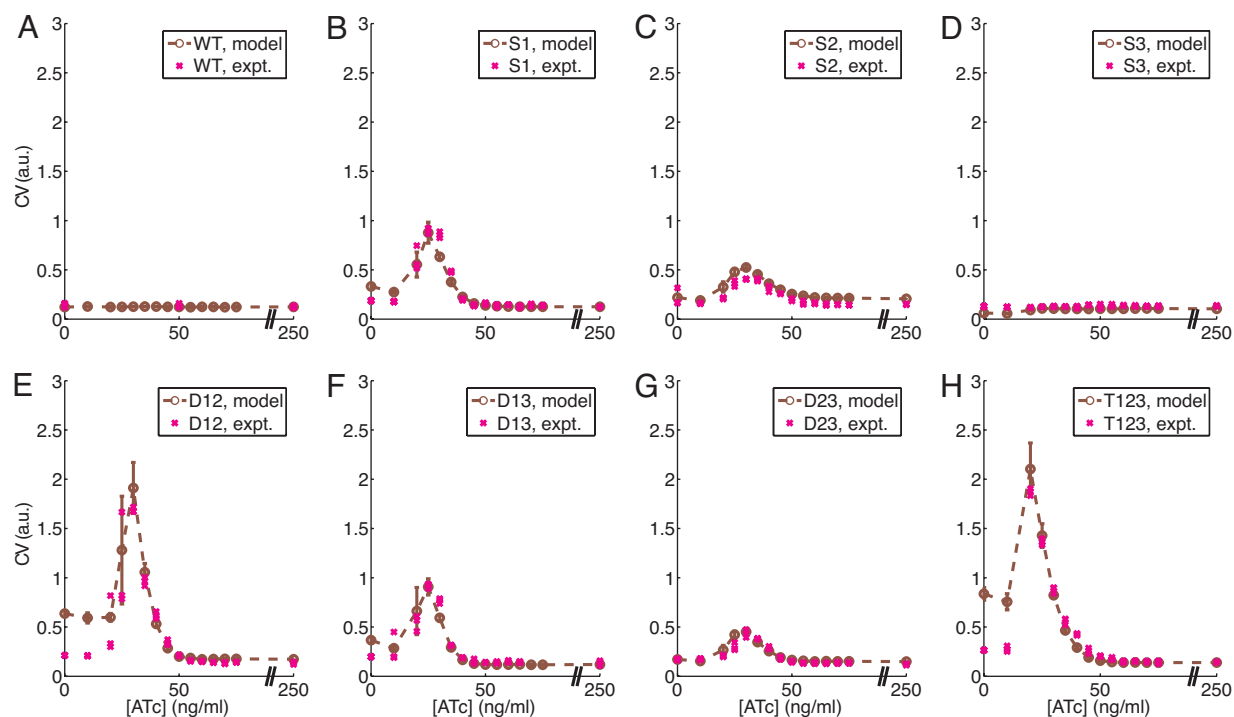
Fig. 4. Gene expression noise from the set of P$_{GAL1*}$ promoters. Experimental (magenta crosses) and simulated (dark red circles) coefficients of variation of the wild-type promoter (WT), single operator-containing promoters (S1, S2, and S3), double operator-containing promoters (D12, D13, and D23), and the triple operator-containing promoter (T123). The error bars indicate standard deviations from 10 different stochastic simulations.

causes a decrease in the steepness of the dose–response curves for low levels of induction, as was observed experimentally.

We studied how the parameters v, n, and L change with the position and multiplicity of operator sites in the various promoters. We found that n drops as the distance between the TATA box and operator site(s) increases for both single and double operator-containing promoters. The other two parameters (v and L) showed no systematic dependence on the location or multiplicity of operator sites within the *GAL1* promoter (see Table S2 in *SI Appendix*).

Repressor Binding Site Location Affects Gene Expression Noise. We used the coefficient of variation (CV, standard deviation/mean) to characterize the effect of combinatorial promoter design on gene expression noise. We found that noise levels, especially peak noise, increased as the single operator site within the promoter was moved closer to the TATA box (Fig. 4 *B–D*). Our double operator-containing promoters show a similar relationship with respect to noise levels (Fig. 4 *E–G*): promoter D12 has the highest levels of peak noise in the double operator set, followed by promoter D13 which has higher peak noise than promoter D23, reflecting the distance of operator sites from the TATA box (Fig. 1*C*).

We also observed differences in gene expression noise when comparing promoters with different number of operators. The triple operator-containing promoter T123 shows the highest level of noise among the seven promoters (Fig. 4*H*, CV = 1.85). A general trend of increasing noise with increasing number of operators can be seen upon comparison of the triple, double and single operator-containing promoters, with some exceptions (e.g., S1 versus D23). This dependence of the noise on the multiplicity of operator sites might reflect the higher repression efficiency of multiple operator-containing promoters, which is exhibited in the basal expression (Fig. 2). These differences in gene expression noise can also be observed when analyzing CV as a function of mean expression.

Importantly, our seven synthetic promoters display significant differences in CV at the same mean expression level, across a broad range of values (see Fig. S12 in *SI Appendix*).

One advantage of the function P (Eq. **2**) compared with the empirical Hill function (1) is that the underlying chemical reaction scheme (Fig. 3) can be used to estimate the noise computationally for both single and multiple operator-containing promoters. Because the parameters obtained from fitting (i.e., v, n, and L) determine only the ratios r/ρ and a/α in Fig. 3, and not the individual rates, we introduced two scaling factors within these ratios, and estimated them using the experimentally measured noise of the wild-type promoter WT and of the single operator-containing promoter S1. Keeping these scaling factors constant, we calculated the reaction rates from the estimated parameters v, n, and L, and used the Gillespie algorithm (58) to simulate the noise for each of the single and multiple operator-containing promoters (Fig. 4). The good agreement between the simulations and experimental data (Fig. 4) indicates the advantage of our simple chemical model compared with a purely empirical function, such as the Hill function.

Computational Modeling of Promoter Repression by Single and Multiple TetR Molecules. We developed more detailed mathematical models and reaction schemes for the multiple operator-containing promoters to determine whether binding of repressors to the single operator-containing promoters S1, S2, and S3 is predictive of the dose–response curves and gene expression noise exhibited by the double and triple operator-containing promoters (see *SI Appendix*).

We replaced the repressed promoter state R in Fig. 3 with three states (R_i, R_j, and R_{ij}, i, j = 1, 2, 3) for the double operator-containing promoters, and with seven states (R_1, R_2, R_3, R_{12}, R_{13}, R_{23}, R_{123}) for the triple operator-containing promoter (see *SI Appendix*). The superposition of independent TetR binding/unbinding dynamics estimated from single operator-containing

promoters was insufficient to explain the dose–response curves and noise levels of the multiple operator-containing promoters. In particular, this assumption gave a decreasing rate of inducer-dependent leakage from the D12 promoter, and could not reproduce the dose–response curve and gene expression noise of the T123 promoter (see Fig. S5 in *SI Appendix*). Therefore, we assumed that TetR dimers can mutually affect each other's binding dynamics on the promoter, and calculated the parameters that describe this potential interaction (see *SI Appendix*). Introducing a new constant to account for the interactions between repressors improved the fit to the experimental data (see Fig. S6 in *SI Appendix*). Interestingly, the values obtained for these interaction constants suggest that repressors bound to sites S1 and S2 tend to stabilize each other, whereas the repressors bound to sites S1 and S3 or S2 and S3 destabilize each other on the DNA (59). Assuming that the interaction parameters are not constants, but depend on the inducer concentration, improved the quality of our fits even further (see *SI Appendix*). In conclusion, we believe that additional interactions are needed, besides independent repressor binding, to explain the behavior of the multiple operator-containing promoters. Spacing-dependent stabilization of DNA-bound repressor proteins has been observed in yeast (60), and additional evidence suggests that multiple TetR dimers can influence each other's operator binding dynamics (61). It will be interesting to explore experimentally whether such interactions occur in the engineered system.

Discussion

To fulfill the promise of synthetic biology, the basic building blocks of engineered gene circuits need to be well characterized, both individually and as components of integrated, complex systems (62–67). With this aim in mind, we chose to study a set of seven engineered promoters, built by inserting one, two, and three TetR-repressible operator sites in the *GAL1* promoter in various configurations. For the single operator-containing promoters, we found that the basal level of gene expression increases, whereas the steepness of the dose–response curve at high induction decreases as the operator site is moved farther from the TATA box within the *GAL1* promoter. We developed a generic chemical reaction scheme to explain the observations for all seven synthetic promoters. We also developed more detailed models, trying to explain the behavior of the multiple operator-containing promoters based on the single operator-containing promoters. We found that the multiple operator-containing promoters are predictable only after making additional assumptions, which indicates that their behavior cannot be explained as a simple superposition of the dynamics of the individual operator sites.

Our finding that the basal expression level increases with the distance of the operator from the TATA box is in agreement with previous studies on other promoters (68). In eukaryotic TetR-repressible promoters, the typical strategy is to insert single or multiple operators in the vicinity of the TATA box or near the transcription start site, with the assumption that DNA-bound TetR will interfere with the binding of general transcription factors or RNA polymerase II (69). However, strategies for *tetO_2* operator placement within promoters are not universally applicable, and different promoters from various eukaryotic species require different operator locations for optimal repression (70). For the *GAL1* promoter of *S. cerevisiae*, we found that greater repression by TetR occurs when operators are placed close to the TATA box, rather than the transcription start site.

Increasing the number of operator sites is another common strategy used to reduce basal expression in the design of TetR-repressible promoters (71). We validated this design approach with our set of promoters, as the triple operator-containing promoter T123 showed lower basal expression than any double or single operator-containing promoters. Still, positional effects

contribute strongly to the effectiveness of TetR-mediated repression and can result in higher basal expression from multiple operator-containing promoters compared with a single operator-containing promoters (e.g., Fig. 2 *B* and *G*: S1 versus D23).

Our seven synthetic *GAL1* promoters show large differences in their levels of gene expression noise, which can have important phenotypic consequences (36–43). Various factors and processes have been shown to influence gene expression noise, including gene positioning along the chromosome (72). We reveal differences in noise levels caused by operator positioning within a promoter sequence. Specifically, we found that gene expression noise typically increases when the operator is moved closer to the TATA box. This position-dependence of noise is likely related to the basal expression level, which contributes to the mean, causing a decrease in the coefficient of variation.

In synthetic gene networks, it is often necessary to reduce basal expression to achieve optimal network performance (1, 2), and to reduce gene expression noise to obtain greater consistency in signal transduction. However, our results indicate that a decrease in the basal expression level leads to an increase in noise and vice versa. These findings may be useful for establishing a cost–benefit relationship between high levels of noise and low basal expression, when designing operator configurations within a given promoter. Specifically, our results show how a commonly used regulatory component (*tetO_2*) can be best used in the design of a gene expression system to balance noise reduction with basal expression levels. Importantly, our findings demonstrate how gene expression noise can be engineered within the design of a given promoter and provide a strategy for the examination of the effects of different noise levels for a given mean value of expression (43); this will be an important tool for future studies that address the biological significance of intrinsic fluctuations.

Our results point to an important difference between electronic and biological circuit design. The integration of basic electronic components into large circuits with predictable behavior is feasible because resistors, capacitors, diodes, etc., are relatively simple and well characterized in their regimes of operation. However, basic biomolecular components can exhibit complex, context-dependent behavior when integrated into larger systems. Due to this inherent complexity, the simple superposition of the dynamics of the individual operator sites was not sufficient to explain their behavior when brought together into the *GAL1* promoter.

Through computational modeling, we were able to augment the experimental description of our biological system, and suggest interactions that might explain the experimentally observed characteristics of our seven promoters. As we show, computational modeling can suggest new interactions between the individual components, and provide possible insights into the origin of complex system behavior. Our findings highlight the utility of integrated computational–experimental approaches for studying simple regulatory elements with the aim of designing and constructing increasingly complex synthetic gene networks with predictable dynamics.

Materials and Methods

Strains and Media. *S. cerevisiae* strain YPH500 (α, *ura3-52*, *lys2-801*, *ade2-101*, *trp1Δ63*, *his3Δ200*, *leu2Δ1*) (Stratagene, La Jolla, CA) served as the host strain for all plasmid chromosomal integrations. Yeast transformations were carried out by a modified lithium acetate procedure (73). The *TRP1* selectable marker gene within the plasmids allowed for initial selection of yeast clones. Individual positive clones were then screened for single integration at the *GAL1-10* promoter region of chromosome II by PCR of isolated gDNA using *Taq* DNA polymerase (New England Biolabs, Ipswich, MA), as well as measurement of *yEGFP* expression by flow cytometry. Cultures of all strains were grown in synthetic drop-out media without tryptophan (SD-TRP) as described (9).

Plasmid Synthetic Promoter Construction. The previously described yeast integrative plasmid pRS4D1 (8) served as the template for creating a set of synthetic tet-repressible *GAL1* promoters. The plasmids used in this study differ with respect to pRS4D1 only in the number and arrangement of $tetO_2$ operator sites inserted downstream of the *GAL1* TATA box (Fig. 1).

The 19-bp $tetO_2$ operator sites were inserted downstream of the *GAL*1 TATA box by standard PCR techniques using Pfu Turbo DNA polymerase (Stratagene) on a PTC-100 Programmable Thermal Controller (MJ Research, Waltham, MA) (see Table S1 in *SI Appendix* for a complete list of primers used for each promoter construct). Each inserted operator site replaced the native promoter sequence at the corresponding positions, thus maintaining constant distance between the TATA box and transcription start site (TSS) in all promoter designs.

All plasmids used were transformed into *Escherichia coli* strain XL-10 Gold (Stratagene). Competent bacterial cells were prepared, transformed, and plated on LB agar plates containing ampicilin for selection (all Fisher BioReagents). Plasmid DNA was recovered from positive bacterial clones by the QIAprep Spin Miniprep kit (Qiagen, Valencia, CA). Proper insertion of $tetO_2$ sites into the *GAL1* promoter was then verified by sequencing (Agencourt, Beverly, MA).

yEGFP Induction Experiments. Single yeast colonies for each strain were picked from SD-TRP plates containing 2% glucose and used to inoculate 3 ml SD-TRP media containing 2% galactose. The selected colonies were then grown at 30°C with 300 rpm orbital shaking until reaching an OD_{600} of 1.0–1.5. A triplicate set of 3-ml SD-TRP cultures containing 2% galactose and anhydrotetracycline (ACROS Organics, Geel, Belgium) at a concentration range of 0–250 ng/ml was then inoculated by the initial culture to an OD_{600} of 0.01 and incubated similarly overnight. After 16–20 h, cultures reached an OD_{600} of 0.5 ± 0.2 and were subsequently assayed for *yEGFP* expression by flow cytometry.

Flow Cytometry and Data Analysis. Flow cytometry measurements were carried out as described (9). Samples were run on a low flow rate until 2,000 cells had been collected within a small forward and side scatter gate, thus reducing extrinsic sources of variation and allowing for examination of cells of similar size, shape, and point in the cell cycle. Flow cytometry data files were then analyzed by using Matlab (The MathWorks, Natick, MA). The original log-binned fluoresense intensity values were linearized, and the mean and standard deviation of these values were calculated for each sample. The noise (coefficient of variation) was computed for each sample as the standard deviation normalized by the mean.

1. Elowitz MB, Leibler S (2000) *Nature* 403:335–338.
2. Gardner TS, Cantor CR, Collins JJ (2000) *Nature* 403:339–342.
3. Becskei A, Serrano L (2000) *Nature* 405:590–593.
4. Becskei A, Seraphin B, Serrano L (2001) *EMBO J* 20:2528–2535.
5. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) *Nat Genet* 31:69–73.
6. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) *Science* 297:1183–1186.
7. Rosenfeld N, Elowitz MB, Alon U (2002) *J Mol Biol* 323:785–793.
8. Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) *Nature* 422:633–637.
9. Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) *Cell* 113:597–607.
10. Isaacs FJ, Hasty J, Cantor CR, Collins JJ (2003) *Proc Natl Acad Sci USA* 100:7714–7719.
11. Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) *Proc Natl Acad Sci USA* 101:6355–6360.
12. Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) *Nat Biotechnol* 22:841–847.
13. Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, Cantor CR, Collins JJ (2004) *Proc Natl Acad Sci USA* 101:8414–8419.
14. Kramer BP, Viretta AU, Daoud-El-Baba M, Aubel D, Weber W, Fussenegger M (2004) *Nat Biotechnol* 22:867–870.
15. You L, Cox RS, III, Weiss R, Arnold FH (2004) *Nature* 428:868–871.
16. Fung E, Wong WW, Suen JK, Bulter T, Lee SG, Liao JC (2005) *Nature* 435:118–122.
17. Hooshangi S, Thiberge S, Weiss R (2005) *Proc Natl Acad Sci USA* 102:3581–3586.
18. Kramer BP, Fussenegger M (2005) *Proc Natl Acad Sci USA* 102:9517–9522.
19. Pedraza JM, van Oudenaarden A (2005) *Science* 307:1965–1969.
20. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) *Science* 307:1962–1965.
21. Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ (2006) *Nature* 439:856–860.
22. Savageau MA (1998) *Genetics* 149:1665–1676.
23. Thieffry D, Huerta AM, Perez-Rueda E, Collado-Vides J (1998) *BioEssays* 20:433–440.
24. Yuh CH, Bolouri H, Davidson EH (1998) *Science* 279:1896–1902.
25. Pilpel Y, Sudarsanam P, Church GM (2001) *Nat Genet* 29:153–159.
26. Buchler NE, Gerland U, Hwa T (2003) *Proc Natl Acad Sci USA* 100:5136–5141.
27. Setty Y, Mayo AE, Surette MG, Alon U (2003) *Proc Natl Acad Sci USA* 100:7702–7707.
28. Wang W, Cherry JM, Nochomovitz Y, Jolly E, Botstein D, Li H (2005) *Proc Natl Acad Sci USA* 102:1998–2003.
29. Balázsi G, Barabási AL, Oltvai ZN (2005) *Proc Natl Acad Sci USA* 102:7841–7846.
30. Ptashne M (2005) *Trends Biochem Sci* 30:275–279.
31. Raser JM, O'Shea EK (2004) *Science* 304:1811–1814.
32. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) *Cell* 123:1025–1036.
33. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N (2006) *Nat Genet* 38:636–643.
34. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS (2006) *Nature* 441:840–846.
35. Chubb JR, Trcek T, Shenoy SM, Singer RH (2006) *Curr Biol* 16:1018–1025.
36. Rao CV, Wolf DM, Arkin AP (2002) *Nature* 420:231–237.
37. Thattai M, van Oudenaarden A (2004) *Genetics* 167:523–530.
38. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) *PLoS Biol* 2:e137.
39. Wolf DM, Vazirani VV, Arkin AP (2005) *J Theor Biol* 234:227–253.
40. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer DV (2005) *Cell* 122:169–182.
41. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) *Nat Rev Genet* 6:451–464.
42. Acar M, Becskei A, van Oudenaarden A (2005) *Nature* 435:228–232.
43. Blake WJ, Balázsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ (2006) *Mol Cell* 24:853–865.
44. Korobkova E, Emonet T, Vilar JM, Shimizu TS, Cluzel P (2004) *Nature* 428:574–578.
45. Ozbudak EM, Thattai M, Lim HN, Shraiman BI, Van Oudenaarden A (2004) *Nature* 427:737–740.
46. Ozbudak EM, Becskei A, van Oudenaarden A (2005) *Dev Cell* 9:565–571.
47. Ko MS (1991) *J Theor Biol* 153:181–194.
48. Swain PS, Elowitz MB, Siggia ED (2002) *Proc Natl Acad Sci USA* 99:12795–12800.
49. McAdams HH, Arkin A (1997) *Proc Natl Acad Sci USA* 94:814–819.
50. Paulsson J (2004) *Nature* 427:415–418.
51. Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS, Hasty J (2006) *Nature* 439:861–864.
52. Tiebel B, Garke K, Hillen W (2000) *Nat Struct Biol* 7:479–481.
53. Meissner M, Brecht S, Bujard H, Soldati D (2001) *Nucleic Acids Res* 29:E115.
54. Hamann L, Buss H, Tannich E (1997) *Mol Biochem Parasitol* 84:83–91.
55. Legewie S, Bluthgen N, Herzel H (2005) *FEBS J* 272:4071–4079.
56. Struhl K (1999) *Cell* 98:1–4.
57. Yean D, Gralla JD (1999) *Nucleic Acids Res* 27:831–838.
58. Gillespie DT (1977) *J Phys Chem* 81:2340–2361.
59. Vashee S, Melcher K, Ding WV, Johnston SA, Kodadek T (1998) *Curr Biol* 8:452–458.
60. Melcher K, Xu HE (2001) *EMBO J* 20:841–851.
61. Kleinschmidt C, Tovar K, Hillen W (1991) *Nucleic Acids Res* 19:1021–1028.
62. Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) *Mol Syst Biol* 2:2006–0028.
63. Chin JW (2006) *Nat Chem Biol* 2:304–311.
64. Endy D (2005) *Nature* 438:449–453.
65. Hasty J, McMillen D, Collins JJ (2002) *Nature* 420:224–230.
66. Isaacs FJ, Dwyer DJ, Collins JJ (2006) *Nat Biotechnol* 24:545–554.
67. McDaniel R, Weiss R (2005) *Curr Opin Biotechnol* 16:476–483.
68. Heins L, Frohberg C, Gatz C (1992) *Mol Gen Genet* 232:328–331.
69. Gossen M, Bonin AL, Bujard H (1993) *Trends Biochem Sci* 18:471–475.
70. Berens C, Hillen W (2003) *Eur J Biochem* 270:3109–3121.
71. Geissendorfer M, Hillen W (1990) *Appl Microbiol Biotechnol* 33:657–663.
72. Becskei A, Kaufmann BB, van Oudenaarden A (2005) *Nat Genet* 37:937–944.
73. Gietz RD, Schiestl RH, Willems AR, Woods RA (1995) *Yeast* 11:355–360.

APPLIED MATHEMATICS

BIOPHYSICS