Predictive Design of Synthetic Microbes

Howard M. Salis[1], Christopher A. Voigt[2]

[1]Pennsylvania State University (AB Engineering & Chemical Engineering)
[2]University of California San Francisco (Pharmaceutical Chemistry)
salish@picasso.ucsf.edu

Controlling Gene Expression

Summary: When engineering a synthetic biological system, such as a metabolic pathway or genetic network, the production rates of its proteins often require tuning to obtain a specific behavior. For large systems, a rational approach is required.

We have created a predictive design method that generates the sequence of a synthetic ribosome binding site to yield a *user-specified* protein production rate. The method has been experimentally validated in *Escherichia coli*. A web interface is available at http://voigtlab.ucsf.edu/software.

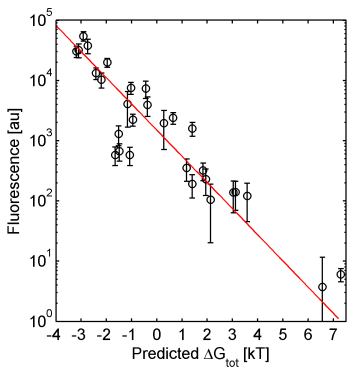
A Statistical Thermodynamic Model of Translation Initiation

Translation initiation is a rate-limiting step in gene expression. We can relate the translation initiation rate of a mRNA to the free energy change ΔG_{tot} according to:

$$r \propto \exp\left(-\beta \Delta G_{tot}\right)$$

Given a specific mRNA sequence, we predict the ΔG_{tot} according to an energy model that quantifies the Ribosome – RNA and the RNA – RNA interactions:

$$\Delta G_{tot} = \left(\Delta G_{mRNA:rRNA} + \Delta G_{spacing} + \Delta G_{start}\right) - \left(\Delta G_{mRNA} - \Delta G_{standby}\right)$$



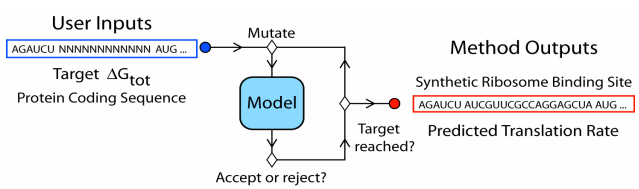
Experimental Validation

We combined Monte Carlo optimization with the model to design 29 synthetic ribosome binding sites to each yield a *specified rate* of fluorescent protein production in *E. coli*.

Comparing this data to the theory, the average error in the prediction is $|\Delta\Delta G|$ = 0.82 kT (R^2 = 0.84).

The design method allows us to *choose* the protein production rate over a 100,000 fold range.

Forward design with Monte Carlo optimization



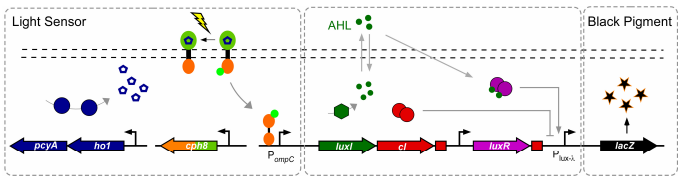
Engineering Synthetic Gene Networks

Summary: A bacteria's genetic code can be programmed to execute complex algorithms. A central goal of synthetic biology is to use quantitative modeling to predict the genetic code that results in a desired algorithm.

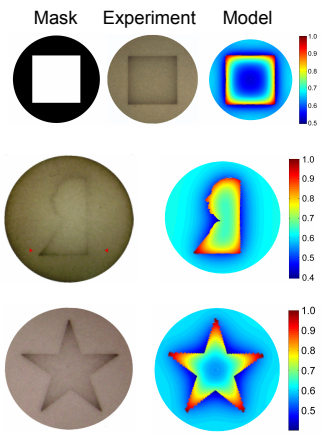
We have constructed a synthetic gene network that combines cell-to-cell communication with logical cellular computing to execute an edge detection algorithm – a lawn of bacteria detects the boundaries between light/dark regions on a Petri dish. We describe its spatiotemporal dynamics using a reaction-diffusion model. This work was carried out with Dr. Jeffrey Tabor.

A Synthetic Genetic Edge Detection Program

A light sensor represses the P_{ompC} promoter in the presence of red light. The P_{ompC} promoter produces LuxI – an enzyme that synthesizes the diffusible signal AHL – and the transcriptional repressor cI. In the presence of AHL and the absence of cI, the $P_{lux-\lambda}$ promoter expresses LacZ – an enzyme that makes black pigment.



Edge Detection Experiments and the Reaction-Diffusion Model:



J. Tabor, H. Salis, Z. Simpson, A. Chevalier, A. Levskaya, E. Marcotte, C.A. Voigt, and A. Ellington, *Cell*, 2009

Model Equations

$$\text{INPUT} \rightarrow L_{(r,\theta)} \qquad f_{light} = \frac{K}{K + L_{(r,\theta)}}\left(\beta_{max} - \beta_{min}\right) + \beta_{min}$$

$$\frac{\partial u_1}{\partial t} = \frac{1}{r}\frac{\partial u_1}{\partial r} + \frac{\partial^2 u_1}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 u_1}{\partial \theta^2} + \kappa_1 f_{light} - \kappa_2 u_1$$

$$u_2 = \kappa_3 f_{light}$$

$$u_3 = \frac{c_0 + c_1 f_{Lux}}{1 + c_1 f_{Lux} + c_2 f_{CI}^{\ n} + c_1 c_2 f_{Lux} f_{CI}^{\ n}} \rightarrow \text{OUTPUT}$$

$$f_{CI} = \frac{u_2}{2} + \frac{1}{8K_{CI}}\left(1 - \sqrt{1 + 8K_{CI}u_2}\right)$$

$$f_{Lux} = \frac{1}{2}\left(\text{LuxR}_{tot} + \frac{K_A^2 K_D}{4u_1^2}\right) - \frac{1}{2}\sqrt{\left(\text{LuxR}_{tot} + \frac{K_A^2 K_D}{4u_1^2}\right)^2 - \text{LuxR}_{tot}^{\ 2}}$$

These equations are solved in dimensionless form with a no-flux boundary condition at r = 1 and a uniformly zero initial condition. *L* is the light input, u_1 is AHL, u_2 is cI, and u_3 is LacZ.

There are 13 parameters total: 7 parameters were measured by characterizing the individual genetic components, 3 parameters were previously measured, and 3 were estimated.

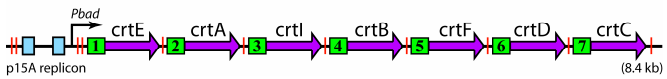
Optimizing Metabolic Pathways

Summary: The introduction of a synthetic metabolic pathway into a microbial host enables the production of fuels, specialty chemicals, and drugs from renewable feedstocks, including cellulose. However, these metabolic pathways often require optimization for their new host.

We use our new ability to control gene expression (left panel) to optimize a seven enzyme metabolic pathway that produces a carotenoid pigment. The pathway is encoded by a completely synthetic operon – the DNA is synthesized by a machine according to *our specifications*.

A Synthetic Carotenoid Pathway

We designed a 8400 nucleotide DNA sequence to encode the seven *crt* enzymes that convert isopentenyl PP to hydroxy-spheroidene. In the process, we removed all extraneous regulation and cryptic natural sequences, codon optimized the enzymes for *Escherichia coli*, and replaced the ribosome binding site sequences.



The flux through a metabolic pathway is controlled by the production rates of its enzyme and the sequences of the ribosome binding sites

Pathway Optimization with Rational Sequence Design

We tested the synthetic carotenoid operon in *E coli* and measured carotenoid pigment accumulation with HPLC/spectroscopy. The first design iteration (v1.0) resulted in accumulation at three intermediates with no final product. Using the predictive design method (left panel), we redesigned a 30 nucleotide sequence to create a second version of the operon (v2.0). The second version accumulates hydroxy-spheroidene – the final product. Additional design iterations are currently in progress with the goal of maximizing the product yield.

Experimental data: HPLC/spectroscopy