Combinatorial Promoter Design for Engineering Noisy Gene Expression

Kevin F. Murphy[1,2], Gábor Balázsi[1], and James J. Collins[1*]

1. Department of Biomedical Engineering, Center for BioDynamics and Center for Advanced Biotechnology, Boston University, Boston, MA, USA
2. Molecular Biology, Cell Biology, and Biochemistry Program, Boston University, Boston, MA, USA

*Email: jcollins@bu.edu

Understanding the behavior of biomolecular components as parts of larger systems is one of the goals of the developing field of synthetic biology[1]. Most promoters controlling gene expression contain regulatory elements for multiple transcription factors, which must integrate a myriad of input signals into an appropriate output gene expression response. Complicating the matter, the combination of regulatory sites in a promoter region can result in behavior that is not predictable from studying the individual sites alone[2,3]. 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. Here we used an integrated computational-experimental approach to characterize how the position and multiplicity of *tet*O_2 operator sites within the *GAL1* promoter affect gene expression levels and gene expression noise in *Saccharomyces cerevisiae*. We built a combinatorial set of seven synthetic promoters containing one, two or three *tet*O_2 operators in all possible configurations of three unique positions within the *GAL1* promoter. We used the coefficient of variation (CV, standard deviation/mean) to characterize the effect of these different *tet*O_2 operator combinations on gene expression noise from the *GAL1* promoter. We observed lower basal expression levels and found that noise levels, especially peak noise, increased as a single operator site within the promoter was moved closer to the TATA box (Figure 1B-D). Similar relationships with respect to basal expression and noise levels were also observed with our double operator-containing promoters (Figure 1E-G). Additionally, a general trend of increasing noise with increasing number of operators was observed when analyzing CV as a function of both inducer concentration (Figure 1B-H) as well as mean expression level (Figure 2) for our set of synthetic promoters. In synthetic gene networks, it is often necessary to reduce basal expression to achieve optimal network performance[4,5], 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. In addition to our experimentation, 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, 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. Additionally, 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.

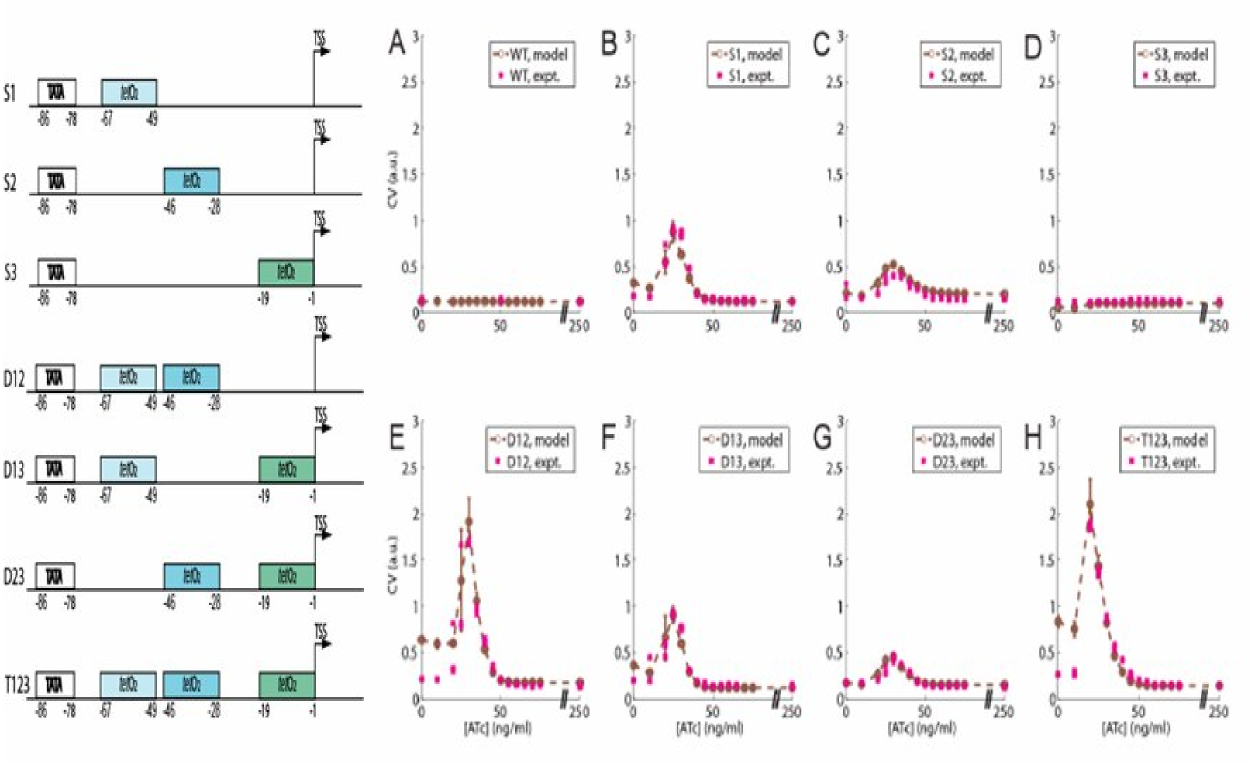

Figure 1. Left, diagram of P_{GAL1*} promoter constructs containing all seven *tet*O_2 operator combinations. The TATA box and *tet*O_2 operator locations are indicated by base position number relative to transcription start site (TSS). The name of each promoter is indicated to its left in the diagram. 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. A-H**.** 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.

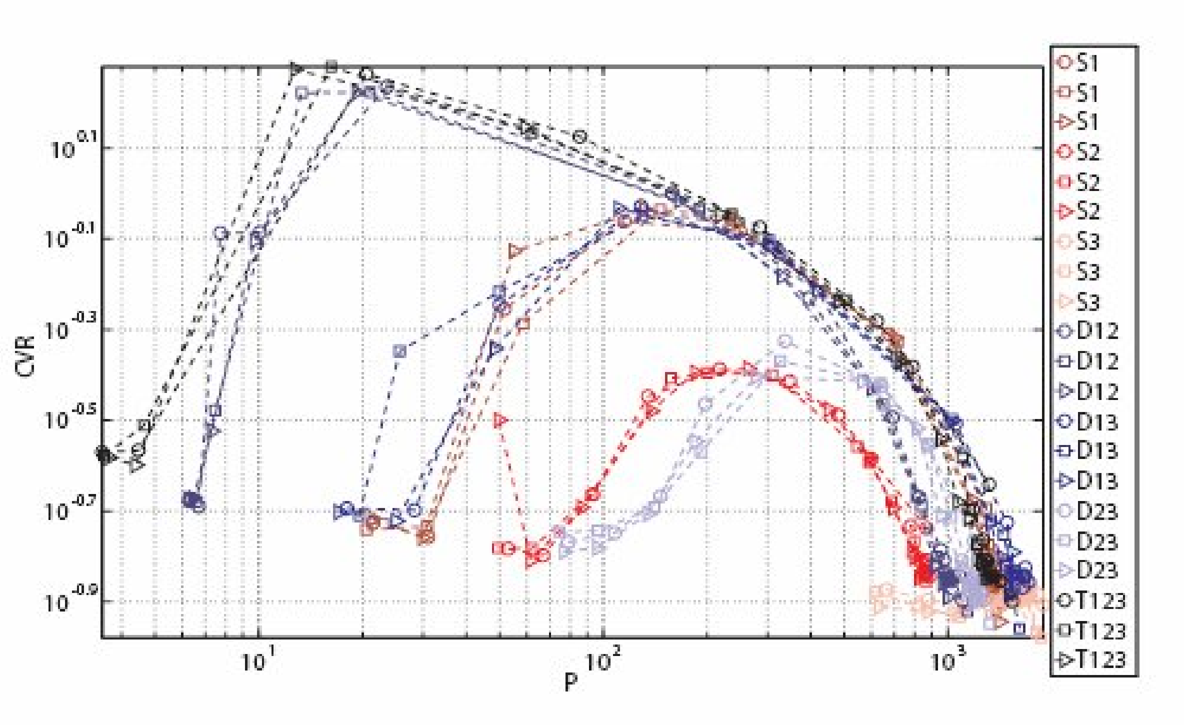

Figure 2. Coefficient of variation (CVR) as a function of mean expression (P) for all seven synthetic *GAL1* promoters. Three independent experiments are shown for each promoter.

References:

1. Murphy, KF, Balázsi, G, Collins, JJ (2007) *Proc Natl Acad Sci USA* **104**:12726-12731.
2. Setty Y, Mayo AE, Surette MG, Alon U (2003) *Proc Natl Acad Sci USA* **100**:7702–7707.
3. Ptashne M (2005) *Trends Biochem Sci* **30**:275–279.
4. Elowitz MB, Leibler S (2000) *Nature* **403**:335–338.
5. Gardner TS, Cantor CR, Collins JJ (2000) *Nature* **403**:339–342.