Combinatorial engineering of intergenic regions to tune expression of multiple genes in operons

Brian F. Pfleger[1], Douglas J. Pitera[1], Christina D. Smolke[1,†], and Jay D. Keasling[1,2,3]**

Note: This work is published in recent issues of Nature Biotechnology and Metabolic Engineering. All figures and tables are presented there.

Pfleger, B.F., Pitera, D.J., Smolke, C.D. and Keasling, J.D., 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol.24 8 1027-32. http://www.nature.com/nbt/journal/v24/n8/abs/nbt1226.html

Pfleger, B.F., Pitera, D.J., Newman, J.D., Martin, V.J.J., and Keasling, J.D., (In Press). Microbial Sensors for Small Molecules: Development of a Mevalonate Biosensor. Metab Eng. http://www.sciencedirect.com/science/journal/10967176

[1] Department of Chemical Engineering,
University of California Berkeley, CA 94720-1462

[2] Department of Bioengineering,
University of California Berkeley, CA 94720-1762

[3] Synthetic Biology Department, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

*To whom correspondence should be addressed. Mailing address: Berkeley Center for Synthetic Biology, 717 Potter Street, Building 977, Mail code 3224, University of California, Berkeley, CA 94720-3224. USA. Phone: (510) 495-2620. Fax: (510) 495-2630. E-mail: keasling@berkeley.edu

†Current address: Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

Many applications of synthetic biology require the balanced expression of multiple genes, and operons are convenient for coordinating the expression of two or more genes in prokaryotes and eukaryotes. Unfortunately, it is difficult to coordinate the many post-transcriptional processes that determine the relative levels of gene expression in operons through *a priori* design. We describe here a novel method for tuning the expression of multiple genes within operons by generating libraries of Tunable InterGenic Regions (TIGRs) by recombining various post-transcriptional control elements and screening for the desired relative expression levels. TIGRs were employed to vary the relative expression of two reporter genes over a 100-fold range and to balance expression of three genes in an operon encoding a heterologous, mevalonate biosynthetic pathway, resulting in a seven-fold increase in production. We anticipate that this technology will be generally useful for tuning expression of multiple genes in synthetic operons, both in prokaryotes and eukaryotes.

The synthesis of natural or unnatural products in microorganisms usually involves the introduction of several genes encoding the enzymes of a metabolic pathway[1, 2]. In order to produce these molecules at commercially relevant levels, the genes need to be expressed in a coordinated fashion at appropriately balanced levels to avoid bottlenecks in the biosynthetic pathway that result in suboptimal yields and can lead to the accumulation of potentially toxic intermediates. Similarly, the introduction into a cell or manipulation of multi-subunit proteins (e.g., F_1F_0-ATPase, proteosomes, ion channels, etc.) usually involves coordinated expression of several genes to produce the subunits at the appropriate stoichiometries[3].

Unfortunately, it is challenging to simultaneously control and balance the expression of several genes to achieve optimal function. Not only is it difficult to predict the specific activities

of every enzyme in a pathway, heterologous genes may contain rare codons or RNase sites that limit their expression. As such, it is nearly impossible to predict the necessary strengths of the promoters and ribosome binding sites (RBSs) required to balance expression of multiple genes. Compounding this prediction problem is the general lack of tools to simultaneously balance and coordinate the expression of several genes. The lack of well-controlled, well-characterized biological components for pathway construction is one of the key drivers for the new area of synthetic biology[4].

Grouping multiple, related genes into operons, as is done naturally in prokaryotes[5], is a convenient means for regulating several genes simultaneously without the need for multiple promoters. Internal ribosomal entry sequences (IRESs) from eukaryotic viruses and host stress response pathways perform a similar function and have been harnessed to create operons for heterologous expression of genes in eukaryotes[6-9]. With a single promoter controlling the transcription of several genes, relative expression of each open reading frame in the operon is controlled by altering post-transcriptional processes (transcription termination[10, 11], mRNA stability[12, 13], and translation initiation[14-16]. Previous work has demonstrated that sequences inserted into the intergenic region (IGR) of bacterial operons can direct the processing and segmental stability of a transcript containing multiple coding regions[17, 18]. This type of directed mRNA processing resulted in differential production of the proteins encoded in the operon depending on the nature of the IGR between the coding regions.

One of the major obstacles to implementing this type of control is the difficulty in designing these control elements because of the many interrelated variables involved in transcription termination, mRNA stability, and translation initiation[19]. While our previous work[17, 18] demonstrated that it is possible to differentially control the protein levels encoded by

two or more genes in an operon using IGR sequences, it is difficult or impossible with the current state of knowledge to choose *a priori* the sequences that will balance expression of the genes in an operon. Here, we demonstrate a method for simultaneously tuning the expression of several genes within operons by generating and screening through large libraries of Tunable InterGenic Regions (TIGRs) containing various control elements (mRNA secondary structures, RNase cleavage sites, RBS sequestering sequences, etc.). An operon reporter system containing the genes encoding the red fluorescent protein DsRed (rfp_{EC})[20, 21] and the green fluorescent protein GFP (gfp_{UV}) was designed to facilitate high-throughput measurement of relative gene expression resulting from the libraries of TIGRs.

A large library of TIGR sequences (>10^4) was assembled combinatorially from four sets of oligonucleotides using polymerase chain reaction (PCR). Each oligonucleotide contained two 15-nt sequences that hybridized to a corresponding sequence in the neighboring oligonucleotide, such that a series of chimeric DNA molecules containing oligonucleotides from each of the four sets was created after several rounds of PCR. Between the hybridization sequences at either end of each oligonucleotide was a variable sequence that provided the diversity of features designed into the library. PCR amplification of this DNA pool with end-specific oligonucleotides enriched the population with full-length TIGRs containing a member from each set of oligonucleotides. Specific restriction sites incorporated into the amplification primers were used to clone the TIGR library between the two reporter genes.

The TIGR pool that resulted from the assembly of the oligonucleotides was designed to contain three regions, two variable hairpin sequences flanking a single-stranded region incorporating various RNase E sites[22, 23]. When transcribed, those TIGR sequences that contained a strong endonuclease site would be cleaved generating two secondary transcripts

whose stability could be individually modulated by the secondary structures flanking the RNase site[17]. The TIGR sequences also incorporated mRNA secondary structures of various lengths, GC contents, asymmetries, and mismatched bulges. Greater than 10^4 possible TIGRs were generated from the nine to eleven oligonucleotides in each of the four sets. Inclusion of degeneracies in some of the oligonucleotides and the use of error-prone PCR conditions increased the number of possible sequence combinations but the actual size of the library was likely determined by the number of clones attained by electroporation.

The described TIGR method utilized combinatorial assembly of oligonucleotides possessing various regulatory sequences to generate libraries of operons whose relative expression varied over a 100-fold range. The sequences used in this study incorporated three mechanisms to control protein production: differential mRNA processing, premature transcription termination, and translation initiation inhibition. These mechanisms are interrelated, and the effects of one can not be easily decoupled from the others[19], making it nearly impossible to design *de novo* operons using these and other post-transcriptional regulatory mechanisms to coordinately regulate the expression of multiple genes. The ability to access multiple regulatory mechanisms simultaneously without specific design is a major strength of the TIGR approach. By accessing multiple regulatory mechanisms the TIGR method was able to assemble an improved version of pBad33MevT that when expressed generated seven-fold more mevalonate. More importantly, the mechanisms behind the improved mevalonate production were counterintuitive. Designed constructs may not have focused on reducing the expression of HMGS and tHMGR. Nonetheless, this solution resulted in significant improvements in mevalonate production. Most importantly, incorporation of this optimized operon into the strain engineered to produce artemisinin or its precursor (amorphadiene) could ultimately reduce the

cost of the anti-malarial drug produced using this pathway. Similarly, by using the same methods described above, it should be possible to control eukaryotic protein synthesis from operons by constructing TIGRs containing IRES elements and RNase sites, the combination of which will be very useful for the design of operons for pathway engineering, multi-subunit protein production, and gene therapy.

References

1. **Khosla, C. & Keasling, J.D. Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov* 2, 1019-1025 (2003).**

2. **Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D. & Keasling, J.D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat Biotechnol* 21, 796-802 (2003).**

3. **White, M.M. Pretty subunits all in a row: using concatenated subunit constructs to force the expression of receptors with defined stoichiometry and spatial arrangement. *Molecular Pharmacology* 69, 407-410 (2006).**

4. **Endy, D. Foundations for engineering biology. *Nature* 438, 449-453 (2005).**

5. **Baga, M., Goransson, M., Normark, S. & Uhlin, B.E. Processed mRNA with differential stability in the regulation of E. coli pilin gene expression. *Cell* 52, 197-206 (1988).**

6. **Komar, A.A. & Hatzoglou, M. Internal ribosome entry sites in cellular mRNAs: mystery of their existence. *J. Biol. Chem.* 280, 23425-23428 (2005).**

7. **Martin, P., Albagli, O., Poggi, M.C., Boulukos, K.E. & Pognonec, P. Development of a new bicistronic retroviral vector with strong IRES activity. *BMC Biotechnology* 6, 4 (2006).**

8. **Chappell, S.A. & Mauro, V.P. The internal ribosome entry site (IRES) contained within the RNA-binding motif protein 3 (*Rbm3*) mRNA is composed of functionally distinct elements. *J. Biol. Chem.* 278, 33793-33800 (2003).**

9. **Fernandez-Miragall, O., Ramos, R., Ramajo, J. & Martinez-Salas, E. Evidence of reciprocal tertiary interactions between conserved motifs involved in organizing RNA structure essential for internal initiation of translation. *RNA* 12, 223-234 (2006).**

10. **Nudler, E. & Gottesman, M.E. Transcription termination and anti-termination in E. coli. *Genes Cells* 7, 755-768 (2002).**

11. **Mandal, M. & Breaker, R.R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* 5, 451-463 (2004).**

12. **Arraiano, C.M. & Maquat, L.E. Post-transcriptional control of gene expression: effectors of mRNA decay. *Mol Microbiol* 49, 267-276 (2003).**

13. **Kushner, S.R. mRNA decay in prokaryotes and eukaryotes: different approaches to a similar problem. *IUBMB Life* 56, 585-594 (2004).**

14. **de Smit, M.H. & van Duin, J. Control of translation by mRNA secondary structure in Escherichia coli. A quantitative analysis of literature data. *J Mol Biol* 244, 144-150 (1994).**

15. **Isaacs, F.J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* 22, 841-847 (2004).**

16. **Majdalani, N., Vanderpool, C.K. & Gottesman, S. Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol* 40, 93-113 (2005).**

17. **Smolke, C.D., Carrier, T.A. & Keasling, J.D. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Appl Environ Microbiol* 66, 5399-5405 (2000).**

18. **Smolke, C.D. & Keasling, J.D. Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. *Biotechnol Bioeng* 80, 762-776 (2002).**

19. **Deana, A. & Belasco, J.G. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev* 19, 2526-2533 (2005).**

20. **Matz, M.V. et al. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol* 17, 969-973 (1999).**

21. **Pfleger, B.F., Fawzi, N.J. & Keasling, J.D. Optimization of DsRed production in Escherichia coli: effect of ribosome binding site sequestration on translation efficiency. *Biotechnol Bioeng* 92, 553-558 (2005).**

22. **Naureckiene, S. & Uhlin, B.E. In vitro analysis of mRNA processing by RNase E in the pap operon of Escherichia coli. *Mol Microbiol* 21, 55-68 (1996).**

23. **Kaberdin, V.R. Probing the substrate specificity of Escherichia coli RNase E using a novel oligonucleotide-based assay. *Nucleic Acids Res* 31, 4710-4716 (2003).**

24. **Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415 (2003).**

25. **Spickler, C. & Mackie, G.A. Action of RNase II and polynucleotide phosphorylase against RNAs containing stem-loops of defined structure. *J Bacteriol* 182, 2422-2427 (2000).**

26. **Gottesman, S. The small RNA regulators of Escherichia coli: roles and mechanisms*. *Annu Rev Microbiol* 58, 303-328 (2004).**

27. **Burke, E. & Barik, S. Megaprimer PCR: application in mutagenesis and gene fusion. *Methods Mol Biol* 226, 525-532 (2003).**

28. Pfleger, B.F., Pitera, D.J., Newman, J.D., Martin, V.J. & Keasling, J.D. Microbial Sensors for Small Molecules: Development of a Mevalonate Biosensor. ***Appl Environ Microbiol*** **(Submitted).**

29. Sambrook, J., Fritsch, E.F. & Maniatis, T. Molecular Cloning. A Laboratory Manual. (Cold Spring Harbor Press, Cold Spring Harbor, NY; 1989).

30. Rodwell, V.W. et al. 3-Hydroxy-3-methylglutaryl-CoA reductase. ***Methods Enzymol*** **324, 259-280 (2000).**