Optimization of Heterologous Gene Expression for In Vitro Evolution

BioTechniques 30:474-476 (March 2001)

Proteins can be evolved in vitro by randomly mutating the corresponding genes, expressing the resulting libraries in a population of *E. coli* cells, and screening for clones that exhibit a desired phenotype (16). In our experience, the most time-consuming part of in vitro or "directed" evolution is the development of the high-throughput assay or "screen". In particular, the wild-type gene should initially be expressed so that the desired activity is at the lower limit of detection, allowing for maximal improvements in phenotype during the course of the experiment. Adjustment of gene expression to achieve this basal activity level is often labor intensive, requiring trial and error with multiple expression vectors, strains, and induction conditions.

For example, we attempted to engineer an expression vector that produces low levels of β-glucuronidase (GUS) for an in vitro evolution experiment. Standard expression vectors, based on the *lac* or T7 promoters, proved unsuitable because they yielded activities near the upper limit of the dynamic range of our assay (data not shown). We used standard recombinant DNA techniques to subclone the *Eco*RI-*Afl*III fragment of *gus*A-pBSΔ containing the *lac* promoter (*lacp*) and the β-glucuronidase gene (*gusA*) (10) into the lower copy number pET20 plasmid (Novagen, Madison, WI, USA). Unfortunately, this *lacp-gusA*-pET20 construct proved to be phenotypically identical to the higher copy number, parental *gusA*-pBSΔ plasmid (data not shown).

We considered modulating *gusA* expression by varying inducer concentrations and induction times, but had previously found that these measures provided limited control. Most transported inducers, including those that de-repress the *lac* and P_{BAD} promoters, yield "all-or-nothing" expression (7) and thus can lead to mixed populations of expressing and non-expressing cells within clonal cultures (2,7). This cell-to-cell variability in turn hampers screens and selections in which enzyme activity within individual cells is assayed (2,13). The growing popularity of these high-throughput screens (12) demands a more systematic and general approach to adjusting expression levels.

Rosenberg and Court (14) have shown that 75% of all mutations that affect promoter function fall within the -35 or -10 hexanucleotide regions upstream of the RNA start site. Mutations within the -35 region tend to greatly diminish transcription, whereas those in the -10 region can have mild to deleterious effects (18). Previous workers have randomized these conserved elements to learn about promoter structure and function (4,11). Jensen and Hammer (6) used a randomization approach to optimize the expression of a *Lactococcus lactis* gene for industrial fermentation. They generated a library of synthetic *L. lactis* promoters that contained conserved consensus elements flanked by 24 nucleotides of random sequence, isolated 38 clones, and found that gene expression levels among these clones varied over a 7000-fold range (5,6). While this approach might work for other promoters, it is technically easier to identify and manipulate core promoter sequences. Therefore, we chose to randomize the six conserved nucleotides in the -10 region of the lac promoter instead of 24 non-conserved ones. By limiting the size of the library (4^6 vs. 4^{24}), it proved possible to examine every combination of nucleotides using the same high-throughput assay developed for the in vitro evolution experiment.

The -10 promoter region was randomized by PCR-based mutagenesis, although any site-directed mutagenesis protocol (reviewed in Reference 9) would likely have worked. We replaced the *gusA* of *lacp-gusA*-pET20 with a 1-kb *Xba*I-*Eco*RI stuffer fragment using standard techniques. The whole *lacp*-stuffer-pET20 expression vector was amplified using Vent® DNA polymerase (New England Biolabs, Beverly, MA, USA) (1) and the primers 5′-NNNNNNCGAGCCGGAAGCATAAAGTGT-

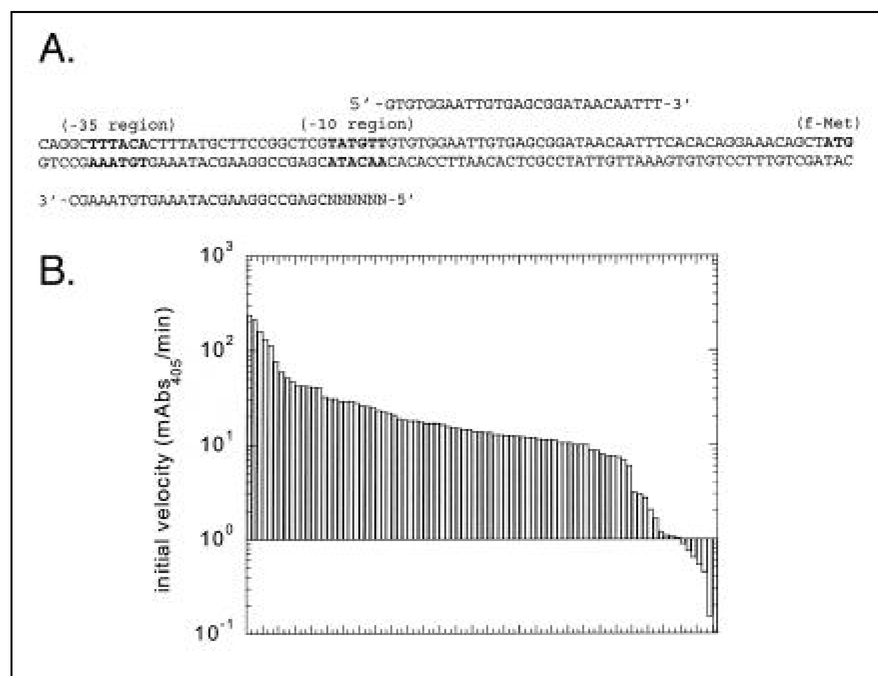

Figure 1. Distribution of GUS activities expressed by a library of randomly mutated promoters. (A.) The -10 region of the *lac* promoter was randomized by PCR-directed mutagenesis using the primers shown. (B) 88 *E. coli* colonies, harboring randomized *lac* promoters that drive the expression of *gusA* reporter genes, were propagated in liquid LB-ampicillin culture. Five microliters of each saturated culture were reacted separately with 1 mM para-nitrophenol β-D-glucuronide in buffer, and the rate of product formation of each reaction was determined using a 96-well spectrophotometer. Each of the 88 bars represents the initial velocity of one reaction, and they are presented left to right in order of descending reaction rate.

AAAGC-3′ and 5′-GTGTGGAATTG-TGAGCGGATAACAATTT-3′ (Figure 1A). The reaction components [1× thermo pol buffer (New England Biolabs), 500 nM primers, 200 µM dNTPs, 200 ng *lacp*-stuffer-pET20 template, and 1 U Vent DNA polymerase] were mixed in 50 µL total volume, and cycled 35 times between 94°C for 30 s and 72°C for 4 min. The end product of this amplification was a linear, blunt-ended plasmid in which each nucleotide of the -10 region (TATGTT) had been replaced with an equimolar mixture of G, A, T, and C. The PCR product was purified from a 0.8% agarose gel (QIAquick™ gel extraction kit; Qiagen, Valencia, CA, USA), phosphorylated at the 5′ ends with T4 polynucleotide kinase (New England Biolabs), and circularized with T4 DNA ligase (Life Technologies, Rockville, MD, USA) (8). The library was butanol precipitated (17) and electroporated (Bio-Rad Laboratories, Hercules, CA, USA) (3) into *E. coli* Inv*a*F′/pREP4, which constitutively expresses the *lac* repressor (10). The transformants were propagated in selective LB-ampicillin media, and the plasmids were purified using ion exchange chromatography (Qiagen).

We subcloned the *Xba*I-*Eco*RI fragment (containing the wild-type *gusA* gene) from *gusA*-pBSΔ (10) downstream of the randomized promoter, and transformed the library into Inv*a*F′ (Invitrogen, Carlsbad, CA, USA). This strain was derived from DH5α, which exhibits little endogenous *gusA* activity (15). The transformants were plated on LB-ampicillin agar plates containing 0.8% (w/v) 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-gluc), a histochemical marker of GUS activity. Inv*a*F′ is deficient in *lac* repressor activity, so expression of *gusA* was constitutive. After 24 h of incubation at 37°C, the colonies exhibited a range of color. Some were white, others were very blue and were surrounded by blue halos, and the rest were various shades of blue (data not shown).

To more quantitatively determine the range of GUS expression, we picked 88 colonies with sterile toothpicks from a "blind" LB-ampicillin plate that did not contain X-gluc, and inoculated 100-µL cultures in a sterile 96-well plate. The cultures were shaken for 48 h at 37°C until they were saturated (A_{600} = approximately 2); some cell lysis may also have occurred. We added 5 µL each culture to 1 mM para-nitrophenyl β-D-glucuronide in 200 µL buffer (50 mM sodium phosphate, pH 7.0, 1 mM EDTA, 5 mM β-mercaptoethanol). We monitored the changes in absorbance at 405 nm in a 96-well spectrophotometer (Bio-Tek® FL600; Bio-Tek, Winooski, VT, USA) at 23°C for 3 h and determined the initial velocity of each reaction (Figure 1B). Even though only a small fraction (88/4096) of the randomized -10 library was sampled, the values varied over three orders of magnitude. The expression construct containing the wild-type promoter was not genetically stable when constitutively expressed, so a reproducible value for its expression level could not be obtained (data not shown).

Constitutive expression from high copy number plasmids favors the propagation of non-expressing cells and tends to be unstable (7). To ascertain the stability of the transformants, several were picked and re-streaked onto LB-ampicillin plates with X-gluc. A few cultures yielded a mixture of white and blue colonies, suggesting that the plasmids were unstable (or possibly that the original picks were polyclonal). We chose two apparently stable clones, one white and one blue, for further propagation and analysis of individual cells via fluorescence-activated cell sorting. The saturated liquid cultures were reacted for 10 min at room temperature with 0.1 mM ELF® 97 β-D-glucuronide (Molecular Probes, Eugene, OR, USA), which has a precipitating fluorescent product. The cultures were centrifuged, resuspended in phosphate buffer, and analyzed using a Cytomation MoFlo® (Cytomation, Fort Collins, CO, USA) fluorescence-activated cell sorter with a 637-nm diode laser for light scatter and an Innova 90® (Coherant, Santa Clara, CA, USA) (350–360 nm) laser for excitation; fluorescence was measured using a 530/40 filter with a Hamamatsu H957-12 (Hamamatsu, Bridgewater, NJ, USA) set at 600 V. The cells in each population exhibited a uniform

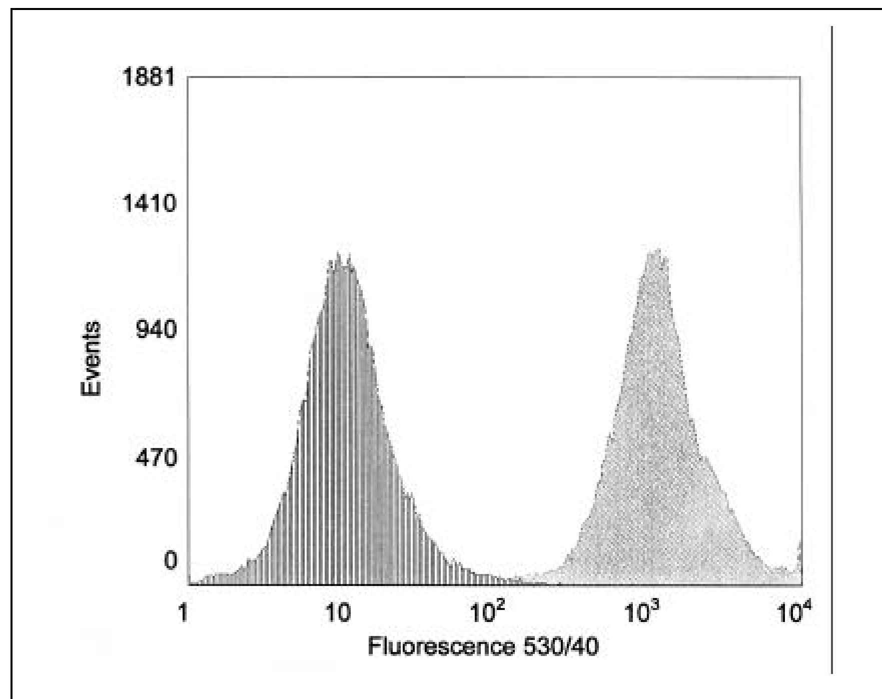

Figure 2. Expression of GUS activity by two clones from the library of randomly mutated promoters. *E. coli* transformed with *gusA* expression vectors with mutated *lac* promoters were plated on LB-ampicillin plates containing X-gluc. One white colony (vertical shading) and one blue colony (slanted shading) were separately propagated in liquid LB-ampicillin culture. One milliliter of each saturated culture was reacted with 0.1 mM ELF 97 β-D-glucuronide for 10 min at 23°C. The cells were washed with buffer and analyzed with a fluorescence-activated cell sorter.

level of fluorescence (Figure 2), confirming that the expression of the reporter enzyme was stable and that the selected constitutive promoter resulted in uniform expression levels. In contrast, fluorescence-activated cell sorting (FACS) analysis of an antibody expressed from the commercially available pBAD vector (Invitrogen) has shown that sub-saturating concentrations of arabinose leads to mixed populations of fully induced and fully uninduced cells (2).

The random mutation of the -10 region is a technically simple way to modulate heterologous gene expression levels for in vitro evolution experiments or other assays that require optimized expression levels, such as screens for pharmaceuticals. We chose to isolate plasmids that expressed GUS under constitutive expression conditions because such vectors could be incorporated into our high-throughput assay most conveniently. The principle virtue of this approach is that it could potentially be applied to any expression system regardless of the promoter, gene, plasmid copy number, host strain, growth media, or induction conditions. This approach will allow virtually any expression system to be adapted for a particular high-throughput assay.

REFERENCES

1. **Byrappa, S., D.K. Gavin, and K.C. Gupta.** 1995. A highly efficient procedure for site-specific mutagenesis of full-length plasmids using Vent DNA polymerase. Genome Res. *5*:404-407.
2. **Daugherty, P.S., M.J. Olsen, B.L. Iverson, and G. Georgiou.** 1999. Development of an optimized expression system for the screening of antibody libraries displayed on the *Escherichia coli* surface. Protein Eng. *12*:613-621.
3. **Dower, W.J., J.F. Miller, and C.W. Ragsdale.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. *16*:6127-6145.
4. **Horwitz, M.S. and L.A. Loeb.** 1986. Promoters selected from random DNA sequences. Proc. Natl. Acad. Sci. USA. *83*:7405-7409.
5. **Jensen, P.R. and K. Hammer.** 1998. Artificial promoters for metabolic optimization. Biotechnol. Bioeng. *58*:191-195.
6. **Jensen, P.R. and K. Hammer.** 1998. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl. Environ. Microbiol. *64*:82-87.
7. **Keasling, J.D.** 1999. Gene-expression tools for the metabolic engineering of bacteria. Trends Biotechnol. *17*:452-460.
8. **King, P.V. and R.W. Blakesley.** 1986. Optimizing DNA ligations for transformation. Focus (Gibco-BRL). *8*:1-3.
9. **Ling, M.M. and B.H. Robinson.** 1997. Approaches to DNA mutagenesis: an overview. Anal. Biochem. *254*:157-178.
10. **Matsumura, I., J.B. Wallingford, N.K. Surana, P.D. Vize, and A.D. Ellington.** 1999. Directed evolution of the surface chemistry of the reporter enzyme β-glucuronidase. Nat. Biotechnol. *17*:696-701.
11. **Oliphant, A.R. and K. Struhl.** 1988. Defining the consensus sequences of *E. coli* promoter elements by random selection. Nucleic Acids Res. *16*:7673-7683.
12. **Olsen, M., B. Iverson, and G. Georgiou.** 2000. High-throughput screening of enzyme libraries. Curr. Opin. Biotechnol. *11*:331-337.
13. **Olsen, M.J., D. Stephens, D. Griffiths, P. Daugherty, G. Georgiou, and B.L. Iverson.** 2000. Function-based isolation of novel enzymes from a large library. Nat. Biotechnol. *18*:1071-1074.
14. **Rosenberg, M. and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. *13*:319-353.
15. **Singh, M. and C. Sharma.** 1999. Detection of plasmid-encoded *gusA* gene in GUS-positive *Escherichia coli*. BioTechniques *26*:261-264.
16. **Sutherland, J.D.** 2000. Evolutionary optimisation of enzymes. Curr. Opin. Chem. Biol. *4*:263-269.
17. **Thomas, M.R.** 1994. Simple, effective cleanup of DNA ligation reactions prior to electro-transformation of *E. coli*. BioTechniques *16*:988-990.
18. **Youderian, P., S. Bouvier, and M.M. Susskind.** 1982. Sequence determinants of promoter activity. Cell *30*:843-853.

We thank Profs. Brent Iverson and George Georgiou for thoughtful advice, and for the use of their fluorescence-activated cell sorter. *Address correspondence to Andrew Ellington, Institute of Cellular and Molecular Biology, ICMB A4800/MBB 3.424, 26th and Speedway, University of Texas, Austin, TX 78712, USA. e-mail: andy.ellington@mail.utexas.edu*

Received 29 August 2000; accepted 1 December 2000.

Ichiro Matsumura, Mark J. Olsen, and Andrew D. Ellington
University of Texas Austin, TX, USA

Vectors for the Generation of FLAG®- or EGFP-Tagged cDNA Constructs and EGFP-Tagged Antisense RNA Constructs

BioTechniques 30:476-482 (March 2001)

Epitope or fusion protein tagging has become a common tool for the recombinant expression of proteins in mammalian cells. The addition of an N-terminal or C-terminal tag allows surveillance of the protein and facilitates identification of transfected cells and purification of the recombinant protein. Two commonly used tags are the FLAG® epitope (Sigma-Aldrich, St. Louis, MO, USA) and the enhanced green fluorescent protein (EGFP). The FLAG epitope consists of eight amino acid residues (DYKDDDDK). FLAG-tagged proteins can be detected with anti-FLAG antibodies in western blots and can be purified by immunoaffinity chromatography (4). EGFP is an optimized variant of the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* that emits bright green fluorescent light upon exposure to UV or blue light (2,3,5,8). EGFP-tagged protein can be detected directly in living cells without fixation or disruption. In addition, EGFP-tagged proteins can be detected in western blots or immunoprecipitated with anti-EGFP antibodies.

Eukaryotic expression vectors for the cloning of epitope-tagged cDNAs are available but are not always suited for direct cloning because they lack appropriate restriction sites or do not allow cloning of toxic genes. Certain toxic cDNAs can only be cloned and propagated in *E. coli* if leaky expression by unspecific transcription and translation is prevented (1). The use of low copy number vectors and/or vectors with several transcription termination signals can keep the nonspecific basal expression low and usually allows cloning and propagation in *E. coli.* Here, we describe the construction of vectors that allow cloning of highly toxic cDNAs in frame with an N-terminal FLAG or EGFP tag, or as EGFP-tagged antisense constructs. The cloned