# Quantitative characterization of synthetic gene regulatory network libraries

Tsz-Leung To, Narendra Maheshri*
Department of Chemical Engineering,
Massachusetts Institute of Technology, Cambridge, USA
*email: narendra@mit.edu

## Abstract

The remarkable success of directed evolution techniques in protein engineering is due to the highly evolvable nature of protein folds [1]. Natural genetic regulatory networks also appear to be evolvable [2, 3], requiring only small genetic changes to lead to qualitatively different behavior. Directed evolution of a simple synthetic genetic circuit [4] and promoter mutagenesis studies [5] suggests its utility in cell and metabolic engineering.

In this work, we propose and test a methodology that adapts the tools of directed evolution to 1) rapidly generate synthetic regulatory networks with a desired dynamical behavior and 2) determine the limits of current kinetic models of these networks. We define one component of a regulatory network as a single gene (with promoter) and any gene-specific transcription factors. Its steady-state behavior is captured using a model defined by a small set of thermodynamic and kinetic parameters. Next, we create a library of this component by mutagenesis (here, of the promoter) - each member then has a potentially different set of parameters (Figure 1A). Rather than looking at individual members, we quantitatively characterize the *distribution* of parameter values in the entire library. Finally, this library is plugged into a larger network whose underlying kinetic model suggests qualitatively different steady-state attractors depending on parameter values. The parameter distribution associated with a library then gives a prediction for the percentage of library members that possess each of the possible steady-state behaviors. We do not expect perfect agreement, but 1) we can bias our chance of success of obtaining a particular behavior by designing mutations to give a favorable parameter distribution, and 2) use unbiased sampling to understand if the model misses unknown interactions that occur in limited parameter regimes.

To demonstrate this approach, we employ the tet-trans activator (tTA) system [6] in *S. cerevisiae* and monitor single cell gene expression state with fluorescent proteins (Figure 1B). Because there is a response distribution even in a genetically homogeneous sample, we used global fluorescent reporters to eliminate non-genetic variation due to plasmid fluctuations or global noise (Figure 1C). A library of *tetO* mutants (diversity ~ $10^3$) was constructed that responded differently to doxycycline (Figure 2A). Analysis of the wild type and individual mutants confirmed that a 4-parameter model could describe steady-state gene expression and mutants exhibited differences in $K_D$, the dissociation constant for tTA and the *tetO* binding site (Figure 2B). We used the response data of the library to estimate the $K_D$ distribution (Figure 2C). The one component network was rewired in a positive feedback loop (Figure 2D) and the $K_D$ distribution successfully predicted the percentage of clones that possessed a growth defect due to high levels of tTA [7] (see Figure 2 legend for details).

We are continuing this approach by measuring parameter distributions of libraries using promoters involving multiple *tetO* sites and building larger regulatory networks with many possible steady-states. Furthermore, we are constructing models that link types of mutagenesis to parameter changes.

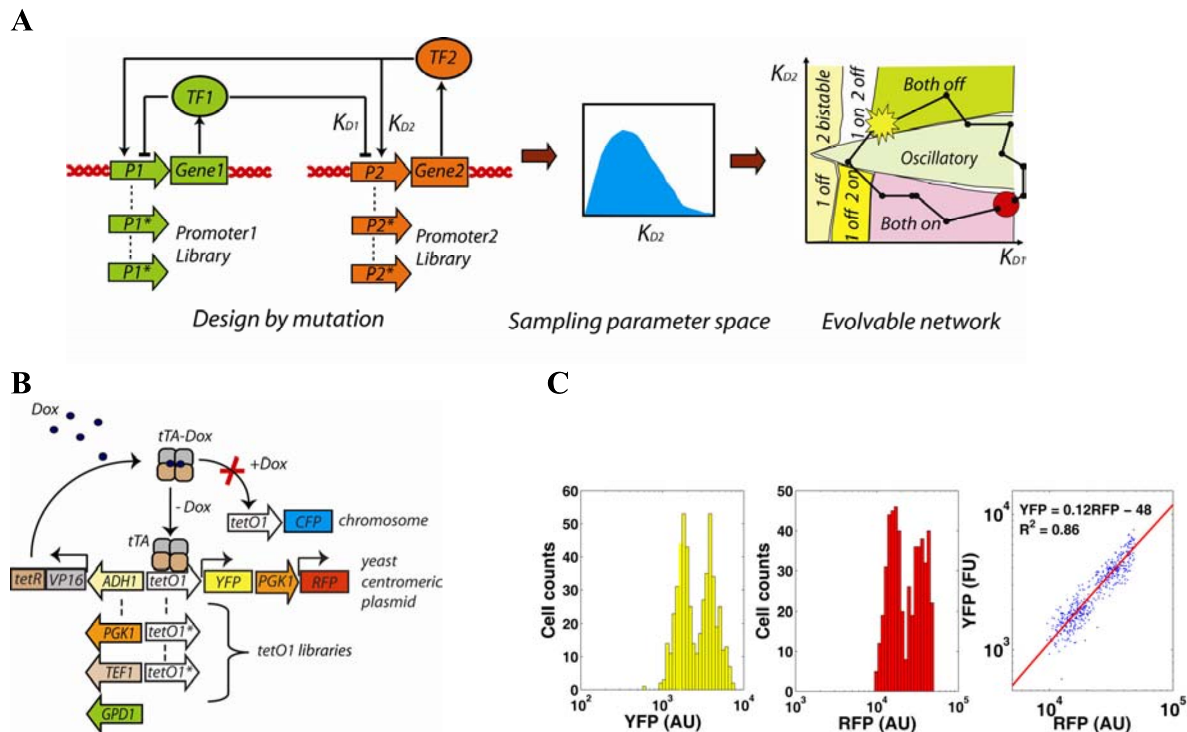
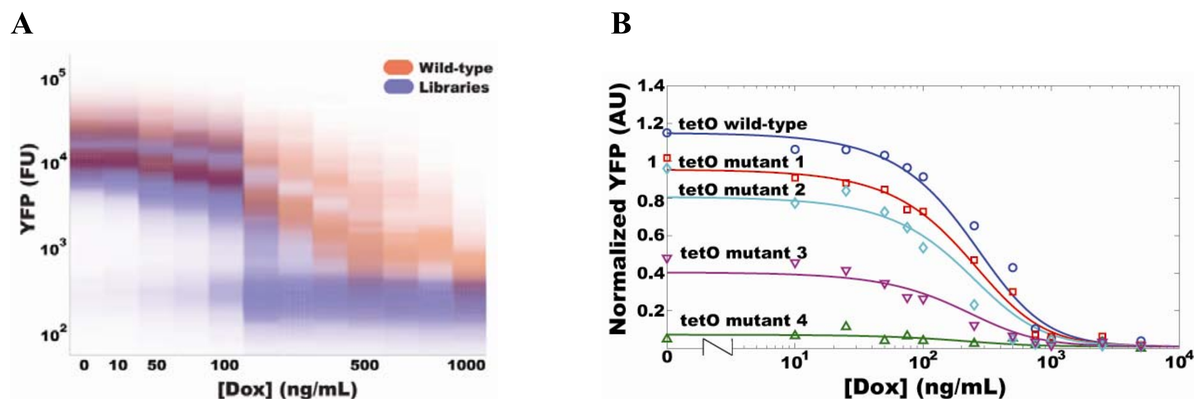
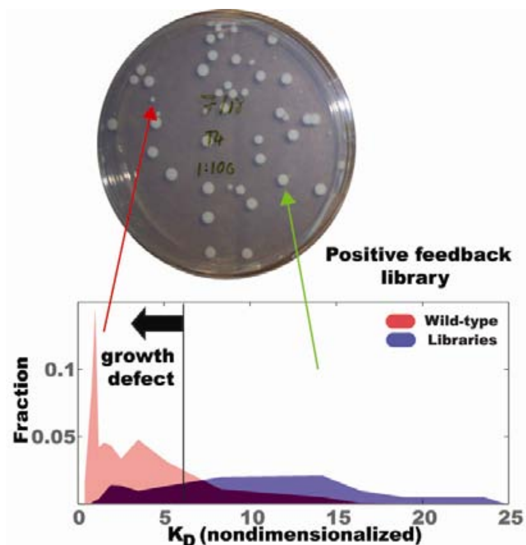



**Figure 1: General concept and the core one component construct.** (**A**) A genetic regulatory network can be described by a dynamical system model, whose parameters determine its steady-state behavior. To experimentally sample the parameter space of this network, we create mutagenized libraries of one component of this network. (**B**) The one component is composed of the tet-trans activator (tTA) and a tTA-inducible promoter (*tetO*$_1$), driving expression of YFP on a low copy yeast centromeric vector. Non-genetically induced fluctuations in gene expression (from plasmid and global noise) are controlled for by the presence of a constitutively expressed RFP on the plasmid. A second tTA-inducible promoter lies in the chromosome and drives CFP expression. Active tTA levels can be controlled by addition of the small molecule doxycycline, which inhibits binding of tTA to the *tetO* binding site. Alternatively, the strength of the constitutive promoter driving tTA expression can be altered. (**C**) Microscopic analysis of measured YFP expression in yeast strains containing the one-component system indicates a bimodal expression profile. This is due to a doubling of plasmid copy number (data not shown). The global reporter RFP is well correlated with YFP and accounts for both the plasmid fluctuation and other sources of cell-to-cell variation. Effectively over 80% of the non-genetic cell to cell variation can be accounted for by the RFP reporter.



C




D

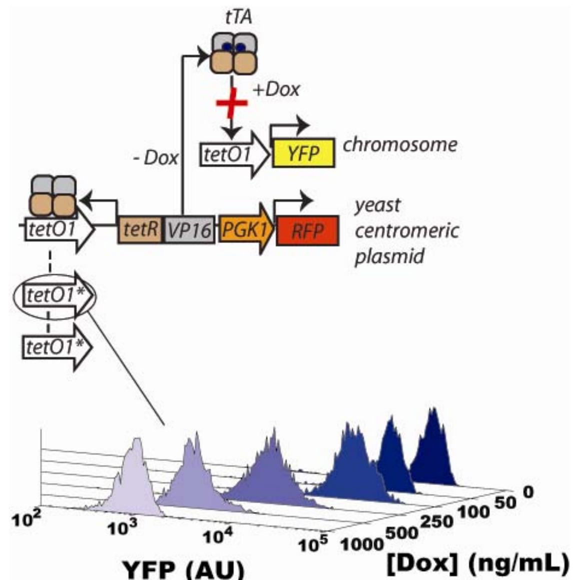



**Figure 2: Characterization and application of one component libraries.** (**A**) Overlapped density plots of steady-state YFP expression in the wild-type and *tetO₁* mutant libraries at various doxycycline concentrations. The genetically homogenous wild-type strain shows less but still significant variation (the two populations correspond to plasmid fluctuations) that can be reduced by over 80% by normalizing with respect to a global reporter (see Figure 1C). (**B**) The dose response curves of the wild-type strain and four selected individual mutant clones along with fits (solid line) to a Hill-like model. The only one parameter varied across these clones is the dissociation constant, $K_D$, between tTA and its DNA binding site. (**C**) Using the data in A, we estimate $K_D$ distribution. The multi-peaked nature of the wild-type distribution is because of plasmid fluctuations that can be eliminated with a global reporter. Nevertheless, even without accounting for these fluctuations, the library $K_D$ distribution spans a much larger range. Moreover, based on toxicity experiments, a growth defect occurs when $K_D < 6$. The $K_D$ distribution correctly predicts that 20-25% of the positive feedback mutants have retarded growth. (**D**) The positive feedback construct has severe growth defect when driven by the wild-type $tetO_1$. Mutants with a larger $K_D$ express lower levels of tTA and do not possess this defect. They show a graded response to dox confirming tTA binding is not cooperative as assumed by the one component model.

## References


1. Kirschner M & Gerhart J. (1998) *Proc Natl Acad Sci U S A* 95: 8420-8427.
2. Mayo AE, Setty Y, Shavit S, Zaslaver A & Alon U. (2006) *PLoS Biol* 4: e45.
3. Voigt CA, Wolf DM & Arkin AP. (2005) *Genetics* 169: 1187-1202.
4. Yokobayashi Y, Weiss R & Arnold FH. (2002) *Proc Natl Acad Sci U S A* 99: 16587-16591.
5. Alper H, Fischer C, Nevoigt E & Stephanopoulos G. (2005) *Proc Natl Acad Sci U S A* 102: 12678-12683.
6. Gossen M & Bujard H. (1992) *Proc Natl Acad Sci U S A* 89: 5547-5551.
7. Brown CE, Howe L, Sousa K, Alley SC, Carrozza MJ, Tan S & Workman JL. (2001) *Science* 292: 2333-2337.