microbial biotechnology

Open Access

Context-dependency of synthetic minimal promoters in driving gene expression: a case study

Liyuan Jin,[1,†] Said Nawab,[1,†] Mengli Xia,[1] Xiaoyan Ma[1*] and Yi-Xin Huo[1,2]

[1]*Key Laboratory of Molecular Medicine and Biotherapy, School of Life Science, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing, 100081, China.*

[2]*UCLA Institute for Technology Advancement (Suzhou), 10 Yueliangwan Road, Suzhou Industrial Park, Suzhou, 215123, China.*

Summary

Synthetic promoters are considered ideal candidates in driving robust gene expression. Most of the available synthetic promoters are minimal promoters, for which the upstream sequence of the 5′ end of the core region is usually excluded. Although the upstream sequence has been shown to mediate transcription of natural promoters, its impact on synthetic promoters has not been widely studied. Here, a library of chromosomal DNA fragments is randomly fused with the 5′ end of the J23119 synthetic promoter, and the transcriptional performance of the promoter is evaluated through β-galactosidase assay, fluorescence intensity and chemical biosynthesis. Results show that changes in the upstream sequence can induce significant variation in the promoter strength of up to 5.8-fold. The effect is independent of the length of the insertions and the number of potential transcription factor binding sites. Several DNA fragments that are able to enhance the transcription of both the natural and the synthetic promoters are identified. This study indicates that the synthetic minimal promoters are susceptible to the surrounding sequence context. Therefore, the upstream sequence should be treated as an indispensable component in the design and application of synthetic promoters, or as an independent genetic part for the fine-tuning of gene expression.

Received 9 July, 2019; accepted 9 September, 2019.
*For correspondence. E-mail xyma@bit.edu.cn; Tel. +86 010 68917657; Fax + 86 010 68917657.
†These authors contributed equally to this work.
Microbial Biotechnology (2019) **12**(6), 1476–1486
doi:10.1111/1751-7915.13489
Funding information
The work was jointly supported by the National Key R&D Program of China (grant no. 2017YFD0201400) and the National Nature Science Foundation of China (grant no. 21676026).

Introduction

Promoters are among the most fundamental components for synthetic biological systems. They affect system behaviour by switching on and off gene transcription in response to various stimuli and by mediating the rate of transcription. Exploiting of these features has enabled the construction of genetic circuits (Nielsen *et al*., 2016) and environmental sensors (Dietrich *et al*., 2013), as well as the overexpression of biosynthetic pathway genes for the production of value-added chemicals (Hwang *et al*., 2018). Some natural promoters have been well characterized and employed to drive gene expression (Dahl *et al*., 2013; Huo *et al*., 2018). However, there are only limited number of available natural promoters, which could interact with the endogenous transcription machinery and disturb the gene transcription of the host. Therefore, the performance of natural promoters is usually context-dependent (Blazeck and Alper, 2013; Gilman and Love, 2016). In other words, their transcription is affected by the choice and arrangement of the surrounding genetic constructs, which can lead to regulatory failures for the engineered biological systems.

Synthetic promoters offer a solution to these drawbacks. By splicing together artificial pieces of promoter elements, novel promoters that do not exist in nature can be constructed rapidly (Blazeck and Alper, 2013). These promoters have the potential to be independent of the endogenous transcription regulation and function more robustly than those naturally derived ones (Gilman and Love, 2016; Portela *et al*., 2017). Some synthetic promoters such as the J231 family in the Registry of Standard Biological Parts have gained wide application in gene circuit construction (Qi *et al*., 2012), tool development (Jervis *et al*., 2019), chemical production (Meng *et al*., 2016; Choi *et al*., 2017; Zhou *et al*., 2017) and the characterization of biological systems (Pasotti *et al*., 2012). However, inconsistencies in their transcriptional strength and responses to environmental signals are often encountered (Pasotti *et al*., 2012; Kosuri *et al*., 2013; Zucca *et al*., 2015; Jervis *et al*., 2019), for which promoters exhibiting high activities in certain conditions might become the weak ones under another context. The lack of robustness on the performance of synthetic promoters can be partially attributed to the host genetic background, the gene copy number and the environmental factors, while the potential influence of the upstream

sequence on the promoter strength has not yet been widely noticed.

Most synthetic promoters contain only the core promoter elements, which are the σ factor binding sites and the spacer (Jervis *et al*., 2019). The σ factor binding sites are usually the derivatives of the naturally defined –10 and –35 boxes, while the spacer is usually a 17-bp random fragment placed in between the two binding sites. Compared with natural promoters, the synthetic core promoters can achieve higher transcriptional activities and broader dynamic ranges (Yim *et al*., 2013; Gilman and Love, 2016). However, the performance of promoters does not depend solely on the core elements. Sequence located upstream of the core region can also mediate the promoter strength. One example is the 'UP element' adjacent to the –35 site of some natural promoters such as the *rrnB*p1 in *E. coli* (Estrem *et al*., 1998). This ~20 bp sequence is able to bind the carboxyl-terminal domain of the α subunit (α-CTD) of the RNA polymerase (RNAP), anchoring RNAP to the promoter and stimulates transcription initiation. Replacing the inherent UP element with other derivatives or removing it from its promoter could alter transcription of up to hundreds of folds (Ross *et al*., 1998; Presnell *et al*., 2019). Another source of influence comes from the transcription factors (TFs) that bind to their regulatory regions reside within the upstream sequence. In most cases, TFs bind to the upstream region can activate transcription by facilitating the binding between RNAP and the promoter through protein–protein interactions, while the case of transcription repression also exists for some TFs (Shimada *et al*., 2011). Therefore, a synthetic promoter may not perform as desired if placed downstream a sequence containing the UP element-like regions or the TF binding sites (TFBSs), and its performance is likely to vary from one context to another. Thus, synthetic promoters containing only the core elements are considered to face similar drawbacks as the natural promoters, and investigating the context-dependency of the synthetic promoters should urge rational design and application of the synthetic promoters for genetic regulation.

Here, we investigated the effects of upstream sequence on the performance of synthetic promoters. The commonly used synthetic promoter J23119 (Part: BBa_J23119) was employed to drive the expression of the *lacZ*α reporter gene. An upstream sequence library was obtained by digesting the *Escherichia coli* genomic DNA and the fragmented DNA pieces were inserted into the upstream of the J23119 promoter. The promoter strength was characterized by both the blue-white screening and the β-galactosidase (β-gal) assay, and its relationship with the upstream sequence was analysed. Sequences leading to significant changes in the β-gal activity were fused with other synthetic and natural promoters to test their effects in mediating protein overexpression and chemical production. We hypothesize that the synthetic promoters are susceptible to their upstream sequences, which could both up- and down-regulate gene expression. The upstream sequences have the potential to mediate core promoter transcription under various scenarios including protein production and biosynthesis.

Results

Generation of upstream sequence library

The plasmid pUC19 was used to investigate the effects of the upstream sequence on the promoter strength. After excluding the restriction enzymes whose recognition sequences were within the pUC19 vector, the BtgI with restriction site of 5′-CCRYGG-3′ was selected for chromosome digestion. Because the R and Y each represented two types of bases, the probability of occurrence for this 6-bp sequence would be $1/4^4 \times 1/2^2$, which is 1/1024. Therefore, the number of this restriction site on the chromosome would be around 4700 (the length of the chromosomal DNA divided by 1024), which agreed with the criteria for enzyme selection (see Experimental procedures). As expected, a total of 4016 BtgI restriction sites were discovered on the DH10B genome (Fig. 1). These sites would segment the chromosome into DNA pieces ranging from 10 to 9747 bp, with an average length of 1171 bp. The digested fragments were inserted into the BtgI restriction site adjacent to the 5′ end of the J23119 and transformed into the DH5α strain. Colonies exhibiting light to dark blue were formed on the X-gal plate (Fig. 2A). Among them, 76 colonies were picked, from which 65 unique upstream fragments were identified (Supporting Data 1). Nearly 80% of the fragments were within 1.0 to 4.0 kb (Fig. 2B), giving an average length of 2.1 kb for the inserted upstream sequences.

The activities of β-gal expressed from promoters with different upstream insertions

A 4.5-fold variation in the enzyme activity was observed for the β-gal expressed from promoter J23119 with the 5′ end flanked by different fragments (Fig. 3A). Compared with the original expression plasmid (U0) with only the BtgI restriction site inserted into the 5′ end of J23119, the two upstream fragments U1 and U2 (numbered according to the rank of β-gal activity from high to low) led to a 50% and a 34% increase ($P < 0.0001$) in the β-gal activity respectively. For the rest fragments, six showed no significant influences (U3 to U8) on the β-gal activity, and the other 57 sequences reduced the enzyme activity by 13.4% to 65.6%.

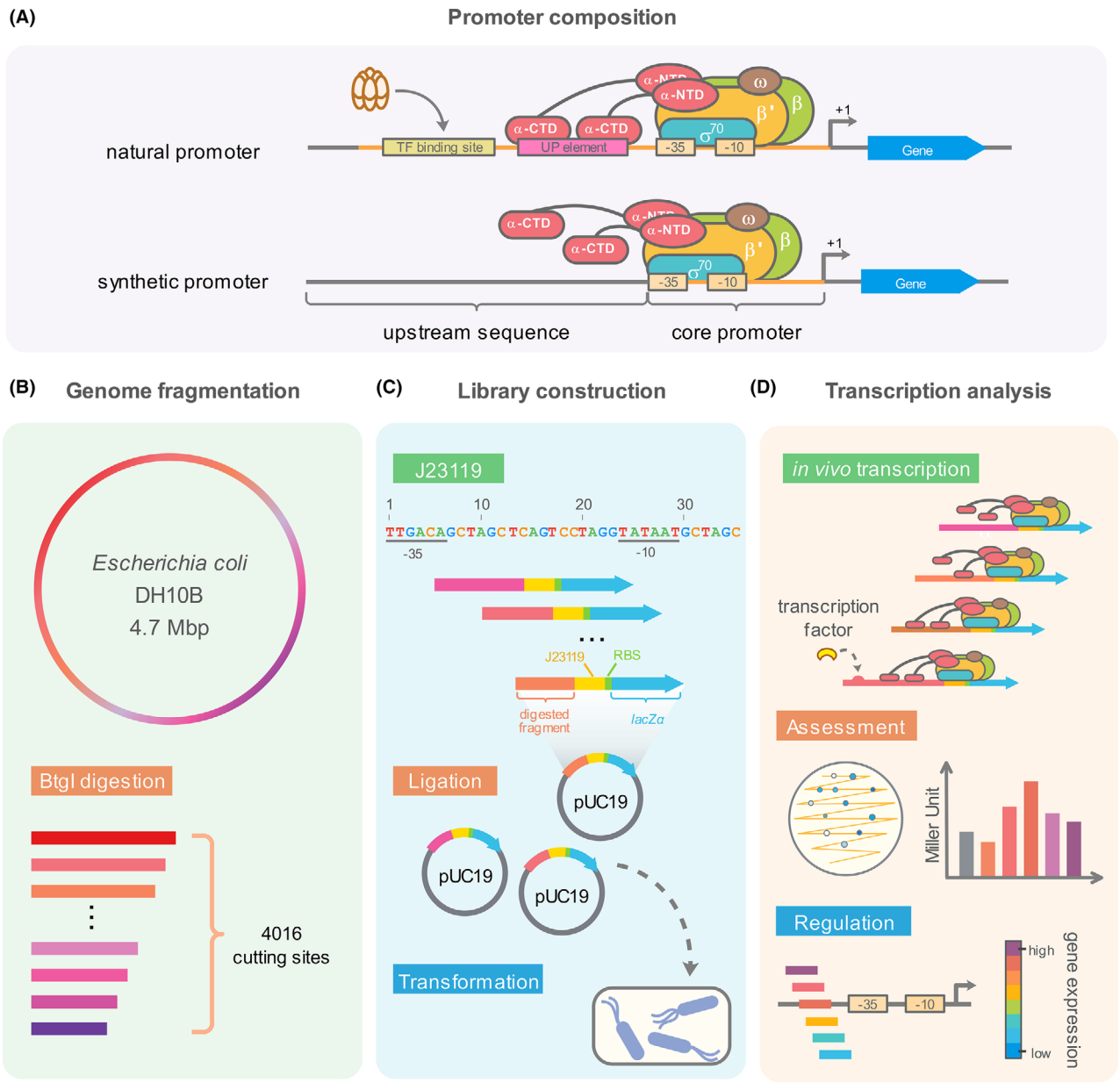

Fig. 1. The experimental scheme for investigating the effects of upstream sequences on the transcription of synthetic promoters.
A. Structures for natural and minimal synthetic promoters and their interactions with RNAP.
B. Generation of the insertion fragments.
C. Construction of the upstream-promoter fusion plasmids.
D. Analysis of the transcriptional strength for the fused promoters.

Analysis of the insertion sequences showed that the ones leading to lower β-gal activities than the initial plasmid U0 were in length ranged from 230 to 4271 bp, the ones showing no particular influences on the enzyme activity spanned from 869 to 3888 bp. Fragment U2 that induced one of the highest β-gal activities had a 504-bp sequence inserted into the BtgI restriction site. In comparison, only a 4-bp insertion was identified for U1 at the upstream of the BtgI restriction site. Besides, both the U1 and the U2 had an identical four-base (GGGA) insertion in between the 3′ end of the BtgI restriction site and the 5′ end of the –35 promoter element.

Effects of fixed-length upstream insertions on promoter activity

No significant correlation was found between the β-gal activity and the fragment length (Table 1). Nevertheless,

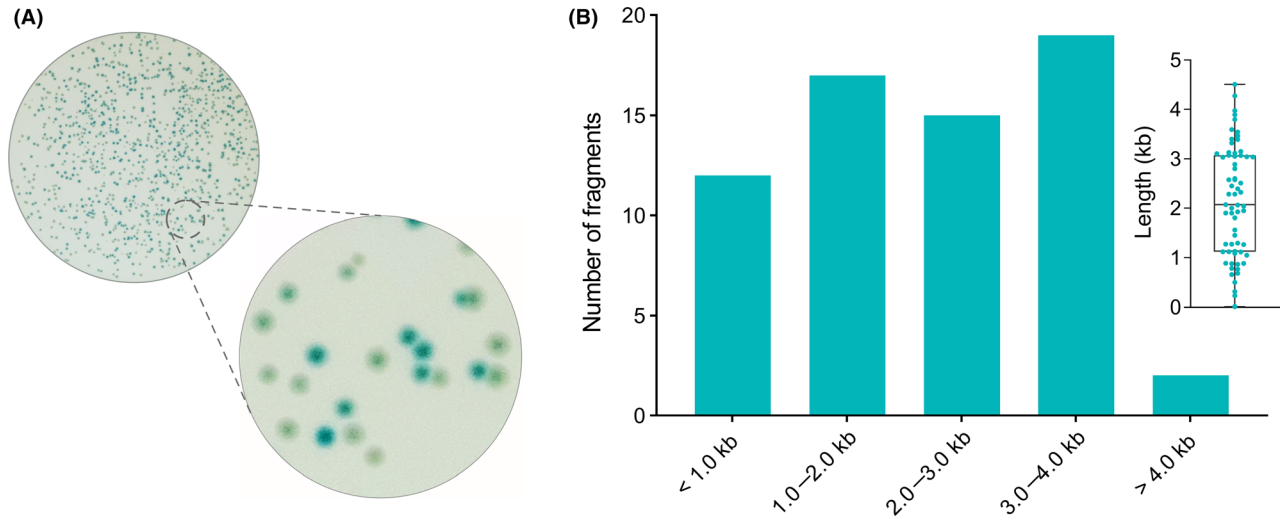

Fig. 2. Screening of the plasmids carrying the upstream insertions.
A. Colonies expressing the β-gal from the synthetic promoters with different upstream sequences.
B. Length distribution of the inserted upstream fragments.

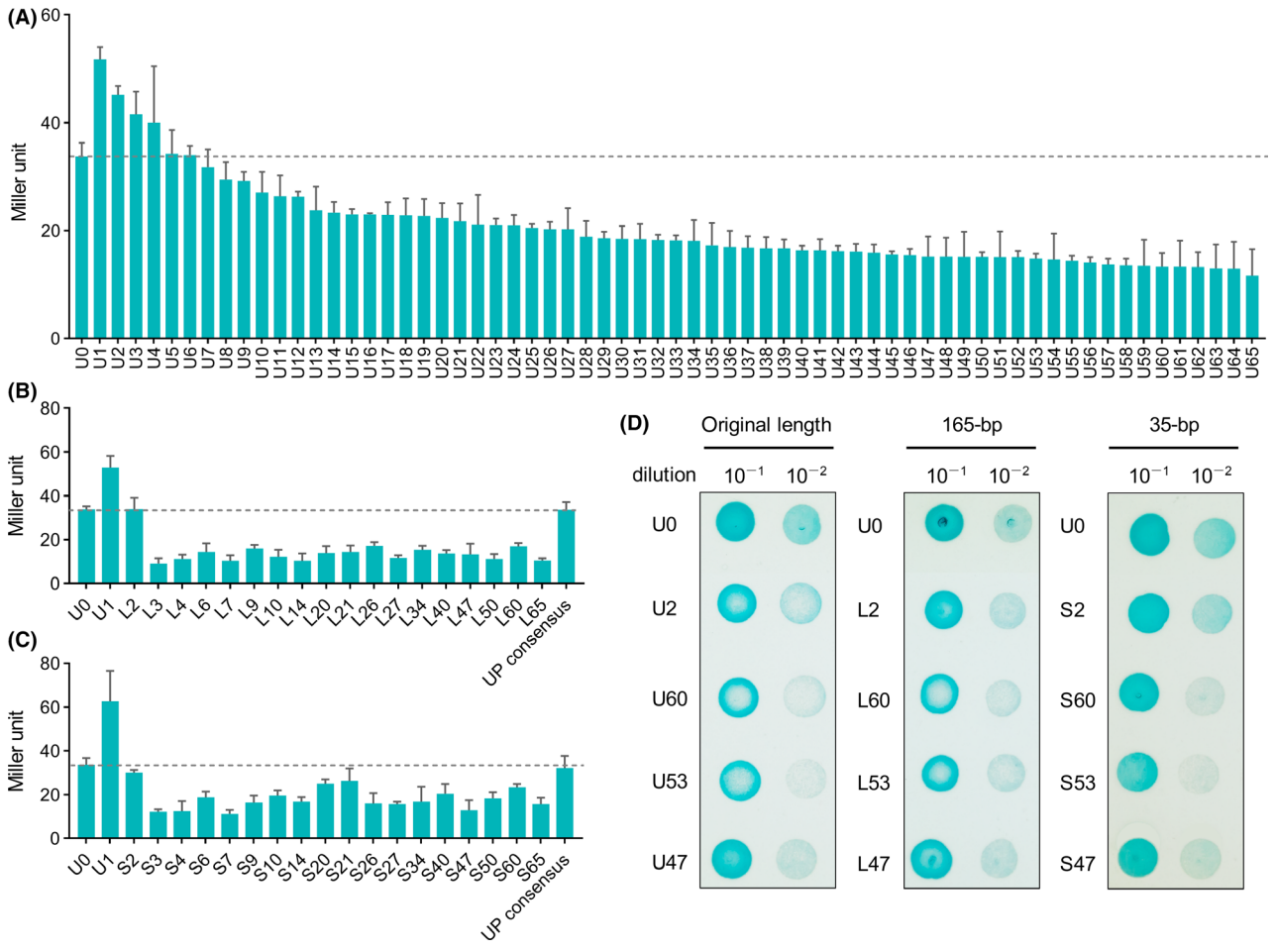

Fig. 3. Miller units for β-gal expressed from J23119 with (A) random length, (B) 165-bp, and (C) 35-bp upstream fragments. (D) Colour development of cells expressing β-gal from promoter J23119 fused with the original and the truncated upstream sequences. Values and error bars represent the mean and the SD (n = 3).

a total of 18 insertion sequences, which contributed to higher or lower β-gal activities than the initial plasmid U0, were selected to investigate the potential influence of plasmid size on transcription. These fragments were truncated to fixed length (Data S1), producing two sets of plasmids containing only the first 165-bp or the first 35-bp region from the 3′-end of the inserted fragments (Table S1). The effects of the fixed-length upstream sequences on the promoter strength were evaluated again by β-gal assay. A reference plasmid with the confirmed UP element consensus sequence (24 bp; Estrem *et al*., 1998) inserted into the promoter upstream was also included for the comparison of promoter activities.

The 165-bp insertions led to significant variation in the Miller units of up to 3.7-fold (L2 vs L3; Fig. 3B and D). Fragment L2, which contributed to one of the highest β-gal activities in its full length (U2), led to Miller units the same levels as the control plasmid and the plasmid carrying the consensus UP element, but was 36% lower than when in its full length. The enzyme activity decreased to 27%–42% of the control by the truncated insertion fragments (L3 to L8), which have no significant effects on β-gal activity in their full length (U3 to U8). The rest of the fragments showing negative effects on the β-gal activity in their original length also led to decreased β-gal activity when truncated to 165 bp. Further reducing the fragments to 35 bp also resulted in significant difference in the Miller units of up to 2.7-fold (S2 vs S3; Fig. 3C and D). The effects of the 35-bp sequences on the β-gal activity were similar as when they were in the length of 165 bp.

To examine the effects of the selected upstream fragments on the activity of other promoters, five 35-bp fragments were inserted into the upstream of the J23106 (Part: BBa_J23106) and the *fumA* promoters (P_{fumA}). The 35-bp fragment (S1) including the –36 to –70 promoter region on the *lac*Z plasmid carrying U1 was also cloned into the upstream of J23106 and P_{fumA}. When using *egfp* as the marker gene, a 3.5- and a 2.5-fold variations in the fluorescence intensity were observed for J23106 (Fig. 4A) and P_{fumA} (Fig. 4B) respectively. For both promoters, fragments S1 and S2 led to the highest fluorescence intensities, and the rest upstream fragments also enhanced *egfp* expression in comparison with the control that did not harbour upstream insertion. Moreover, compared with the consensus UP element, a 1.4- and a 1.6-fold increases in the fluorescence intensity were achieved by S1 and S2, respectively, after inserting into the upstream of J23106. The fluorescence intensities for these two fragments were on the same level ($P < 0.05$) as that of the consensus UP element when inserted into the upstream of P_{fumA}.

Analysis of the upstream sequences

A total of 2 to 7 potential TFBSs were discovered for each of the 35-bp fragments (Fig. 5A). Among the corresponding TFs, five (AscG, IclR, MraZ, NanR and PutA) acted as transcriptional repressors when bound to the core promoter regions of their targeted genes. Three TFs including CspA, Dan and MalT can activate transcription when bind to the upstream of their targeted promoters. The rest were dual regulators that might act as either activators or repressors. For fragment S1, five potential MraZ binding sites were identified within the promoter region from –36 to –49. Besides MraZ, TFBSs for IclR and NanR were also discovered within S2. For fragments S7 and S3 that led to the two lowest Miller units, the binding sites for four repressors including NanR, MraZ, AscG and IclR were discovered within the promoter upstream region from –44 to –54. However, the β-gal activity did not show significant correlation with the number of TFBSs for either the full-length insertions or the truncated upstream fragments (Table 1).

Multiple sequence alignment showed significant differences among these insertion fragments. The selected 35-bp sequences were grouped into three lineages (Fig. 5B). Fragment S1 that elicited the highest β-gal activity was closest to the upstream sequence of the control plasmid U0. The major difference between these two sequences was within the first 19 bases (–35 to –54) due to the four-base insertion (GGGA) at the 3′ end of the BtgI site within S1. Fragment S2 that contributed to a comparable β-gal activity to that of the control plasmid was assigned to another lineage. It shared the highest similarity to the consensus UP element among all the 35-bp sequences, for which the –47 to –52 region of

Table 1. Pearson correlation between the β-gal activities and the characteristics of the inserted upstream fragments.

β-gal activity	Fragment length	AT content	No. of TFBSs	$\beta\text{-gal}_{ori}$	$\beta\text{-gal}_{165}$	$\beta\text{-gal}_{35}$
$\beta\text{-gal}_{ori}$	−0.168	−0.345**	−0.081	–	0.291	0.160
$\beta\text{-gal}_{165}$	–	−0.451	0.271	0.326	–	0.689**
$\beta\text{-gal}_{35}$	–	0.016	0.221	0.004	0.689**	–

$\beta\text{-gal}_{ori}$, β-gal activities derived from plasmids with the original insertion sequences; $\beta\text{-gal}_{165}$, β-gal activities derived from plasmids with the 165-bp truncated insertion fragments; $\beta\text{-gal}_{35}$, β-gal activities derived from plasmids with the 35-bp truncated insertion fragments.
**Significant correlation at the 0.01 level (two-tailed).

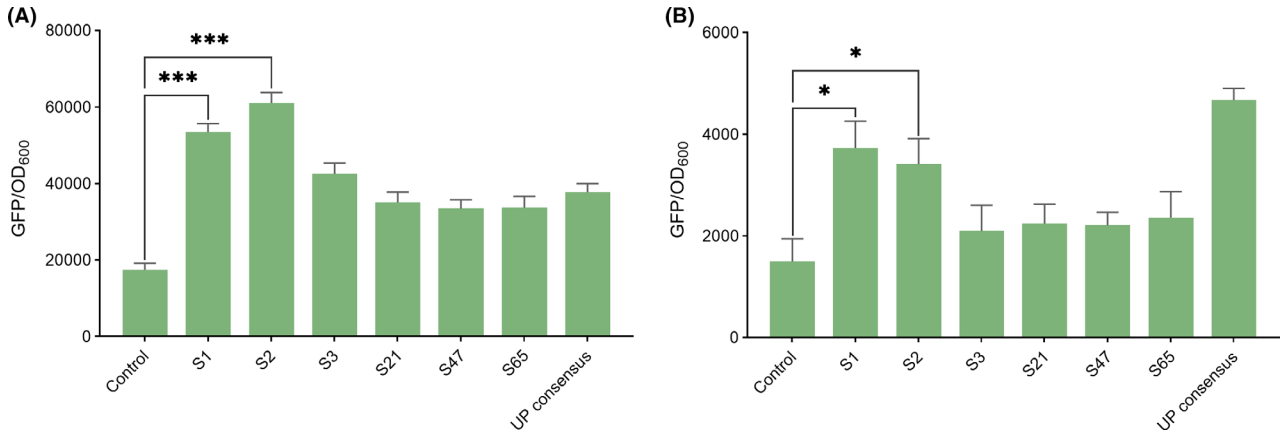

Fig. 4. Fluorescence intensities for EGFP expressed from (A) the synthetic promoter J23106 and (B) the natural promoter P_{fumA} fused with different 35-bp upstream fragments. Values and error bars represent the mean and the SD (n = 3). *P < 0.05 and ***P < 0.001 as determined by unpaired t-test.

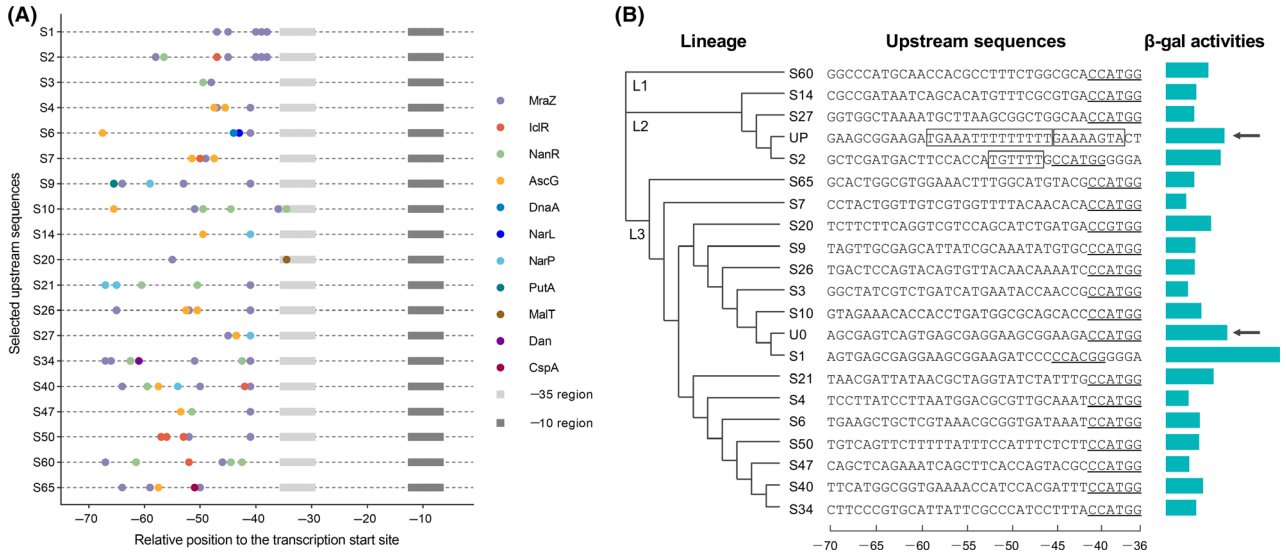

Fig. 5. Sequence analysis of the upstream fragments.
A. Predicted TFBSs within the 35-bp upstream fragments, dots indicate the centre of the binding sites.
B. Multiple sequence alignment of the upstream sequences. The selected upstream fragments were grouped into three lineages from L1 to L3, the blank boxes indicate the UP element subsites and the UP element-like sequence, the restriction sites are underlined. Arrows indicate the control plasmid (U0) and the plasmid carrying the consensus UP element (UP).

S2 and the corresponding distal subsite of the consensus UP element were both thymine rich. In addition, the first 6 bases of S2 also differed significantly from most of the fragments due to the same four-base insertion as S1.

Effects of the upstream sequences on chemical production

To test the effects of upstream sequences on the expression of biosynthetic pathways, fragment S1 and S2 were each fused with the 5′ end of the 35-bp P_{fumA} and inserted into the upstream of the isobutanol biosynthetic pathway (Table S1). This pathway contained five genes that first converted pyruvate into 2-ketoisovalerate through acetolactate synthase (AlsS), ketol-acid reductoisomerase (IlvC) and dihydroxy-acid dehydratase (IlvD) and then into isobutanol through ketoisovalerate decarboxylase (KivD) and alcohol dehydrogenase (AdhA; Fig. 6A). The isobutanol production was greatly improved by these two fragments. Compared with the control P_{fumA} without any upstream fusions, fragment S1 led to a 2.4-fold increase in the isobutanol titre after 72 h of fermentation (Fig. 6B). The highest isobutanol production was achieved by fragment S2, for which a final titre

of 5.9 g l^{-1} was detected. This titre was 10.8-fold and 4.6-fold higher than that of the control and the S1 fragment respectively.

Discussion

This study reveals that the upstream sequence is a critical factor in mediating the performance of synthetic promoters, leading to significant variations in protein expression, specific enzyme activity and chemical production. Therefore, the upstream sequence should be taken into consideration when designing and employing the synthetic promoters. The upstream sequence can be used as an independent genetic part in tuning gene expression.

The influence of upstream sequence on promoter activity suggests that the commonly used synthetic promoters are not context-independent. However, the current design of synthetic promoters, especially for prokaryotic systems, has widely ignored the sequence context surrounding the core promoter. Besides the J231 family, other synthetic promoters such as the K137 family (e.g. K137030 and K137031) are also only contain the core region (Table S2). Therefore, the predefined performance of these synthetic promoters may not be reproduced once their sequence contexts are varied in actual situations, and discrepancies among different studies have been witnessed. For example, the J23105 promoter has shown transcriptional activities 95% lower than the commonly used promoter $P_L lacO_1$ (Qi *et al.*, 2012), or comparable to the $P_L lacO_1$ (Jervis *et al.*, 2019). Ruling out the interferences from factors such as the copy number, the type of the reporter protein and the stage of the cell cycle, difference in the upstream sequences among these studies is a non-negligible factor that jeopardizes the robustness of the synthetic promoters. The lack of robustness for synthetic promoters would fail their applications. For instance, when used to express the isobutanol synthetic pathway, the J23100, which is characterized to be stronger than $P_L lacO_1$ (Jervis *et al.*, 2019), can only produce 0.1 g l^{-1} isobutanol, which is significantly lower than that of the $P_L lacO_1$ and the natural promoter *rrnB*p1 (Fig. S1). Besides the design and application of the synthetic promoters, the upstream sequence is also ignored in the prediction of promoter strength (De Mey *et al.*, 2007; Rhodius and Mutalik, 2010). Taking the upstream sequence as an extra variable should improve the prediction accuracy and facilitate the rational design of a suitable promoter.

The upstream region affecting promoter strength can span from tens to thousands of base pairs in length. The β-gal activities derived from plasmids with long (β-gal_{ori}) or truncated insertions (β-gal_{165} or β-gal_{35}) do not show significant correlations (Table 1), suggesting that the distal sequence might have extra effects, which differ from the influences exerted by the proximal sequence, on promoter strength. Nevertheless, the similar effects of the 165-bp insertions and the 35-bp insertions on β-gal

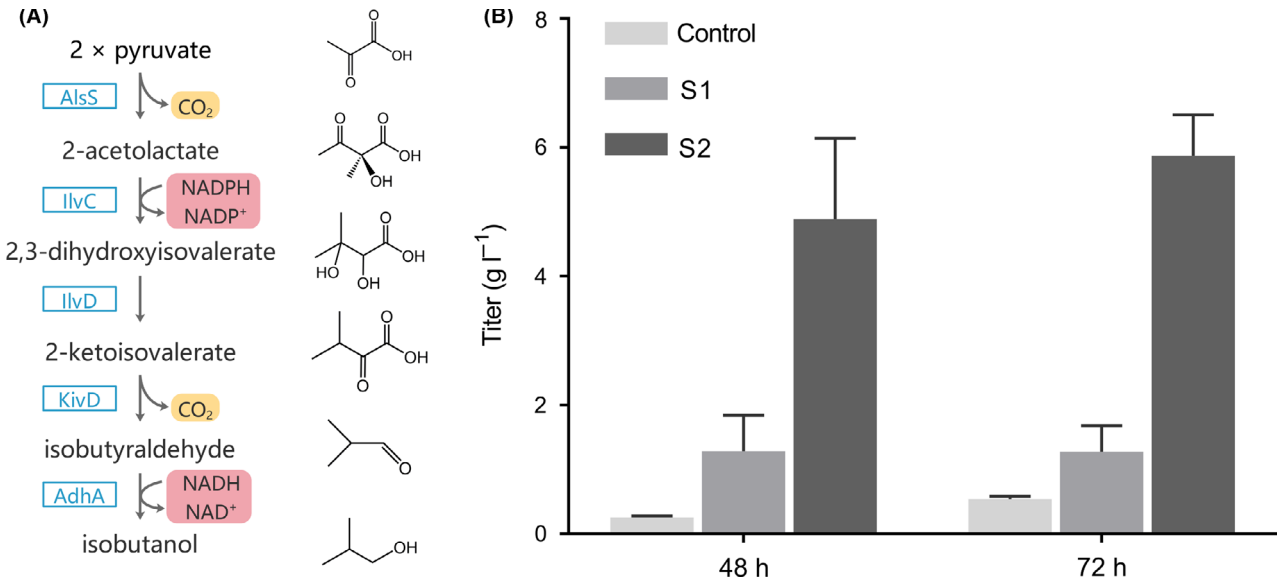

Fig. 6. Effects of the selected upstream sequence on chemical production.
A. Isobutanol biosynthetic pathway. AlsS: acetolactate synthase; IlvC: ketol-acid reductoisomerase; IlvD: dihydroxy-acid dehydratase; KivD: ketoisovalerate decarboxylase; and AdhA: alcohol dehydrogenase.
B. Isobutanol production driven by P_{fumA} fused with the selected upstream sequences. Values and error bars represent the mean and the SD (n = 3).

activities, as represented by the significant correlation between β-gal$_{165}$ and β-gal$_{35}$ (Table 1), indicate that the upstream region most adjacent to the core promoter plays crucial role in mediating the transcriptional activity. This is supported by a predictive modelling of the promoter strength, which shows that the adjacent upstream sequence explains more variance in promoter expression, than the distal region located at the upstream of the position −120 (Urtecho *et al*., 2018). Regardless of the difference in the contributions from different upstream segments to the promoter strength, the upstream sequence as a whole is equally important as the spacer region in tuning gene expression (Urtecho *et al*., 2018). In general, the upstream sequence mediates promoter strength by interacting with the RNA polymerase. Besides the commonly known AT-rich UP element that binds the α-CTD, the potential TFBSs and the promoter-like sequence (Huerta *et al*., 2006) reside within the upstream region, as well as the DNA conformation of the upstream sequence can also affect transcriptional activity (Bikard *et al*., 2010; Ma and Wang, 2014).

A total of 19 sets of upstream fragments are characterized in this study, which lead to changes in the promoter strength of up to 5.8-fold. No significant correlations are discovered between the promoter strength and the numbers of TFBSs for activators, repressors or dual regulators within the upstream fragments (Table 1). One explanation is that the role of TFs in mediating the promoter strength is trivial for these fragments, another explanation is that non-linear interaction exists between the TFs and the promote strength (Urtecho *et al*., 2018). The enhancer fragment S1 contains a 21-bp region that is identical to the upstream sequence of the initial plasmid (U0), but is located at a different position compared with the corresponding region in U0. Therefore, the positioning effects of the upstream sequence may play a role in tuning the promoter strength. The UP element-like sequence is not the sole option for enhancing the promoter strength, as fragment S1, which shares little resemblance to the consensus UP sequence, can also lead to higher or comparable transcriptional activities. The two enhancer fragments (S1 and S2) share a unique 6-bp sequence in the 3′ end adjacent to the core promoter, suggesting that the key region affecting the promoter expression may reside within several bases upstream the 5′ end of the −35 site.

As synthetic promoters are gaining wide applications in the programmable control of gene expression (Lee *et al*., 2016; Jo *et al*., 2019), achieving predictable transcriptional activity is crucial for the robust performance of the biological systems. Considering the susceptibility of the core promoters to the sequence context, the well-characterized natural promoters with defined upstream sequences should be better choices than those minimal synthetic promoters. If synthetic promoters have to be used, their transcriptional strengths should be re-characterized under specific sequence context. To reduce context-dependency, inclusion of an upstream sequence in the design of synthetic promoters is advised. For instance, leaky transcription from the upstream promoter-like sequences can be reduced by fusing the 5′ end of the synthetic promoters with transcription terminators (Engstrom and Pfleger, 2017). Insulating sequences in length of tens of base pairs are also able to buffer the promoter from the influences of the neighbouring region (Davis *et al*., 2011; Carr *et al*., 2017). Referring to the strategies used in promoter design, the upstream sequences can be generated by direct cloning from genomic non-promoter regions (Urtecho *et al*., 2018), random mutagenesis (Blazeck and Alper, 2013) and the integration of natural regulating elements (Davis *et al*., 2011). The produced upstream sequences can be fused with minimal promoters to add an extra level of regulation to gene transcriptions, just like the upstream activation or repressive sequences in eukaryotic promoters (Hahn and Young, 2011). As shown in this study, these sequences are able to fine-tune protein expression and chemical production for both the synthetic and the natural promoters, adding new genetic part to the toolbox for synthetic biology.

Experimental procedures

Digestion of E. coli *chromosome*

The fragments inserted into the upstream of the synthetic promoters were derived by enzymatically digesting the chromosome of *E. coli* strain DH10B. To produce a library of fragments with an average length of 1.0 kb, the 4.7 Mb chromosome had to be segmented into approximately 4700 pieces. Therefore, the restriction enzyme used for digestion should have around 4700 restriction sites resided within the *E. coli* chromosome. Five sets of 2 μg genomic DNA were enzymatically digested by 2 U of the restriction enzyme in a total reaction volume of 50 μl at 37°C for 3 h and purified using the GeneJET PCR Purification Kit (Thermo Fisher).

Plasmid construction

The primer UP-U0-F containing the intact J23119 sequence and the primer UP-U0-R embedded with the restriction site at the 5′ end of J23119 (Table S3) were used to amplify the pUC19 plasmid, producing a linear fragment with the original *lac* promoter, the *lac* operator, the CAP binding site and the *lac*I residue at the

upstream of *lacZ*α replaced with the synthetic promoter J23119 (Fig. S2a and b). The generated linear fragment had a 28-bp overlap, which was the –8 to –35 region of the J23119, at both ends, and thus could be ligated using the *In Vivo* Assembly (IVA) method by transforming into the *E. coli* strain DH5α (García-Nafría *et al.*, 2016). Plasmid pUP-U0 isolated from the DH5α cells would carry the restriction site at the –36 to –41 region of J23119. To insert the upstream fragments, 2 μg of the plasmid pUP-U0 was digested using 2 U of the restriction enzyme in a reaction volume of 50 μl at 37°C for 2 h, and the digested products were purified using the GeneJET PCR Purification Kit. For ligation, a total of 150 ng DNA fragments with an insert-to-vector molar ratio of 1:1 were ligated by 1 Weiss U of T4 DNA ligase (Thermo Fisher) in a total volume of 20 μl at 22°C for 1 h. The plasmids (pUP-series) were extracted (Table S1 and Fig. S2c), producing a library carrying different insertions at the upstream of the J23119. To obtain plasmids carrying the truncated chromosomal DNA insertions, the plasmid regions excluding the sequence between the 5′ ends of the inserted fragments and the defined sites from their 3′ ends were amplified from the pUP-series (Table S3 and Fig. S3). Each of the linear plasmids had a 20-bp overlap region at the ends and was ligated using the IVA method. Two pUC19-derived EGFP expression plasmids with J23106 and P_{fumA} (5′-ttgttaaaaaagtgtgtaggatattgttactcgct-3′) as the promoters, respectively, were employed to further investigate the effects of upstream insertions on promoter expression (Table S1). Fragments in fixed-length were inserted to the –36 position of the J23106 or P_{fumA} through primers using PCR. The primer design was similar to that used for constructing the plasmids carrying the truncated upstream insertions. For the forward primer, the first 27-bp region from its 5′ end was homologous to the first 27-bp region from the 3′ end of the 35-bp chromosomal DNA fragment (the insertion), and the first 27-bp region from the 5′ end of the reverse primer was homologous to the 5′ end of the 35-bp insertion. This would generate a 19-bp overlap sequence at the 5′ ends for both primers (Table S3). The rest 20-bp region for the forward primer was homologous to the 5′ end of the promoter, while the rest region for the reverse primer was homologous to the vector sequence adjacent to the promoter. The amplified linear fragments were ligated again using the IVA method, producing the pEGFP-P1 and the pEGFP-P2 plasmids (Table S1 and Fig. S4a). The same strategy was also employed to insert the targeted DNA fragments into the upstream of the operons that consisted of the isobutanol biosynthetic pathway (Table S1 and Fig. S4b and c). Detailed procedures for constructing each plasmid were included in Table S3.

Blue-white screening

Strains transformed with the *lacZ*α expression plasmids were plated onto LB agar medium containing 0.2 mg ml^{-1} X-gal, 1.0 mM IPTG and 100 μg ml^{-1} ampicillin, and incubated at 37°C for 24 h. Colonies exhibiting light to dark blue were picked and incubated in LB medium at 37°C overnight, and the cell cultures were spotted on X-gal agar plates in serial dilutions (0.1 and 0.01). Strain harbouring the control plasmid was also plated with the same dilution factor.

β-gal activity assay

Strains screened out from the X-gal plate were grown to mid-log phase and harvested by centrifuging at 4000 r.p.m. for 5 min. The cells were washed with 0.9% NaCl solution and resuspended in chilled Z buffer (Miller, 1972) containing 0.05 M β-mercaptoethanol to a final volume of 1.0 ml. The cells were permeabilized by adding 20 μl chloroform and 3.75 μl 1% SDS and equilibrated in a 28°C water bath for 5 min. The reaction was initiated by adding 200 μl of 4 mg ml^{-1} ortho-Nitrophenyl-β-galactoside and incubated at 28°C. A hundred microlitre of 1.0 M Na_2CO_3 was added into the tube to stop the reaction when sufficient yellow colour has developed. The ODs at both 420 nm and 550 nm were measured, and the Miller units were calculated as previously described (Miller, 1972).

Measurement of fluorescence intensity

The *E. coli* strains overexpressing EGFP were inoculated into 5.0 ml LB medium and incubated at 37°C for 20 h. The fluorescence was excited at 488 nm and measured at 510 nm, and the OD_{600} was also measured. Strain harbouring the vector plasmid was cultured under the same condition, and the fluorescence was measured as the background. The ratio of EGFP to OD_{600} was used to represent the florescence intensity.

Isobutanol fermentation and detection

Strains harbouring the isobutanol synthetic pathway were inoculated into 5 mL LB medium and incubated at 37°C in a shaker at 250 r.p.m. Four hundred microlitre of the overnight seed culture was inoculated into 20 ml M9 medium containing 40 g l^{-1} glucose, 10 g l^{-1} yeast extract, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 10^{-4}% thiamine, 100 μg ml^{-1} ampicillin and 50 μg ml^{-1} kanamycin. Fermentation was performed in a capped 250 ml flask at 30°C in a shaker at 250 r.p.m.

The produced isobutanol was detected by gas chromatograph. A DB-FFAP capillary column (30 m × 0.32 mm × 0.25 μm; Agilent Technologies, Santa Clara, CA, USA) was used to separate isobutanol from other compounds in the fermentation broth with n-pentanol as the internal standard. The GC oven was first held at 80°C for 3 min, increased to 230°C at 115°C min^{-1} and held for 1 min. The isobutanol was then detected by a flame ionization detector at 280°C.

Sequence analysis

The inserted DNA fragments were sequenced and blasted against the *E. coli* DH10B genome (GenBank: CP000948.1) to check for the presence of natural promoters and terminators. The inserted sequences were then scanned against the *E. coli* TFBSs assigned by RegulonDB using the tfscan tool embedded in EMBOSS 6.5.7 (Rice *et al*., 2000). The potential UP elements within the upstream region were also predicted by referring to the UP element consensus sequences (Estrem *et al*., 1998; Gourse *et al*., 2000). Multiple sequence alignment was performed using CLUSTALW 2.1, and a tree was generated by the neighbour-joining method (Jukes and Cantor, 1969; Larkin *et al*., 2007).

Acknowledgements

The work was jointly supported by the National Key R&D Program of China (grant no. 2017YFD0201400) and the National Nature Science Foundation of China (grant no. 21676026).

Conflict of interest

The authors have no conflict of interest.

Author contributions

X.Y.M. designed and supervised the experiments. Y.X.H. proposed the project. L.Y.J., S.N. and M.L.X. performed the experiments. X.Y.M., L.Y.J. and S.N. analysed the data. X.Y.M. and Y.X.H. drafted the manuscript. All authors read and approved the final manuscript.

References

Bikard, D., Loot, C., Baharoglu, Z., and Mazel, D. (2010) Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiol Mol Biol Rev* **74:** 570–588.

Blazeck, J., and Alper, H.S. (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J* **8:** 46–58.

Carr, S.B., Beal, J., and Densmore, D.M. (2017) Reducing DNA context dependence in bacterial promoters. *PLoS ONE* **12:** e0176013.

Choi, S.Y., Kim, W.J., Yu, S.J., Park, S.J., Im, S.G., and Lee, S.Y. (2017) Engineering the xylose-catabolizing Dahms pathway for production of poly(d-lactate-co-glycolate) and poly(d-lactate-co-glycolate-co-d-2-hydroxybutyrate) in *Escherichia coli*. *Microb Biotechnol* **10:** 1353–1364.

Dahl, R.H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T.S., Redding-Johanson, A.M., *et al*. (2013) Engineering dynamic pathway regulation using stress-response promoters. *Nat Biotechnol* **31:** 1039.

Davis, J.H., Rubin, A.J., and Sauer, R.T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* **39:** 1131–1141.

De Mey, M., Maertens, J., Lequeux, G.J., Soetaert, W.K., and Vandamme, E.J. (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol* **7:** 34.

Dietrich, J.A., Shis, D.L., Alikhani, A., and Keasling, J.D. (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth Biol* **2:** 47–58.

Engstrom, M.D., and Pfleger, B.F. (2017) Transcription control engineering and applications in synthetic biology. *Synth Syst Biotechnol* **2:** 176–191.

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.

García-Nafría, J., Watson, J.F., and Greger, I.H. (2016) IVA cloning: a single-tube universal cloning system exploiting bacterial *In Vivo* Assembly. *Sci Rep* **6:** 27459.

Gilman, J., and Love, J. (2016) Synthetic promoter design for new microbial chassis. *Biochem Soc Trans* **44:** 731–737.

Gourse, R.L., Ross, W., and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol* **37:** 687–695.

Hahn, S., and Young, E.T. (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189:** 705–736.

Huerta, A.M., Francino, M.P., Morett, E., and Collado-Vides, J. (2006) Selection for unequal densities of σ^{70} promoter-like signals in different regions of large bacterial genomes. *PLoS Genet* **2:** e185.

Huo, Y.-X., Guo, L., and Ma, X. (2018) Biofuel production with a stress-resistant and growth phase-independent promoter: mechanism revealed by in vitro transcription assays. *Appl Microbiol Biotechnol* **102:** 2929–2940.

Hwang, H.J., Lee, S.Y., and Lee, P.C. (2018) Engineering and application of synthetic *nar* promoter for fine-tuning the expression of metabolic pathway genes in *Escherichia coli*. *Biotechnol Biofuels* **11:** 103.

Jervis, A.J., Carbonell, P., Taylor, S., Sung, R., Dunstan, M.S., Robinson, C.J., *et al*. (2019) SelProm: a queryable

and predictive expression vector selection tool for *Escherichia coli*. *ACS Synth Biol* **8:** 1478–1483.
Jo, M., Noh, M.H., Lim, H.G., Kang, C.W., Im, D.-K., Oh, M.-K., and Jung, G.Y. (2019) Precise tuning of the glyoxylate cycle in *Escherichia coli* for efficient tyrosine production from acetate. *Microb Cell Fact* **18:** 57.
Jukes, T.H., and Cantor, C.R. (1969) Evolution of protein molecules. In *Mammalian Protein Metabolism*. Munro, H.N. (ed). New York, NY: Academic Press, pp. 21–132.
Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., *et al*. (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc Natl Acad Sci USA* **110:** 14024–14029.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., *et al*. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.
Lee, Y.J., Hoynes-O'Connor, A., Leong, M.C., and Moon, T.S. (2016) Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Res* **44:** 2462–2473.
Ma, J., and Wang, M. (2014) Interplay between DNA supercoiling and transcription elongation. *Transcription* **5:** e28636.
Meng, J., Wang, B., Liu, D., Chen, T., Wang, Z., and Zhao, X. (2016) High-yield anaerobic succinate production by strategically regulating multiple metabolic pathways based on stoichiometric maximum in *Escherichia coli*. *Microb Cell Fact* **15:** 141.
Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: CSH Laboratory Press.
Nielsen, A.A.K., Der, B.S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E.A., *et al*. (2016) Genetic circuit design automation. *Science* **352:** aac7341.
Pasotti, L., Politi, N., Zucca, S., Cusella De Angelis, M.G., and Magni, P. (2012) Bottom-up engineering of biological systems through standard bricks: a modularity study on basic parts and devices. *PLoS ONE* **7:** e39407.
Portela, R.M.C., Vogl, T., Kniely, C., Fischer, J.E., Oliveira, R., and Glieder, A. (2017) Synthetic core promoters as universal parts for fine-tuning expression in different yeast species. *ACS Synth Biol* **6:** 471–484.
Presnell, K.V., Flexer-Harrison, M., and Alper, H.S. (2019) Design and synthesis of synthetic UP elements for modulation of gene expression in *Escherichia coli*. *Synth Syst Biotechnol* **4:** 99–106.
Qi, L., Haurwitz, R.E., Shao, W., Doudna, J.A., and Arkin, A.P. (2012) RNA processing enables predictable programming of gene expression. *Nat Biotechnol* **30:** 1002.
Rhodius, V.A., and Mutalik, V.K. (2010) Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, σ^E. *Proc Natl Acad Sci USA* **107:** 2854–2859.
Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. *Trends Genet* **16:** 276–277.
Ross, W., Aiyar, S.E., Salomon, J., and Gourse, R.L. (1998) *Escherichia coli* promoters with up elements of different strengths: modular structure of bacterial promoters. *J Bacteriol* **180:** 5375–5383.
Shimada, T., Yamamoto, K., and Ishihama, A. (2011) Novel members of the Cra regulon involved in carbon metabolism in *Escherichia coli*. *J Bacteriol* **193:** 649–659.
Urtecho, G., Tripp, A.D., Insigne, K., Kim, H., and Kosuri, S. (2018) Systematic dissection of sequence elements controlling σ70 promoters using a genomically encoded multiplexed reporter assay in *Escherichia coli*. *Biochemistry* **58:** 1539–1551.
Yim, S.S., An, S.J., Kang, M., Lee, J., and Jeong, K.J. (2013) Isolation of fully synthetic promoters for high-level gene expression in *Corynebacterium glutamicum*. *Biotechnol Bioeng* **110:** 2959–2969.
Zhou, L., Ding, Q., Jiang, G.-Z., Liu, Z.-N., Wang, H.-Y., and Zhao, G.-R. (2017) Chromosome engineering of *Escherichia coli* for constitutive production of salvianic acid A. *Microb Cell Fact* **16:** 84.
Zucca, S., Pasotti, L., Politi, N., Casanova, M., Mazzini, G., Cusella De Angelis, M.G., and Magni, P. (2015) Multifaceted characterization of a novel LuxR-repressible promoter library for *Escherichia coli*. *PLoS ONE* **10:** e0126264.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 List of plasmids and strains.
Table S2. The commonly used synthetic promoters consisting of only the core promoter region.
Table S3. List of primers.
Fig. S1. Isobutanol productions driven by the $P_L lacO_1$, *rrnB*p1 and J23100 promoters after 48 h of fermentation in M9 medium containing 40 g l^{-1} glucose.
Fig. S2. Construction of the β-gal expression plasmids with different upstream insertions.
Fig. S3. Truncation of the upstream insertions to fixed length.
Fig. S4. Maps for plasmids used in investigating the effects of the fixed-length upstream insertions on gene expression and chemical production.
Data S1. Nucleotide sequences for the inserted upstream fragments.

