Published online 15 September 2010 *Nucleic Acids Research, 2011, Vol. 39, No. 3* ***1131–1141***
doi:10.1093/nar/gkq810

Design, construction and characterization of a set of insulated bacterial promoters

Joseph H. Davis, Adam J. Rubin and Robert T. Sauer*

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received May 11, 2010; Revised August 27, 2010; Accepted September 1, 2010

ABSTRACT

We have generated a series of variable-strength, constitutive, bacterial promoters that act predictably in different sequence contexts, span two orders of magnitude in strength and contain convenient sites for cloning and the introduction of downstream open-reading frames. Importantly, their design insulates these promoters from the stimulatory or repressive effects of many 5′- or 3′-sequence elements. We show that different promoters from our library produce constant relative levels of two different proteins in multiple genetic contexts. This set of promoters should be a useful resource for the synthetic-biology community.

INTRODUCTION

The introduction of novel genetic components and pathways into cells has proven useful in biotechnology and as a tool to study and improve our understanding of natural systems (1–4). For some applications, achieving the proper steady-state levels of each gene product can be critical in optimizing the function of an entire biosynthetic pathway, whereas, in other cases, assaying the consequences of altered expression levels is important for probing native gene function (5–7).

In principle, steady-state protein levels can be controlled by using libraries of variable-strength promoters to change transcription rates, by employing different ribosome-binding sites to alter translation efficiency, and by appending degradation tags to adjust rates of protein turnover (5,8,9). Often, the process of finely tuning each of these parameters is laborious and relies on trial-and-error, a problem that has led some to utilize directed-evolution to guide the optimization process (10–12). Others have begun to characterize individual genetic parts rigorously (e.g. the transcriptional strength of a promoter) with the hope that such information might guide and expedite the refinement process (13–15). The success of reusing well-characterized components relies on a critical assumption that such devices are functionally composable, that is the properties of the device in one test context are predictive of those properties in a new context.

Building on the work of others, we sought to design a set of variable-strength constitutive bacterial promoters that are insulated from influences of genomic context. Although, it may not be possible to insulate any biological component completely, a wealth of information is available that can be exploited to limit context-dependent behavior. Bacterial transcription can be decomposed into three phases; binding, initiation and elongation. Extensive biochemical and structural studies have helped elucidate the promoter and polymerase components that control each of these steps (16). Bacterial RNA polymerase (RNAP) is composed of a core polymerase ($\beta\beta'\alpha_2\omega$), which is competent for transcriptional elongation, and a σ subunit, which is utilized to define promoter specificity during binding and initiation but is dispensable for elongation. In the initial binding step, the σ subunit contacts two hexameric DNA sequences located 10 and 35 base pairs 5′ of the transcription start site (named the −10 and −35 boxes, respectively) (17). At some promoters, additional contacts are formed between the α_2 subunits and A/T-rich promoter sequences residing as many as 60 bp 5′ of the transcription-start site (known as an 'UP' sequence). These contacts facilitate polymerase binding and can enhance promoter activity up to 300-fold in a manner that depends upon the sequence distance from the core recognition elements (18–20). Once bound, the enzyme–promoter complex must isomerize from a 'closed' complex in which the DNA is double stranded to an 'open' complex in which base pairs from approximate positions −10 to +2 melt or separate into single strands (17). The conformational equilibrium between closed and open complexes depends, in part, on the sequence of the base pairs that melt (21). From the open complex, the polymerase can undergo a repetitive process, termed abortive transcription, in which it initiates transcription, releases a short RNA transcript (<10 nt), and

*To whom correspondence should be addressed. Tel: +1 617 253 3163; Fax: +1 617 258 0673; Email: bobsauer@mit.edu

then returns to the site of initiation. This process continues in a stochastic fashion until the polymerase clears the promoter, releases the σ subunit, and continues the elongation phase of transcription (often referred to as promoter escape). The identity of the 20-nt downstream of the transcription-start site (defined as +1 to +20) strongly influences the efficiency with which a bacterial polymerase escapes from the promoter and continues elongation (22,23).

To respond to external signals, transcription factors have evolved mechanisms to up-regulate or down-regulate each of the steps described earlier. For example, cyclic-AMP receptor protein (CRP) binds the promoter upstream of the −35 box, simultaneously making favorable contacts with the C-terminal domain of the polymerase α subunits (24,25). These contacts enhance polymerase binding and increase transcription at promoters where binding is rate limiting (26). Promoters can also be regulated at the isomerization step of transcription. For example, the merR transcriptional repressor acts by inhibiting conversion of the closed to open complex (27). Lastly, the efficiency of promoter escape can be regulated. At some promoters, for example, the phage φ29 protein p4 binds tightly to the polymerase α subunit, which slows promoter escape and downregulates productive transcription (27). Interestingly, at promoters where polymerase binding is limiting, this same protein can activate transcription by enhancing binding (28). Such non-uniform effects are particularly important when considering the design of synthetic variable-strength promoter libraries.

Many synthetic promoters have been generated and characterized, but they often show activities that vary with the genetic locus or gene transcribed (5,14,29–32). As discussed earlier, this context dependence is not surprising. For example, increased activity can arise from 5′-UP-sequence elements or as a consequence of read-through from upstream promoters (20,33). One solution to the latter problem is to include a transcriptional terminator at the 5′-boundary of the synthetic promoter, but AT-rich terminators share some sequence similarity with UP elements and may themselves increase transcription of downstream genes. Promoter fusions to different genes may also affect transcription efficiency if the 5′-end of the mRNA contains a sequence that changes the rate of promoter escape.

The strength of any minimal promoter, containing sequences from the −35 hexamer to the site of initiation, is likely to vary depending on neighboring sequences. As a consequence, we generated promoters in which adjoining upstream and downstream sequences, which potentially could alter transcription initiation and promoter escape, were included in the promoter cassette (Figure 1). These 'insulated' promoter cassettes extend

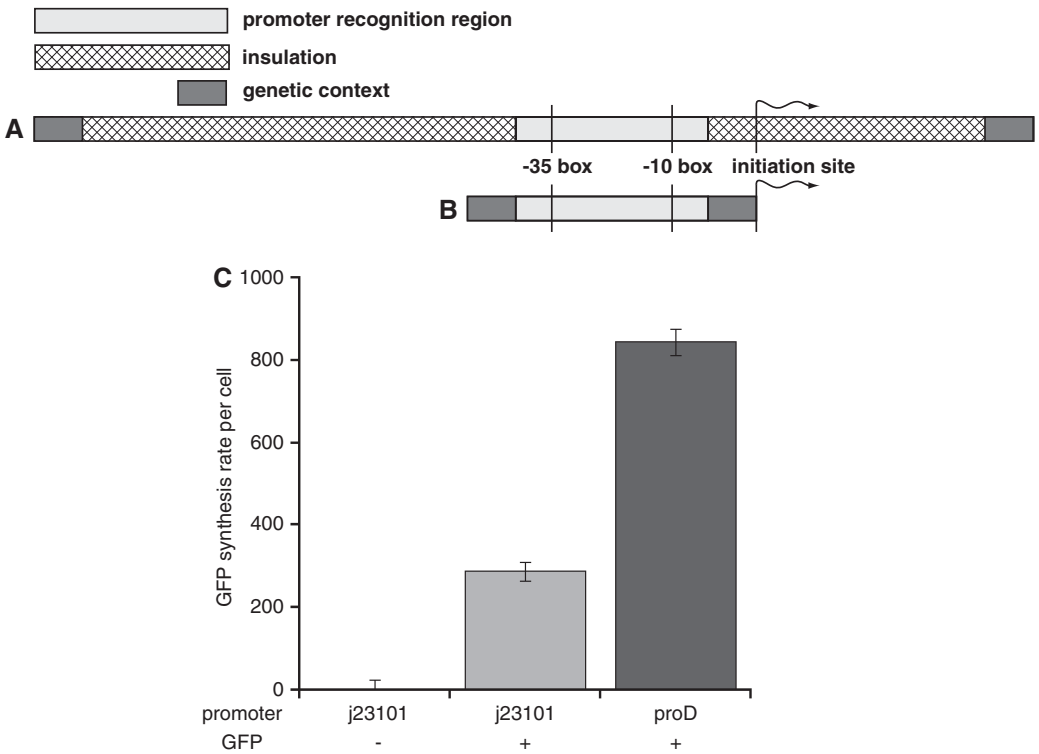

Figure 1. Comparison of promoter organization. Schematic of an insulated promoter (**A**) and a minimal uninsulated promoter (**B**). The promoter recognition region (PRR) containing the −10 and −35 RNAP binding determinants is shown in light gray, the transcription-initiation site (+1) is represented by arrow. In the insulated promoter, the surrounding genetic context (dark gray) is separated from the PRR by insulation sequences (diamond-filled pattern). Most elements known to effect transcription initiation and promoter escape are contained within the insulated promoter cassette boundaries. Because of its smaller size, the genetic context surrounding the minimal promoter is more likely to contain sequences that can effect transcription initiation, thereby increasing the possibility of context-dependent activity. (**C**) GFP synthesis rates per cell were measured for a control construct lacking GFP, a minimal promoter (j23101) or an insulated promoter (proD).

from position −105 to +55, a span which includes the majority of transcription factor-binding sites in natural bacterial promoters as well as most elements that affect transcription initiation and promoter escape (34). We find that such promoters exhibit transcriptional activities that are significantly more predictable in varied genetic contexts.

MATERIALS AND METHODS

Promoter sequences

The promoter sequences used in this work are listed below. The scar sequences generated by standard BioBrick assemblies are in lowercase font, and the expected start site is underlined. For each promoter set (insulted or minimal), only the −35 and −10 hexamers (shown in large, bold font) vary between library members (Table 1).

proD (insulated promoter)
ttctagagCACAGCTAACACCACGTCGTCCCTATCTG
CTGCCCTAGGTCTATGAGTGGTTGCTGGATA
AC**TTTACG**GGCATGCATAAGGCTCG**TATAAT**A
TATTCAGGGAGACCACAACGGTTTCCCTCTAC
AAATAATTTTGTTTAACTTTtactagag

j23101 (minimal promoter)
ttctagag**TTTACA**GCTAGCTCAGTCCTAGG**TATAAT**
GCTAGCtactagag

Plasmids and strains

Experiments with plasmid-borne reporters were performed in *Escherichia coli* strain DH5α [F^-, λ^-, φ80*lac*ZΔM15, Δ(*lac*ZYA-*arg*F)U169, *deo*R, *rec*A1, *end*A1, *hsd*R17(rk^-, mk^+), *pho*A, *sup*E44, *thi*-1, *gyr*A96, *rel*A1], whereas experiments with chromosomally encoded reporters were performed in *E. coli* strain W3110 [F^-, λ^-, IN(*rrn*D-*rrn*E), *rph*-1]. With the exception of assays for promoter function on the chromosome, all experiments utilized plasmid pSB3C5 and the construct of interest was cloned via standard procedures between the BioBrick cloning sites (35). Each construct contained the Bba_B0032 ribosome binding site and the Bba_B0015 transcriptional terminator. The sequences for the fluorescent reporter proteins, GFP (Bba_E0040), dsRed (Bba_E1010), Gemini (Bba_E0051), the ribosome binding site (Bba_B0032) and the terminator (Bba_B0015) can be found at the Registry of Standard Biological Parts (www.partsregistry.org).

Plasmid constructs contained the following elements: promoter-TACTAGAG-B0032-TACTAG-ORF(dsRed, GFP, Gemini)-TACTAGAG-B0015, where the ORF was exchanged using standard PCR-based techniques. For chromosomal insertions, test constructs were fused to a kanamycin resistance marker (Bba_P1003, Registry of Standard Biological Parts, www.partsregistry.org) using SOEing PCR (36). PCR products were recombined onto the chromosome using the λ-red recombination system, encoded on the plasmid vector pSIM5, as described earlier (37). After verification of successful cassette insertion by sequencing, pSIM5 was cured from the strain. For the *tonB* locus, the SOEing primers used are as follows (with homology to the locus listed in bold).

tonB-BioBrickPrefix-fwd
AAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTT
TTGACTgaattcgcggccgcttctag

BioBrickSuffix-rev
cgaacttttgctgagttgaaggatcagCTGCAGCGGCCGCTACT
AGTA

BioBrickSuffix-fwd
TACTAGTAGCGGCCGCTGCAGctgatccttcaactcagcaa
aagttcg

P1003-*tonB*-rev
GATCCTGAAGGAAAACCTCGCGCCTTACCTGTTG
AGTAATttattagaaaaactcatcga

The UP sequence (GAGAAAATTATTTTAAATTTCCT C) was introduced upstream of the promoter constructs using standard techniques resulting in a BioBrick scar (AC TAGA) between the UP sequence and the promoter. The anti-sequence (ATCCGGAATCCTCTGGATCCTC) was introduced in a similar fashion resulting in constructs of the form: promoter-TACTAGAG-anti-B0032-TACTAG -GFP-TACTAGAG-B0015.

Two strains were used to control for cellular autofluorescence. DH5α transformed with pSB3C5 was used as a negative control for experiments using plasmid-based constructs. For experiments testing promoter function from the *tonB* locus, the kanamycin resistance marker with no reporter construct was recombined downstream of the *tonB* locus using primers *tonB*-BioBrickPrefix-fwd, P1003-*tonB*-rev.

Table 1. Promoter strengths

Promoter	−35 hexamer	−10 hexamer	RPUD
proA	tttacg	taggct	0.030
proB	tttacg	taatat	0.119
proC	tttacg	tatgat	0.278
proD	tttacg	tataat	1.000
pro1	tttacg	gtatct	0.009
pro2	gcggtg	tataat	0.017
pro3	tttacg	gaggat	0.017
pro4	tttacg	gatgat	0.033
pro5	tttacg	taggat	0.050
pro6	tttacg	taaaat	0.193
j23113	ctgatg	gattat	*0.005*•α
j23150	tttacg	tattat	*0.077*•α
j23151	ttgatg	acaatg	*0.192*•α
j23101	tttaca	tattat	*0.345*•α

The −10 and −35 hexamers for insulated (pro series) and uninsulated (j series) promoters are listed. RPU_D were calculated by normalizing the GFP-synthesis rate to that of the ProD promoter (see text). Minimal promoters are listed in italics. The α-scaling factor accounts for the fact that the transcripts for the two promoter sets are different, which could cause differences in degradation or translation rates.

Promoter activity assays

All GFP-, dsRed- and gemini-based assays were performed and analyzed as described (14) with the minor modifications listed below. Individual test colonies along with a negative control strain were picked in triplicate and grown overnight in 5-cm culture tubes in LB broth supplemented with 35 μg/ml chloramphenicol (for experiments using pSB3C5) or 10 μg/ml kanamycin (for experiments using chromosomal insertions). Cultures were then diluted 100-fold into M9 media (M9 salts, 1 mM thiamine hydrochloride, 0.2% casamino acids, 2 mM $MgSO_4$, 0.1 mM CaCl2, 0.4% glycerol) supplemented with appropriate antibiotics and grown at 37°C for 4 h. Cultures were aliquoted (150 μl) into a 96-well plate (Greiner Bio-One) in which OD (600 nm) and fluorescence (GFP, Gemini: excitation 467 nm, emission 511 nm; dsRed: excitation 560 nm, emission 590 nm) were read using a SpectraMax M5 fluorescence plate reader (Molecular Devices). Cultures were allowed to continue growing in tubes for an additional 1.25 h at 37°C at which time OD_{600} and fluorescence were read again. For each sample, the change in fluorescence signal between the two readings was divided by the average OD_{600}. This measure of promoter activity (per cell synthesis rate) was corrected for background auto-fluorescence by subtracting the per cell synthesis rate of the negative control. For most experiments, the corrected synthesis rate was normalized to the average synthesis rate of the proD promoter, resulting in RPU_D-relative promoter units (14). In experiments using dsRed, proD displayed diminished activity and the average synthesis rate of the j23101 promoter was used to normalize activity, resulting in RPU_{j12101}-RPUs. Errors bars shown in all figures represent the standard deviation of triplicate measurements.

RESULTS

Basic promoter design

Our initial construct was based on *E. coli* rrnB P1, a strong σ^{70}-dependent promoter with near consensus −10 (TATAAT) and −35 (TTtACg) elements. A 17-bp sequence (GGCATGCATAAGGCTCG) separates the −10 and −35 boxes, resulting in the optimal spacing for near-consensus promoters (38–40). To provide insulation, we also defined the flanking sequences from position −105 to +55, using elements described in 'Materials and Methods' section. Our promoter design extends beyond the transcription initiation start site and thus creates a specific and invariant 5′-mRNA terminus. This feature was incorporated to improve the predictability of promoter strength, mRNA stability, and the site of transcriptional initiation, but it may preclude experiments in which this region of the transcript must be of a particular sequence (41). We note, however, that the ribosome binding site and translation start codon reside downstream of this element, and thus the resulting protein product is not affected by the insulation sequence. It is also interesting to note that the vast majority of natural transcripts in *E. coli* are predicted to encode a 5′-untranslated region >20 nt in length (34).

We transformed cells with a plasmid vector bearing our first-generation insulated promoter (called proD) driving production of a GFP reporter gene (Bba_E0040, Registry of Standard Biological Parts, www.partsregistry.org) and measured GFP synthesis rates as a proxy for promoter strength. Briefly, cells were grown in culture tubes at 37°C to mid-log phase, OD_{600} was measured as a surrogate of cell number, and GFP fluorescence was determined. After an additional 1.25 h of growth, we again measured OD_{600} and GFP fluorescence. The GFP synthesis rate for the promoter was calculated using Equation 1 (14).

$$\text{Synthesis rate}_x = \frac{\text{GFP}(x)_{tp2} - \text{GFP}(x)_{tp1}}{\text{OD}_{600}(x)_{\text{average}}} \quad (1)$$

To determine the strength of proD relative to another promoter, we performed the same procedure using a minimal length, constitutive promoter (Bba_j23101: Registry of Standard Biological Parts), which has been previously characterized (14). In this comparison, the proD construct exhibited a greater GFP synthesis rate than the Bba_j23101 construct (Figure 1C).

Generation and characterization of an insulated promoter library

Next, we generated a promoter library by using degenerate oligonucleotides to randomize either the −35 or the −10 element of our insulated promoter in the plasmid vector and transformed *E. coli* cells. This library was enriched for active promoters using fluorescence activated cell sorting (data not shown). Measurements of GFP synthesis rates from individual colonies of the enriched pool were then used to identify 10 clones that exhibited varied GFP expression. Measurement of GFP synthesis rates showed that this set of variants encoded promoter strengths spanning two orders of magnitude. To aid in comparison with previously characterized minimal promoters, we also measured the activity of a set of uninsulated promoters that spanned a similar range of activity (Bba_j23113, Bba_j23150, Bba_j23151 and Bba_j23101 from The Registry of Standard Biological Parts, www.partsregistry.org; Figure 2).

As described in Kelly *et al.* (14), relative measures of promoter activity can greatly reduce assay-to-assay variance and are straightforward to calculate for a set of promoters that produce identical transcripts. We therefore determined relative promoter strength by normalizing the GFP-synthesis rate of each newly isolated promoter to that of the insulated proD promoter, as shown in Equation 2. To calculate RPUs of the uninsulated promoters, a scaling factor, α, was introduced. This constant, which accounts for potential differences in the degradation rates or translation rates of transcripts produced by the insulated and uninsulated promoter sets, cancels out when promoters within a set are compared and thus is only relevant when comparing promoters between the two sets. The sequences of the −35 and −10 hexamers of each promoter variant as

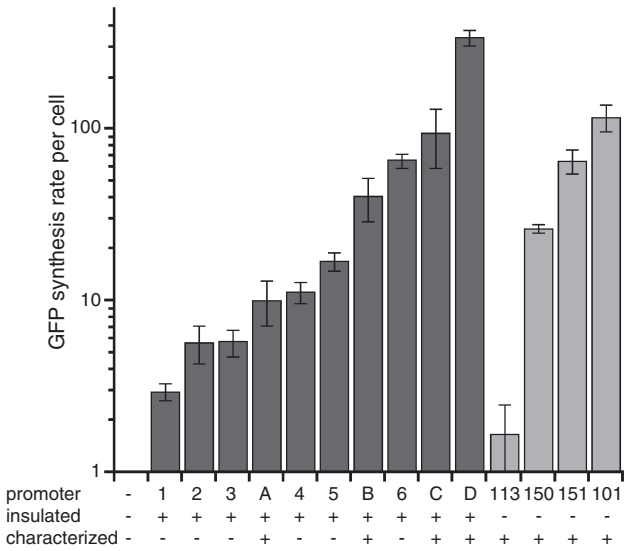

Figure 2. Promoter strength. The strength of each promoter was measured in triplicate in *E. coli* DH5α grown in minimal media using GFP as a reporter (14). The GFP-synthesis rate is reported on a $\log_{10}$ scale as a surrogate of promoter strength. Promoter-strength measurements relative to proD are listed in Table 1.

well as their relative promoter strengths are listed in Table 1.

$$\mathrm{RPU}_D(x) = \frac{\text{Synthesis rate}_x}{\text{Synthesis rate}_D} = \left(\frac{\mathrm{GFP}(x)_{tp2} - \mathrm{GFP}(x)_{tp1}}{\mathrm{OD}_{600}(x)_{\text{average}}} \Big/ \frac{\mathrm{GFP}(D)_{tp2} - \mathrm{GFP}(D)_{tp1}}{\mathrm{OD}_{600}(D)_{\text{average}}}\right) \quad (2)$$

From our set of 10 insulated variants, we selected proD and three additional promoters (proA, proB and proC), which spanned the activity range, for more detailed characterization (Figure 2). These promoters were named in ascending order of activity.

Characterization of promoter insulation

The UP element from *rrnB* P1 was chosen to test the efficacy of the 5′-insulation. This 24-nt sequence (AGAA AATTATTTTAAATTTCCTCA) has been shown to activate transcription from some promoters (33). We inserted the UP element at the 5′-boundaries of the insulated promoter cassettes (proA, proB, proC, proD) and the uninsulated promoter cassettes (j23113, j23101, j23150, j23151). Using the GFP-reporter assay, we determined the relative strength of each promoter either with or without the UP element (Figure 3A). Introduction of the UP sequence slightly reduced transcription from each insulated promoter compared to the same promoter with no UP element, whereas it increased transcription from the uninsulated promoters in a highly variable manner (Figure 3A, inset).

Next, we tested downstream insulation by inserting an anti sequence (ATCCGGAATCCTCTGGATCCTC) at the 3′-boundaries of insulated and uninsulated promoters (Figure 3B). This portable sequence decreases the rate of promoter escape when present at positions +1 to +22 of many transcripts (22). For each set of promoters, this sequence was inserted using the available restriction sites downstream of the promoter element. This strategy resulted in the same scar between the promoter and the downstream sequence that would be present if the promoter were used to drive the production of a new transcript (see 'Materials and Methods' section for a description of scar sequences). The position of the 'anti' sequence was from +47 to +69 for the insulated promoters and from +7 to +31 for the uninsulated promoters. Again, relative GFP synthesis rates were measured as a surrogate for promoter activity. As shown in Figure 3B, insertion of the anti sequence had almost no effect on the insulated promoters and had variable effects on the uninsulated promoters. The strongest promoter, j23101 was down regulated ~2-fold whereas the weaker promoter, j23150, showed no change. Interpreting activity differences between promoters with and without the anti sequence is difficult because the insertions alter the mRNA and thus could affect mRNA stability and translation efficiency in addition to promoter activity. We note, however, that for each set of insulated or uninsulated promoters, the mRNA transcribed is independent of the particular promoter assayed and thus is constant within that set. When normalized for relative promoter activity as shown in Figure 3B, it is clear that the insulated promoters are resistant to the effects of inserting this sequence.

Promoter activity when driving disparate open reading frames

Is promoter strength predictable following fusion to different open-reading frames? To address this question, we fused the insulated and uninsulated promoters to sequences encoding two additional fluorescent reporters. The first was Gemini, which contains an N-terminal LacZα sequence and a C-terminal GFP domain (32). The second was dsRed, which is an engineered fluorescent protein with little homology to GFP (42). Although it is difficult to directly compare expression levels of GFP with those of dsRed or Gemini in a meaningful way, the relative strengths of different promoters driving expression of each type of protein should be predictable if there is sufficient insulation from the effects of the initially transcribed sequence.

We first measured the synthesis rate of Gemini for each promoter (proA, proB, proC, proD, j23113, j23150, j23151, j23101) in a manner similar to that described for GFP. Gemini could be readily assayed using fluorescence from its GFP domain. Importantly, the N-terminus of Gemini (LacZα) is different from that of GFP, and thus one might expect differences in transcription from uninsulated promoters. As reported earlier, we observed a decrease in the absolute synthesis rates of Gemini relative to those of GFP (32; Figure 4). For each promoter, we calculated the relative promoter strength compared to proD driving the same open reading frame using Equation (2). Figure 4A plots the relative promoter

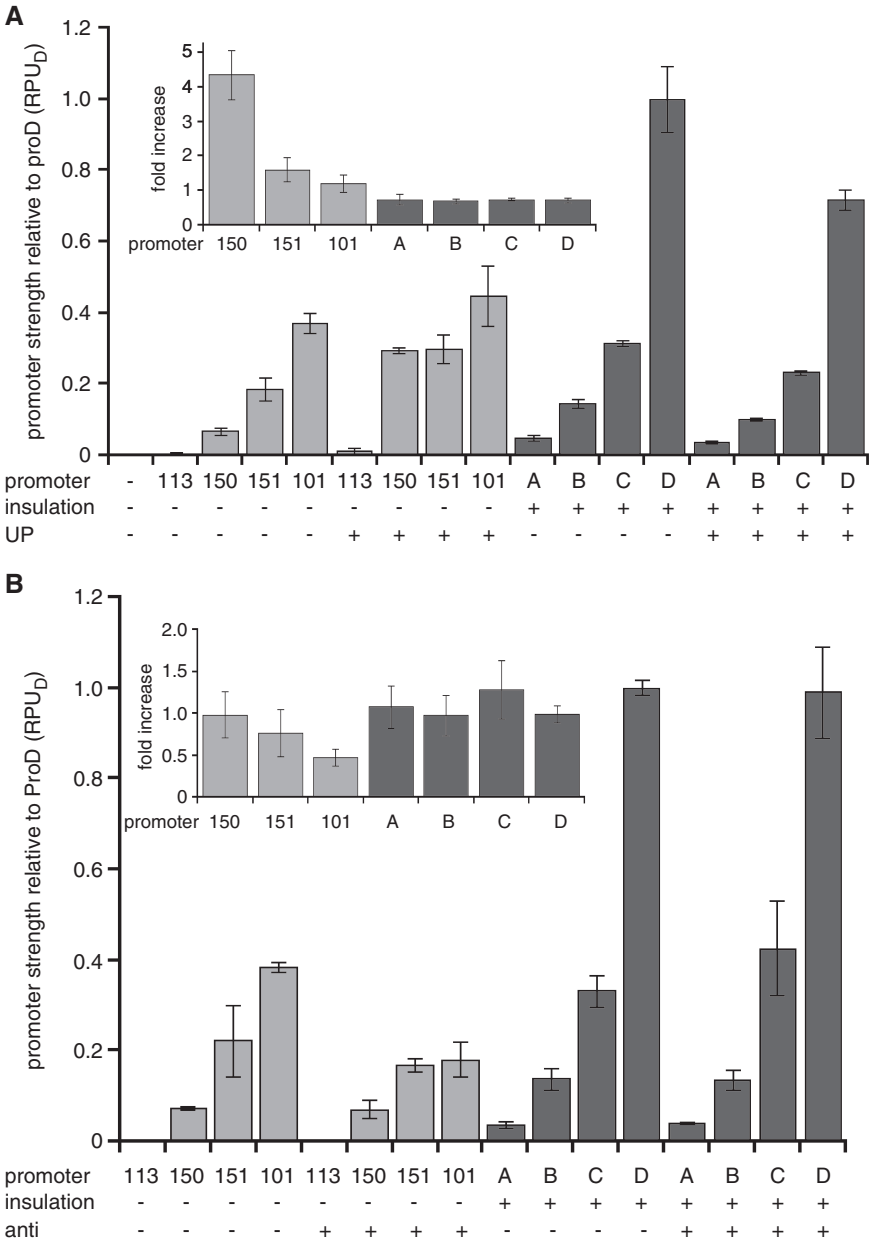

Figure 3. Effect of sequences inserted 5′ or 3′ of the promoter. (**A**) The UP sequence was cloned upstream of either insulated (proA, proB, proC, proD) or uninsulated (j23113, j23150, j23151, j23101) promoters and promoter strength was measured using the GFP-reporter assay. Promoter strength was normalized to the strength of proD resulting in RPU. The inset shows relative strength of each promoter with the UP sequence normalized to the strength of the parental promoter. A value of 1 indicates no change in promoter strength. The promoter j23113 was excluded from comparative analysis due to its weak promoter strength and relatively large colony-to-colony variation. (**B**) The 'anti' sequence was cloned downstream of either insulated or uninsulated promoters and the promoter strength of each construct was measured using the GFP-reporter assay. For each promoter that could be robustly measured, the inset reports the ratio of promoter strength with the anti sequence to that without the insertion.

strength (RPU_D) determined using GFP (x-axis) against that determined using Gemini (y-axis). Promoters not altered by the introduction of Gemini, approximately fall on the line $y = x$. Interestingly, the insulated promoters show a ~1:1 relationship between relative activity measured either by GFP or by Gemini. In contrast, the stronger uninsulated promoters (j23151, j23101) show diminished apparent relative activity when driving production of Gemini (Figure 4A, right). We do not believe that the decreased activity of the uninsulated promoters occurs as a consequence of translation or Gemini folding becoming rate limiting, because the absolute activity of even the strongest uninsulated promoter is weaker than that of the strongest insulated promoter (proD). The fact that neither proC nor proD show diminished activity when driving Gemini production argues that transcription from the strongest uninsulated promoters becomes limiting in this assay. Again, the uninsulated promoters show greater context-dependent activity.

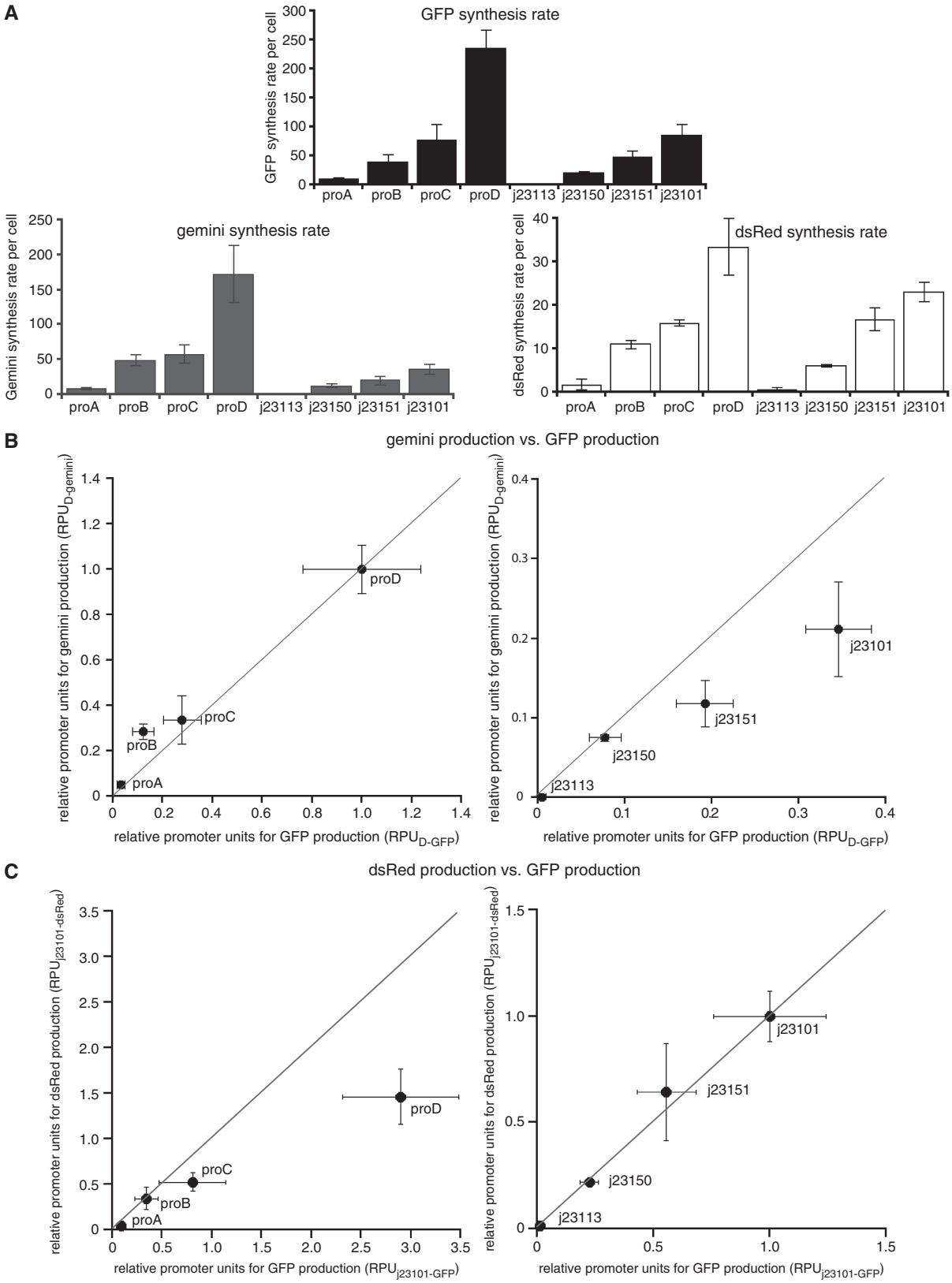

Figure 4. Aparent promoter activities driving production of GFP versus Gemini or dsRed. (**A**) Protein-synthesis rates for GFP (top), Gemini (left) or dsRed (right) were determined by measuring fluorescence and using Equation (1). (**B**) Relative promoter strength was calculated from the protein-synthesis rates for a set of insulated (left) or uninsulated (right) promoters driving the production of GFP (*x*-axis) or Gemini (*y*-axis). To allow comparison between open-reading frames, each synthesis rate was normalized to that of proD driving production of the same open-reading frame [Equation (2)]. (**C**) Promoter strength was measured as described above for either GFP (*x*-axis) or dsRed (*y*-axis). Because production of dsRed from proD was decreased, each synthesis rate was normalized to that of j23101 driving production of the same open reading frame.

We next measured the synthesis rate of dsRed for proA, proB, proC, proD, j23113, j23150, j23151 and j23101 promoters in a manner similar to that described for Gemini (see 'Materials and Methods' section for details). For each promoter driving production of each open reading frame, we calculated promoter strength relative to j23101 driving production of the same open reading frame. As described earlier, this analysis allows for comparison between promoter strength determined using either GFP or dsRed. For promoters with weak activity, we observed a 1:1 correspondence between apparent promoter strength measured using GFP and dsRed. For the strongest promoter tested, proD, we saw decreased relative activity when driving production of dsRed (Figure 4C). As described earlier, this could result from altered transcription indicating the insulation is insufficient or could arise because a process otherthan - transcription becomes limiting for this elevated level of dsRed production. Interestingly, for all of the uninsulated promoters tested, the relative promoter strength was maintained when driving production of dsRed.

Promoter activity from a chromosomal locus

In the experiments described so far, the promoters were carried on the medium copy number plasmid, pSB3C5, which bears a p15A origin of replication (35). To further investigate the efficacy and predictability of these promoters, we moved the insulated proA, proB, proC, proD promoters, the uninsulated j23113, j23150, j23151, j23101 promoters, and a non-fluorescent control cassette to a chromosomal locus. The promoter-RBS-GFP-terminator construct was fused to a kanamycin resistance marker using PCR and the entire cassette was site-specifically recombined at the chromosomal *tonB* locus using recombineering techniques ('Materials and Methods' section) (37). This site was chosen because of the presence of an apparent AT-rich terminator at the 3′-end of the *tonB* gene, which we expected might influence nearby promoter activity.

Promoter activities from the *tonB* locus were measured using the fluorescent signal from the encoded GFP. Figure 5A shows the synthesis rates for each promoter. As expected, the absolute activity of each promoter was decreased when placed on the chromosome, presumably a result of decreased copy number. To allow for comparison of promoter strength between these two loci, the synthesis rate was converted to RPUs by normalization to the synthesis rate of proD [Equation (2)]. As described earlier, this normalization masks the effects of copy number (14). For each promoter, Figure 5B shows the relative promoter strength measured from the plasmid versus the relative promoter strength measured from the chromosome. Interestingly, both sets of promoters exhibited an ~1:1 correspondence in relative promoter activity at the chromosomal versus plasmid loci.

DISCUSSION

We have generated a library of constitutive bacterial promoters with activities that span two orders of magnitude. These promoters contain sequences extending beyond the core-polymerase binding region in both the 5′- and 3′- directions. By testing these promoters in a variety of sequence contexts, both chromosomal and plasmid-based, we have demonstrated that their activity is highly predictable with minimal effects from the surround genetic context. As such, the promoter's activity measured in one context should be predictive of their function in a new context. Such functional composition should facilitate the engineering of biological systems.

Our promoters are 160 bp in length, substantially larger than the minimal 51-bp promoters we have used for comparison. Although our promoters were less affected by the stimulatory and repressive effects of many sequence elements, it should be noted that some long-range regulation by transcription-factors might still affect these promoters. Given that such regulation is probably unavoidable with any amount of insulation, we compromised on a promoter size that shows improved context-independent behavior but remained small enough to be easily incorporated into genetic pathways.

To decrease the possibility of the initially transcribed sequence altering promoter activity, our promoters include insulation extending beyond the transcriptional start site. This design results in several desirable features. First, unlike a minimal promoter, the transcription initiation site is clearly defined and invariant for the library irrespective of the downstream gene. Second, the defined 5′-untranslated sequence may facilitate the measurement and prediction of mRNA degradation rates for transcripts generated by these promoters. Lastly, because some prediction algorithms for the strength of ribosome-binding sites make use of surrounding sequence information, the invariant 5′-termini could improve the prediction of translational initiation rates (9).

Although the 5′-termini encoded by our promoters have no effect on the resulting protein product, it will prohibit some applications. For example, natural and synthetic riboregulators located in the 5′-untranslated region allow for small-molecule control of translation (43,44). The initially transcribed sequence generated by our promoters may be incompatible with such devices. For most applications, however, this 5′-sequence should pose no problem, as it affects neither the ribosome-binding site nor the resultant protein product.

We find that our insulated promoters are not perturbed by the introduction of stimulatory UP sequences 5′ of the promoter, or repressive 'anti' sequences 3′ of the promoter. The UP sequence likely has no effect on these promoters because the extended insulation sequence precludes the polymerase from simultaneously forming favorable contacts with the core promoter and the UP sequence. In contrast, we observe strong activation of the weak minimal promoters by the UP sequence, suggesting that this sequence improves polymerase recruitment. For the strongest minimal promoter, the UP element had no effect and thus some process after polymerase recruitment is probably rate limiting. Such non-uniform effects for the minimal promoters limits their predictably in different sequence contexts.

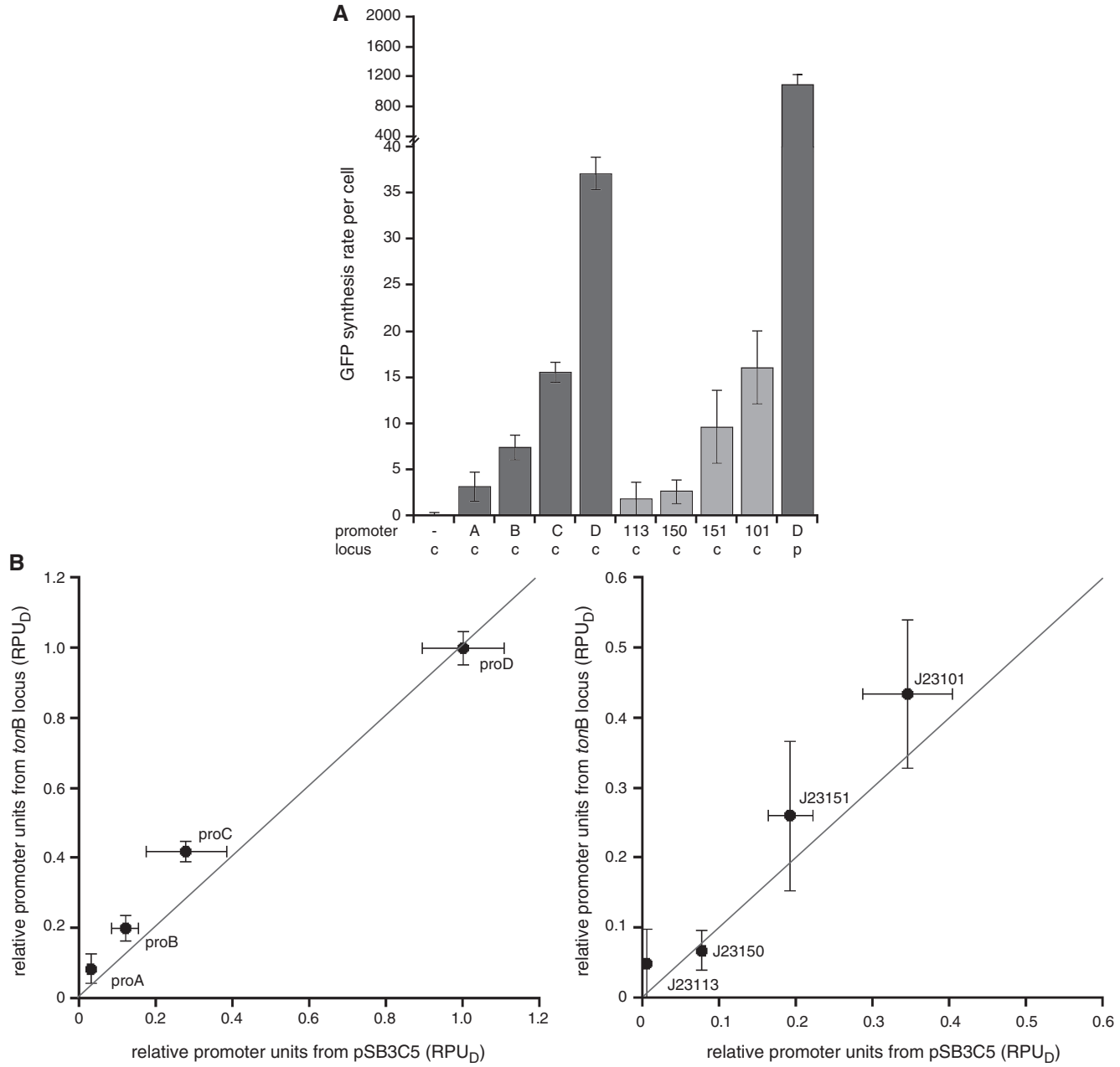

Figure 5. Promoter activity from a chromosomal locus. Promoters driving the expression of GFP were recombined onto the chromosome downstream of the *tonB* locus, and the activity of the promoter was measured using GFP fluorescence. (**A**) GFP synthesis rate per cell. (**B**) Plots correlating the relative promoter activity either from the plasmid, pSB3C5 or the chromosome (at the *tonB* locus) for the insulated promoters (left) or the uninsulted promoters (right).

The anti sequence also affects the minimal promoters non-uniformly, down regulating the strongest promoter 2-fold but not affecting other promoters. For these unaffected promoters, a process other than promoter escape (e.g. polymerase binding) is likely to be rate limiting. Interestingly, none of the insulated promoters are affected by the anti sequence, which causes transcriptional repression in a σ^{70}-dependent fashion (22). It seems likely that the insulated promoters are unaffected, because the σ subunit dissociates from the core polymerase before encountering the anti sequence.

In one instance, we found that the insulated promoters predictably drive production of a different gene product (Gemini), whereas the minimal promoters did not. Again, it appears that for the strongest uninsulated promoters, the initially transcribed sequence strongly affects promoter activity. The fact that even the strongest insulated promoter was not affected implies that transcription is rate limiting for production of GFP and Gemini. However, for the strongest insulated promoter driving dsRed production, we observed decreased apparent promoter activity. In this instance, translation or

chaperone-dependent folding may have become limiting for protein synthesis. In such cases, one can no longer assume that protein production will depend linearly on promoter activity and instead, a more complicated transfer function to correlate promoter activity to protein production will have to be determined. When a process other than transcription becomes rate limiting for production, no amount of insulation will mitigate non-linear effects. When promoters react non-uniformly to the introduction of a new ORF, it can be difficult to ascertain which one or ones have been perturbed. We encountered this problem for the insulated promoters driving expression of the dsRed ORF. Normalizing activity to proD, our standard reference, suggested that the proA, proB and proC promoters had anomalously high activities. In contrast, normalizing activity to j23101 suggested strongly that proD had anomalously low activity in this context. The same conclusion would emerge by normalizing to one of the other insulated promoters.

Finally, we demonstrated that the insulated promoters act predictably when driving GFP expression from a single-copy locus in the chromosome. By using RPUs, we were able to compare promoter strength between expression from a plasmid vector and from the chromosome, further demonstrating the value of this relative measurement technique. We anticipate that the use of these promoters in the chromosome will be of value to the synthetic biology community.

Achieving proper steady-state protein levels can be accomplished using libraries of constitutive promoters as described here, or by using promoters whose activity can be titrated using a small-molecule inducer (13,45). Inducible systems are particularly appealing for applications in which only one-gene product must be regulated and for those systems in which it is advantageous to regulate gene activity dynamically (e.g. induction and repression of essential or toxic genes). The use of insulation sequences to decrease context-dependent promoter activity could easily be extended to regulated promoters.

ACKNOWLEDGEMENTS

We thank Jason Kelly, Drew Endy, the MIT Synthetic Biology Working Group and members of the Sauer lab for helpful discussion and materials.

FUNDING

Funding for open access charge: National Institutes of Health grants (AI-15706 and AI-16892).

Conflict of interest statement. None declared.

REFERENCES

1. Ro,D.K., Paradise,E.M., Ouellet,M., Fisher,K.J., Newman,K.L., Ndungu,J.M., Ho,K.A., Eachus,R.A., Ham,T.S., Kirby,J. *et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**, 940–943.
2. Atsumi,S. and Liao,J.C. (2008) Metabolic engineering for advanced biofuels production from Escherichia coli. *Curr. Opin. Biotechnol.*, **19**, 414–419.
3. Elowitz,M.B. and Leibler,S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–338.
4. Stricker,J., Cookson,S., Bennett,M.R., Mather,W.H., Tsimring,L.S. and Hasty,J. (2008) A fast, robust and tunable synthetic gene oscillator. *Nature*, **456**, 516–519.
5. Alper,H., Fischer,C., Nevoigt,E. and Stephanopoulos,G. (2005) Tuning genetic control through promoter engineering. *Proc.Natl Acad. Sci. USA*, **102**, 12678–12683.
6. Anthony,J.R., Anthony,L.C., Nowroozi,F., Kwon,G., Newman,J.D. and Keasling,J.D. (2009) Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metab. Eng.*, **11**, 13–19.
7. Lutke-Eversloh,T. and Stephanopoulos,G. (2008) Combinatorial pathway analysis for improved L-tyrosine production in Escherichia coli: identification of enzymatic bottlenecks by systematic gene overexpression. *Metab. Eng.*, **10**, 69–77.
8. Wong,W.W., Tsai,T.Y. and Liao,J.C. (2007) Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol. Syst. Biol.*, **3**, 130.
9. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.
10. Basu,S., Gerchman,Y., Collins,C.H., Arnold,F.H. and Weiss,R. (2005) A synthetic multicellular system for programmed pattern formation. *Nature*, **434**, 1130–1134.
11. You,L., Cox,R.S. 3rd, Weiss,R. and Arnold,F.H. (2004) Programmed population control by cell-cell communication and regulated killing. *Nature*, **428**, 868–871.
12. Yokobayashi,Y., Weiss,R. and Arnold,F.H. (2002) Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA*, **99**, 16587–16591.
13. Canton,B., Labno,A. and Endy,D. (2008) Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.*, **26**, 787–793.
14. Kelly,J.R., Rubin,A.J., Davis,J.H., Ajo-Franklin,C.M., Cumbers,J., Czar,M.J., de Mora,K., Glieberman,A.L., Monie,D.D. and Endy,D. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.*, **3**, 4.
15. Endy,D. (2005) Foundations for engineering biology. *Nature*, **438**, 449–453.
16. Browning,D.F. and Busby,S.J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, **2**, 57–65.
17. deHaseth,P.L., Zupancic,M.L. and Record,M.T. Jr (1998) RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J. Bacteriol.*, **180**, 3019–3025.
18. Ross,W., Ernst,A. and Gourse,R.L. (2001) Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev.*, **15**, 491–506.
19. Estrem,S.T., Ross,W., Gaal,T., Chen,Z.W., Niu,W., Ebright,R.H. and Gourse,R.L. (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev.*, **13**, 2134–2147.
20. Estrem,S.T., Gaal,T., Ross,W. and Gourse,R.L. (1998) Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl Acad. Sci. USA*, **95**, 9761–9766.
21. Pemberton,I.K., Muskhelishvili,G., Travers,A.A. and Buckle,M. (2000) The G+C-rich discriminator region of the tyrT promoter antagonises the formation of stable preinitiation complexes. *J. Mol. Biol.*, **299**, 859–864.
22. Chan,C.L. and Gross,C.A. (2001) The anti-initial transcribed sequence, a portable sequence that impedes promoter escape, requires sigma70 for function. *J. Biol. Chem.*, **276**, 38201–38209.
23. Hsu,L.M. (2008) *Escherichia coli and Salmonella*. ASM Press, Washington, DC.
24. de Crombrugghe,B., Busby,S. and Buc,H. (1984) Cyclic AMP receptor protein: role in transcription activation. *Science*, **224**, 831–838.

25. Lawson,C.L., Swigon,D., Murakami,K.S., Darst,S.A., Berman,H.M. and Ebright,R.H. (2004) Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.*, **14**, 10–20.
26. Ebright,R.H. (1993) Transcription activation at Class I CAP-dependent promoters. *Mol. Microbiol.*, **8**, 797–802.
27. Rojo,F. (1999) Repression of transcription initiation in bacteria. *J. Bacteriol.*, **181**, 2987–2991.
28. Mencia,M., Monsalve,M., Rojo,F. and Salas,M. (1996) Transcription activation by phage phi29 protein p4 is mediated by interaction with the alpha subunit of Bacillus subtilis RNA polymerase. *Proc. Natl Acad. Sci. USA*, **93**, 6616–6620.
29. Anderson,J.C. (2009) *Registry of Standard Biological Parts*, [www.partsregistry.org].
30. Jensen,P.R. and Hammer,K. (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl. Environ. Microbiol.*, **64**, 82–87.
31. Hammer,K., Mijakovic,I. and Jensen,P.R. (2006) Synthetic promoter libraries–tuning of gene expression. *Trends Biotechnol.*, **24**, 53–55.
32. Martin,L., Che,A. and Endy,D. (2009) Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. *PLoS ONE*, **4**, e7569.
33. Meng,W., Belyaeva,T., Savery,N.J., Busby,S.J., Ross,W.E., Gaal,T., Gourse,R.L. and Thomas,M.S. (2001) UP element-dependent transcription at the Escherichia coli rrnB P1 promoter: positional requirements and role of the RNA polymerase alpha subunit linker. *Nucleic Acids Res.*, **29**, 4166–4178.
34. Mendoza-Vargas,A., Olvera,L., Olvera,M., Grande,R., Vega-Alvarado,L., Taboada,B., Jimenez-Jacinto,V., Salgado,H., Juarez,K., Contreras-Moreira,B. *et al.* (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. *PLoS ONE*, **4**, e7526.
35. Shetty,R.P., Endy,D. and Knight,T.F. Jr (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.*, **2**, 5.
36. Horton,R.M., Cai,Z.L., Ho,S.N. and Pease,L.R. (1990) Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques*, **8**, 528–535.
37. Datta,S., Costantino,N. and Court,D.L. (2006) A set of recombineering plasmids for gram-negative bacteria. *Gene*, **379**, 109–115.
38. Rossi,J.J., Soberon,X., Marumoto,Y., McMahon,J. and Itakura,K. (1983) Biological expression of an Escherichia coli consensus sequence promoter and some mutant derivatives. *Proc. Natl Acad. Sci. USA*, **80**, 3203–3207.
39. Aoyama,T., Takanami,M., Ohtsuka,E., Taniyama,Y., Marumoto,R., Sato,H. and Ikehara,M. (1983) Essential structure of E. coli promoter: effect of spacer length between the two consensus sequences on promoter function. *Nucleic Acids Res.*, **11**, 5855–5864.
40. Paul,B.J., Ross,W., Gaal,T. and Gourse,R.L. (2004) rRNA transcription in Escherichia coli. *Annu. Rev. Genet.*, **38**, 749–770.
41. Win,M.N., Liang,J.C. and Smolke,C.D. (2009) Frameworks for programming biological function through RNA parts and devices. *Chem. Biol.*, **16**, 298–310.
42. Campbell,R.E., Tour,O., Palmer,A.E., Steinbach,P.A., Baird,G.S., Zacharias,D.A. and Tsien,R.Y. (2002) A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA*, **99**, 7877–7882.
43. Winkler,W., Nahvi,A. and Breaker,R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
44. Suess,B., Fink,B., Berens,C., Stentz,R. and Hillen,W. (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.*, **32**, 1610–1614.
45. Lee,S.K. and Keasling,J.D. (2005) A propionate-inducible expression system for enteric bacteria. *Appl. Environ. Microbiol.*, **71**, 6856–6862.