Transcription timing and network analysis of galactose utilization network of *Escherichia coli*

– doctoral theses –

Author: Péter HORVÁTH,
Supervisor: Szabolcs SEMSEY, PhD

Doctoral School: Eötvös Loránd University, Biology Doctoral School
Head of School: Prof. Anna ERDEI, CMHAS

Doctoral Program: Classic and Molecular Genetics Program
Head of Program Prof. László OROSZ, CMHAS

Introduction

The function of living cells is controlled by complex regulatory networks that are built of a wide diversity of interacting molecular components. The sheer size and intricacy of molecular networks of even the simplest organisms are obstacles toward understanding network functionality. In this work we address two important network features, regulation of timing of gene expression and the flexibility of gene regulation in integrating environmental signals.

We use the galactose regulatory system of *Escherichia coli* as a model system because the elements of this network are well characterized and a simplified mathematical model describing the behavior of the network has been previously published. [1] The *gal* system of *E. coli* contains genes involved in the transport (*galP*, *mglBAC*) and amphibolic utilization (*galETKM*) of the sugar D-galactose. Genes of the *gal* regulon belong to different operons. t4 This setup allows differential regulation of functions when needed. Regulation of the *gal* system is governed by two similar regulators, GalR and GalS, which are regulated in different ways. [2-4] Besides sensing the intracellular sugar level, galactose utilization is also regulated by the cAMP-CRP complex. cAMP is a signal of carbon shortage and sensed by CRP (cAMP Receptor Protein). [5]

Enzymes encoded by the *galETKM* operon are responsible for the amphibolic utilization of galactose. The operon has two overlapping promoters: $P1_{galE}$ and $P2_{galE.}$ All of the Gal enzymes are required for galactose metabolism, however,only GalE is needed for making substrates for biosynthetic glycosylation reactions in the absence of external galactose.Transcription from the $P1_{galE}$ promoter results in equimolar expression of the Gal enzymes, serving the catabolic requirements, while transcription from the $P2_{galE}$ promoter results in dis-coordinated expression (moreproduction of GalE and GalT than thepromoter-distal GalK and GalM), whichsuits the biosynthetic requirements [6-9].

Proper timing of gene expression is crucial in regulation of critical biological phenomena, i.e., cellular adaptation, differentiation, and development. [10, 11] Biological systems evolved two fundamentally different mechanisms for the regulation of timing. One possibility is that after an initial decision a regulatory network proceeds autonomously executing a pre-defined program. [11, 12] The other possibility is that the network monitors specific signals in real-time and regulates timing accordingly. In the natural environment bacterial cells have to adjust their metabolism to alterations in the availability of food sources. Nutrients are often limited in the environment (famine) but there are nutrient rich periods as well (feast). The metabolic and morphological changes accompanying the transitions in the feast-famine cycles

have been studied in detail [13]. The goal of our work was to understand the principles governing the timing of gene expression in sugar networks when cells are grown in a batch culture containing a single sugar as a carbon source.

The galactose network, similar to many other sugar networks, is a natural two-input system, regulated by a specific sugar (D-galactose) and the global regulatory signal cAMP. Therefore it is an optimal system to study signal integration at promoters. Gene regulation is based on simple building blocks such as promoters, transcription factors (TFs) and their binding sites on DNA. Previous studies suggested that gene regulatory sequences and interactions between TFs play a major role in network flexibility.[14, 15] The question is how diverse the functions are that can be obtained by different arrangements of promoters and TF binding sites.

Methods

For construction of reporter strains we used the recombineering method based on the λ *red* system [16] to recombine our promoter constructions replacing the 5' regulatory region of *uidABC* (also known as *gus*) operon. We constructed a vector (pSEM2027) for fast and easy swapping of promoter sequences. A PCR primer pair was designed for amplifying the different promoter constructions from the different pSEM2027 based vectors. The PCR product contained a zeo^R gene, the $rrnBT_1T_2$ terminator, the promoter sequence between *Eco*RI and *Pst*I sites, and two homologous sites at the two ends for the recombination.

We simulated the famine-feast cycle of bacterial growth by diluting stationary phase *E. coli* cells in fresh medium containing galactose as the sole carbon source. At different time points samples were taken and β-glucuronidase enzyme activity (UidA activity) was measured, and changes in promoter activities were calculated.

We used simulated annealing method for modeling the intracellular cAMP and galactose levels.

Results:

1. We find that galactose metabolism has a global effect on gene expression (e.g. through influencing nucleotide triphosphate levels), and also a specific effect on the transcription of genes belonging to the galactose regulon (through the intracellular D-galactose and cAMP levels, influencing the activity of GalR, GalS and CRP).

2. We computed changes in the promoter activities of the galactose regulon genes and in the intracellular Dgalactose and cAMP-CRP levels. The galactose system monitors the intracellular D-galactose and cAMP-CRP levels in realtime and computes the promoter activities accordingly.

3. Based on the experimental results, we built a mathematical model to calculate intracellular D-galactose and cAMP-CRP levels. We found that the cell responds to a decreasing external galactose level by increasing the internal galactose level, which is achieved by limiting galactose metabolism and increasing the expression of transporters. We showed that the cell alters gene expression based primarily on the current state of the cell, and not on monitoring the level of extracellular galactose in real-time. Some decisions have longer term effects, therefore the current state does subtly encode the history of food availability. Our measurements of timing of gene expression in the galactose system suggest that the system has evolved to respond to environments where future galactose levels are unpredictable, rather than regular feast and famine cycles.

4. We measured transcription of the *gal* regulon promoters by stationary phase (σ^{38}) and vegetative (σ^{70}) RNA polymerase Because the *galR* and *galS* genes are poorly transcribed by σ^{38} RNAP, we suggest that repression of the *gal* regulon promoters in the stationary phase is maintained by the repressor proteins accumulated earlier.

5. Limitation in galactose transport due to the decreasing extracellular galactose concentration can be compensated by increasing the concentration of transporters. To obtain higher expression of transporters, cells need to build up a higher level of intracellular galactose, which in turn also allows increased production of the enzymes involved in galactose utilization. If the system did not reduce the rate of galactose

utilization when the galactose influx decreases, the cell would quickly use up the intracellular galactose pool, leading to switching off of promoters by reactivated GalR and GalS and therefore reduced galactose transport and utilization. This effect can be overcome by changing the relative strengths of the transport and metabolism feedback loops.

6. We constructed synthetic regulatory regions using promoter elements and binding sites of two non-interacting TFs (GalR and CRP), each sensing a single environmental input signal. We showed that simply by combining these three kinds of elements, we can obtain 11 out of the 16 Boolean logic gates that integrate two environmental signals in vivo. Further, we demonstrated how combination of logic gates can result in new logic functions.

7. Our results suggest that simple elements of transcription regulation form a highly flexible toolbox that can generate diverse functions under natural selection. We find that transcription activation, i.e., recruitment of RNAP to promoter sequences is a crucial determinant of the flexibility of regulatory logic.

Publications:

Horvath, P. Hunziker, A. Erdossy, J. Krishna, S. Semsey, S., *Timing of Gene Transcription in the Galactose Utilization System of Escherichia coli.* J Biol Chem, 2010. **285**(49): p. 38062-8.

Hunziker, A., Tuboly, C. **Horvath, P**. Krishna, S. Semsey, S., *Genetic flexibility of regulatory networks.* Proc Natl Acad Sci U S A, 2010. **107**(29): p. 12998-3003.

Semsey, S. Krishna, S. Erdossy, J. **Horvath, P**. Orosz, L. Sneppen, K. Adhya, S., *Dominant negative autoregulation limits steady-state repression levels in gene networks.* J Bacteriol, 2009. **191**(14): p. 4487-91.

References

1. Krishna, S., et al., *Relation of intracellular signal levels and promoter activities in the gal regulon of Escherichia coli.* J Mol Biol, 2009. **391**(4): p. 671-8.
2. Geanacopoulos, M. and S. Adhya, *Functional characterization of roles of GalR and GalS as regulators of the gal regulon.* J. Bacteriol., 1997. **179**(1): p. 228-234.
3. Semsey, S., et al., *Signal integration in the galactose network of Escherichia coli.* Mol Microbiol, 2007. **65**(2): p. 465-76.
4. Weickert, M.J. and S. Adhya, *Control of transcription of gal repressor and isorepressor genes in Escherichia coli.* J Bacteriol, 1993. **175**(1): p. 251-8.
5. Weickert, M.J. and S. Adhya, *The galactose regulon of Escherichia coli.* Mol Microbiol, 1993. **10**(2): p. 245-51.
6. Lee, H.J., et al., *Establishment of an mRNA gradient depends on the promoter: an investigation of polarity in gene expression.* J Mol Biol, 2008. **378**(2): p. 318-27.
7. Ullmann, A., E. Joseph, and A. Danchin, *Cyclic AMP as a modulator of polarity in polycistronic transcriptional units.* Proc. Natl. Acad. Sci. U. S. A., 1979. **76**(7): p. 3194-3197.
8. Semsey, S., K. Virnik, and S. Adhya, *Three-stage regulation of the amphibolic gal operon: from repressosome to GalR-free DNA.* J Mol Biol, 2006. **358**(2): p. 355-63.
9. Adhya, S., *Suboperonic regulatory signals.* Sci. STKE, 2003. **2003**(185): p. pe22.
10. Gaudet, J., et al., *Whole-genome analysis of temporal gene expression during foregut development.* PLoS Biol, 2004. **2**(11): p. e352.
11. Amir, A., et al., *Noise in timing and precision of gene activities in a genetic cascade.* Mol Syst Biol, 2007. **3**: p. 71.
12. Johnson, C.H., *Precise circadian clocks in prokaryotic cyanobacteria.* Curr Issues Mol Biol, 2004. **6**(2): p. 103-10.
13. Kolter, R., D.A. Siegele, and A. Tormo, *The stationary phase of the bacterial life cycle.* Annu Rev Microbiol, 1993. **47**: p. 855-74.
14. Buchler, N.E., U. Gerland, and T. Hwa, *On schemes of combinatorial transcription logic.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5136-41.
15. Mayo, A.E., et al., *Plasticity of the cis-regulatory input function of a gene.* PLoS Biol, 2006. **4**(4): p. e45.
16. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.