Method for engineering artificial bidirectional promoters

Maxime Fages-Lartaud (✉ maxime.fages-lartaud@ntnu.no)

Norwegian University of Science and Technology

Weiyan Prone

Norwegian University of Science and Technology

Yasmin Mueller

Norwegian University of Science and Technology

Rahmi Lale

Norwegian University of Science and Technology

Research Article

Keywords:

Posted Date: February 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1266589/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Bidirectional promoters (BDPs) are ubiquitous as they facilitate gene co-expression within functional groups of genes or metabolic pathways. They help to balance enzymatic expression to maintain a stoichiometric relationship and avoid cellular toxicity. While BDPs offer several benefits over single-gene expression cassettes, functionally defined BDPs are scarce and usually unsuited for multi-gene expression in synthetic biology and metabolic engineering efforts. In this paper, we describe the fabrication of BDPs in bacteria using DNA with a random nucleotide composition. We created 230 artificial BDPs using a direct selection method (antibiotic resistance) and 168 gene-specific BDPs using a double fusion protein device (fluorescent reporters along with antibiotic resistance). The method's simplicity, gene-specificity, and universality make it an unprecedented tool for creating artificial BDP libraries representing a wide range of expression levels. This technique addresses the challenges in obtaining balanced co-expression of heterologous genes or pathways in various hosts. Furthermore, it provides insights that help to understand the architecture and features of BDPs.

Background

Genetic information contained in DNA is expressed through transcription into mRNA and its subsequent translation into proteins. In transcription, RNA polymerase holoenzymes (RNAP) recognize specific DNA sequences in the promoter regions and proceed to mRNA polymerization. Promoters contain several DNA sequence elements that are targets for interactions with sigma factors, co-factors and core RNA polymerase^{1,2}. In translation, ribosomes bind to specific mRNA motifs (*e.g.* Shine-Dalgarno or Kozak sequence) present within the 5' untranslated regions (5' UTRs) and translate the downstream coding sequences (CDSs)^{3,4}.

Promoters have long been thought to be unidirectional, that is, to drive transcription in a single direction, due to the arrangement of its core DNA components^{1,5}. However, pervasive transcription of antisense non-coding RNAs (ncRNAs) near promoters has been observed in a variety of taxa, including prokaryotes⁶⁻¹⁰, yeast¹¹⁻¹⁴, mouse^{15,16}, fruit fly¹⁷, plants¹⁸ and humans^{19,20}. The antisense transcription of ncRNAs surrounding promoters has been proposed as one of the genomic regulatory mechanisms²¹, indicating that bidirectionality appears to be a fundamental characteristic of promoters¹². While some short antisense transcripts originating from unidirectional promoters are stable, the majority are rapidly degraded^{14,22}. In contrast, a family of so-called bidirectional promoters (BDPs) not only produce mature transcripts in both orientations, but both transcripts also code for proteins^{23,24}.

BDPs were identified in the human genome as transcribing a divergent gene pair with transcription start sites (TSS) separated by less than 1,000 base pairs^{23,24}. BDPs are involved in the transcription of 10% of human genes²⁴ and 20% to 25% of genes in *E. coli*^{9,25}. In eukaryotes, BDPs show characteristic chromatin modification^{26,27}, and in prokaryotes, super-helical deplete signatures associated with transcription initiating regions²⁵. Interestingly, BDPs were linked to the requirement to coordinately

express a gene pair within a functional group^{23,28}. For instance, to balance the co-expression of genes belonging to a certain metabolic pathway²⁹⁻³¹, to maintain a stoichiometric relationship (e.g. histones genes)³²⁻³⁴, to respond to heat shock induction³⁵, or timing expression during the cell cycle³⁶.

Inspired by the benefits offered by the aforementioned natural systems, BDPs have been used to co-express heterologous genes in *E. coli*³⁷, filamentous fungi³⁸, yeast^{39,40}, plants^{41,42} and mammalian cells⁴³⁻⁴⁵ in various synthetic biology-related research efforts. Despite the numerous efforts, co-expression of several enzymes from the same metabolic pathway is difficult to achieve, as transcription and translation must be coordinated to ensure the balanced expression of multiple genes. Additionally, translation optimization is also required to fine-tune the cumulative protein yield, relative enzymatic activity, and the regulation of these factors⁴⁶⁻⁵⁰. Otherwise, the host may face growth deprivations as a result of burdensome protein over-expression or accumulation of toxic intermediate metabolites, leading to systemic failure. To avoid such detrimental consequences, it is critical to adjust and balance the co-expression of genes involved in the same metabolic pathway. There are currently no techniques in synthetic biology that allow for the rational creation of entire promoters and 5' UTRs (5' regulatory sequences, RES). In the absence of such tools, existing techniques rely mostly on the combinatorial assembly of pre-characterized 5' RESs employed in series to express CDSs⁵¹.

BDPs provide various advantages, such as lowering the number of genetic parts required, simplifying screening experiments, jointly profiting from the benefits of local DNA architecture, while providing solutions to enzyme co-expression and metabolic engineering challenges. BDPs have been cleverly employed to improve the stability of a transgene by coupling its transcription with an essential gene³⁷. While natural BDPs are functional in their native cellular host, they are not always adaptable to heterologous gene sequences and do not provide tunable expression. Examples of synthetic biology efforts involving BDPs focused on creating a handful of BDPs and combining their core elements³⁷⁻⁴⁵. To realize the promise of BDPs, larger and more diverse libraries covering a wide range of expression levels have to be created. So far, the only large library available was created in yeast using a combinatorial strategy to obtain a total of 168 BDPs³⁹. Even though such a combinatorial strategy allows for a wide range of expression, the libraries lack genetic diversity and adaptation to heterologous genes.

Complementarity between genetic components is critical, particularly between the 5' UTRs and the CDSs. Efficient translation is heavily reliant on mRNA structures formed within the 5' UTR and in proximity to the CDS⁵²⁻⁵⁵. When coupled to a new CDS, the utilization of a small set of conventional regulatory DNA parts may not offer efficient expression for instance due to formation of undesired mRNA structures. Computational techniques for predicting 5' UTR sequences specific to downstream CDSs have been established⁵⁵⁻⁵⁷, however, never in a bidirectional context. As a result, there are currently few strategies that allow for the balanced co-expression of heterologous genes in a metabolic pathway in different hosts.

We previously demonstrated the use of 200 nucleotides with random DNA composition for the construction of artificial gene expression systems in seven different microorganisms spanning from bacteria to yeast^{58,59}. The random DNA approach provides a straightforward and efficient option for simultaneously addressing transcriptional and translational expression constraints. In *E. coli*, 30 to 40% of random DNA sequences yielded successful protein expression while providing a wide range expression profiles⁵⁸. This approach has various advantages, it can be used in generating new artificial gene expression elements specific to the targeted CDS, while the expression range provides candidates with various cellular toxicity profiles. Additionally, the universality of the method allows it to be used in a variety of organisms without the need for redesigning genetic expression components.

In this study, we present and discuss how we exploited random DNA libraries to generate bidirectional 5' RES in *E. coli*. We propose a reliable universal approach for producing artificial BDPs, as well as characterized DNA components that should be of interest to the synthetic biology community. The method can also be used to obtain crucial insights into the properties of BDPs, to generate organism- and gene-specific DNA parts, and to find applications in metabolic pathway engineering and heterologous protein co-expression.

Results

Before we present our findings, we would like to clarify the nomenclature used to describe the bidirectional gene expression. Because the term bidirectional promoter omits the process of bidirectional translation, we refer to these genomic regions as Bidirectional Gene Expression Sequences (BiGESSs) rather than BDPs.

The construction of bidirectional regulatory sequences.

In a previous study, a random sequence of 200 nucleotides was shown to be capable of generating functional Gene Expression Sequences (GESSs) (promoter and 5' UTR) to drive the expression of various CDSs in a range of microorganisms⁵⁸. In this study, we sought to assess whether random DNA sequences could lead to expression of genes in both sense and antisense orientation simultaneously. To test the idea, we constructed two libraries by cloning random library of 50 (50N) and 200 (200N) nucleotides between the start codons of two genes facing opposite directions. We used the original 50N and 200N random DNA sequences from the Lale *et al.* study⁵⁸ that contained an upstream 5'-TGCC-3' Bsal scar and a downstream 5'-NATG-3' Bsal scar (See *Supplementary Figure S1*). In this study, for the construction of both plasmid DNA libraries, the upstream 5'-TGCC-3' Bsal scar of the random DNA fragments were modified to 5'-CATN-3' (5'-NATG-3' on the reverse strand) to match the start codon on the reverse strand. We used a two-step PCR method to generate the double 5'-NATG-3' Bsal scar on both sides of the DNA library and cloned it into the receiving plasmids (pSK and pSKGC) by Golden Gate assembly (see *Methods* and Fig. 1a). The Golden Gate assembly mixture was transformed into competent *E. coli* cells and the transformants were plated on agar plates containing 100 µg/mL ampicillin to determine the

total number of colonies in the library (library size). The transformation yielded between 5,000 and 10,000 colonies per 20 μ L of Golden Gate assembly mixture. Sanger DNA sequencing revealed the proper insertion of the random DNA. We were able to insert the modified 50N or 200N fragment into the pSK plasmid to create a large library. We could then test the functionality of these DNA sequences for bidirectional gene expression activity.

Direct selection method for generating bidirectional 5' regulatory sequences.

We created two plasmid DNA libraries by cloning random DNA sequences (50N and 200N) between two antibiotic resistance genes, *aadA* and *nptII*, conferring resistance to spectinomycin and kanamycin, respectively, to enable a functional screening (Fig. 1b). The resulting plasmid libraries were named pSK-50N and pSK-200N based on the length of the random DNA insert. The control plasmid, named pSK-Gibson, containing a fixed DNA sequence derived from cloning sandwiched between the two CDSs (see *Methods* and *Supplementary Figure S4a*). The transformation of pSK-Gibson to *E. coli* competent cells did not provide any colonies on agar plates containing 25 μ g/mL kanamycin, rare clones on 25 μ g/mL spectinomycin and no colonies on plates containing both antibiotics (25 μ g/mL kanamycin and spectinomycin). We performed the selection of functional BiGESs for the two plasmid libraries pSK-200N and pSK-50N on agar plates containing both antibiotics.

Given that 30 to 40% of the random 200N DNA library drives gene expression in one direction⁵⁸, one could expect approximately 9–16% of random sequences to generate bidirectional gene expression (0.3×0.3 and 0.4×0.4). However, the fraction of random DNA sequences producing bidirectional gene expression was estimated to 1.60% for 200N and 0.74% for 50N of the library size based on library colony counts (see *Supplementary Table S2*). While 55% of the pSK-200N library led to spectinomycin resistant clones, just 4.6% of the same library led to kanamycin resistant clones. As for the pSK-50N library, only 0.74% of the library was leading to functional expression of both antibiotic markers. The efficacy of the approach is clearly reliant on the length of the random DNA insert, the gene of interest and the antibiotic concentrations employed for selection, which explains why the projected predictions fall short of the experimentally determined number of clones.

To determine the strength of the BiGESs expression, 192 colonies from the pSK-200N and 48 colonies from the pSK-50N libraries were grown overnight in 96-well plates in LB supplemented with 25 μ g/mL spectinomycin and kanamycin. After 100-time dilution, cells were spotted on agar plates supplemented with increasing antibiotic concentrations (25 to 400 μ g/mL spectinomycin and kanamycin) to determine their resistance profile. Figure 2 shows the maximum concentration of spectinomycin and kanamycin that each clone sustained. For pSK-200N colonies (Fig. 2a), the gradient was started at 50 μ g/mL spectinomycin and kanamycin to avoid potential false positives. From the pSK-200N library, we obtained 9 clones resistant to 50 μ g/mL, 47 clones resistant to 100 μ g/mL, 92 clones resistant to 200 μ g/mL, and 44 clones resistant to 400 μ g/mL of spectinomycin and kanamycin. None of the clones grew on higher

concentrations of antibiotics tested (500 to 1,000 µg/mL). Whereas from the pSK-50N library, we obtained 12 clones resistant to 25 µg/mL, 9 clones resistant to 50 µg/mL, 10 clones resistant to 100 µg/mL, and 7 clones resistant to 200 µg/mL of spectinomycin and kanamycin and 10 clones failed to grow. As expected, a shorter random sequence of 50 nucleotides yielded less colonies than 200 nucleotides.

Lale *et al.*⁵⁸ shown that multiple transcription start sites occur within 200N sequences; thus, the longer the random sequence, the higher the probability to discover clones that lead to functional expression of both genes of interest. The results demonstrate the ability of a completely random DNA sequence of 200 or 50 bp to drive successfully bidirectional gene expression. This constitutes a rapid and efficient method to generate large amounts of unique, artificial BiGESs. However, because transferring genetic parts to a different context can affect their functionality, we sought to develop a device that allows for the selection of functional BiGESs as well as adaptation to any CDS.

Genetic device for the creation of gene-specific bidirectional 5' regulatory elements.

Since gene regulatory elements, particularly the 5' UTR and CDS, are highly reliant on genetic context, it is critical to be able to create BiGESs specific to a gene pair that cannot be utilized for direct selection. To do so, we created a second plasmid, pSKGC, in which the genes of interest are translationally fused with a respective antibiotic resistance gene to produce two selectable fusion proteins.

On one side, the super-folder green fluorescent protein (sfGFP) CDS was fused to the spectinomycin resistance gene via the P2A self-cleaving peptide (inactive in prokaryotes, only serves as a linker). On the other side, the red fluorescent protein (mCherry) CDS was fused to the kanamycin resistance gene by the T2A self-cleaving peptide (also inactive in prokaryotes, used as a linker) (See Fig. 1c). Thus, if a functional BiGES is generated between *sfGFP* and *mCherry*, it will result in two fusion proteins providing resistance to spectinomycin and kanamycin while allowing quantification of the promoter and 5' UTR strength by fluorescence intensity measurements.

Two pSKGC DNA libraries were created by cloning the 50 and the 200N sandwiched between the *sfGFP* and *mCherry* CDSs, which are situated on opposing DNA strands facing opposite directions. *E. coli* transformants were selected on 25 µg/mL spectinomycin and kanamycin. Positive clones were grown overnight in two 96-well plates, from which 167 out of 184 clones survived. About half of these colonies originated from pSKGC-50N and the rest from pSKGC-200N. Cells were diluted appropriately, and the intensity of GFP and mCherry fluorescence was measured for each strain (Fig. 3 and Fig. 4, respectively).

Clones from the pSKGC-200N library produced higher fluorescence signal for both GFP and mCherry. A first finding was that all 167 clones expressed mCherry than the control (pSKGC-Gibson, see *Methods* and *Supplementary Figure S4b*), and the intensity range expanded over two orders of magnitude. In contrast, the GFP signal range of the selected clones is limited to less than one order of magnitude.

Additionally, about a third of samples are close to the error range of the control. Since pSKGC-Gibson contains *sfGFP*, there was a possibility for leaky expression of the gene. However, it was ruled out by fluorescence microscopy analysis of pSKGC-Gibson (Fig. 5) and by running pSK-Gibson as a new control, giving similar GFP fluorescence values. Another possibility is that *sfGFP* contains elements that express the downstream *aadA* providing false selection on 25 µg/mL spectinomycin. Indeed, when plated on 25 µg/mL spectinomycin and kanamycin pSKGC-Gibson provided colonies, although no GFP fluorescence was detected and occasional slight fluorescence for mCherry was observable (see control Fig. 5). It points out, either the possibility of internal promoters and/or 5' UTRs within pSKGC genes providing antibiotic resistance, or that the control sandwich sequence acts as gene expression elements (*Supplementary Figure S4b*), or a combination of both. However, this artifact was abolished after inserting the random library, indicating that the control sequence was involved to some extent. Nevertheless, it raised concerns about the actual functionality of some of the 167 BiGES. Therefore, we selected 55 clones for further characterization under a fluorescent microscope (clones with asterisks on Fig. 3).

Confirmation of functional BiGESs by fluorescence microscopy.

Since GFP signals from some of the clones were in close range with the control, we investigated if GFP was actually expressed in all the 55 selected clones with a fluorescence microscope. In addition, we needed to confirm that GFP and mCherry were expressed in the same cell. The selected colonies are marked with an asterisk on Fig. 3. We found that eight clones seemed to express only mCherry, four samples had only a few BiGES-expressing cells mixed with mCherry-expressing cells, a couple samples were composed of two cell populations expressing GFP and mCherry separately, and the others were expressing both GFP and mCherry (Fig. 5 and *Supplementary Figure S5*). Interestingly, two clones with low GFP expression were nevertheless confirmed as containing both fluorescent proteins indicating BiGESs functionality in the low GFP range. Overall, most of the clones presented active BiGESs indicating that the two fusion protein genetic device is a valuable system for generating gene-specific BiGES. The BiGESs from confirmed clones expressing both fluorescent proteins were sequenced. After cleaning duplicates, 32 clones with unique BiGESs were identified (for sequences: see *Supplementary Table S3*). Their respective fluorescence profile is presented in Fig. 6. Interestingly, confirmed BiGESs present different expression intensity on each DNA strand. For example, BiGES_29 presents strong GFP and mid mCherry intensities while for BiGES_09 it is the opposite, low expression of GFP and high mCherry. The results presented so far, satisfies metabolic engineering requirements by producing a wide range of expression and varied expression ratios for the two coupled enzymes.

The assessment of bidirectional regulatory elements transfer to other genes.

To assess if BiGESs can be transferred to a new gene pair while conserving activity, we extracted twenty BiGESs from pSK direct selection onto pSKGC and monitored GFP and mCherry fluorescence levels. The extracted BiGESs of pSK-200N and pSK-50N are indicated on Fig. 2 with orange squares. For each pSKGC clones generated with a transferred BiGES, ten colonies were grown in 96-well plate for fluorescence analysis. Fluorescence profiles of the twenty clones are shown in Fig. 7. Interestingly, four clones presented two equally represented distinct fluorescent profiles, indicated with stripped lines on Fig. 7. All tested clones exhibited mCherry fluorescence and at least half of them expressed GFP significantly in comparison to the control. The relatively high GFP value for the control comes from an artifact of normalization by OD₆₀₀. However, nine out of 24 clones showed very low mCherry expression, ten had mid-expression and five were intense. Since BiGES transfer is non-directional (matching 5'-NATG-3' BsaI scar), the emergence of two populations after some transfers can show either a failure to drive expression in one direction, a non-compatibility of the 5' UTR with one of the gene, or a strength difference of the promoters and 5' UTRs present in a particular BiGES. Another explanation is that the selected pSK strain sustained two or more plasmid versions since sequence differences are minimal. Thus, during the extraction and transfer process, several BiGES sequences are present, generating clones with different expression profiles. Noteworthy, all BiGESs transferred from pSK-50N (pSK_T 16 to 20) displayed low to mid GFP and mCherry signals. Either it can indicate the dependence on the 5' gene sequences or that some elements of gene expression are not comprised within the 50 nucleotides sequence. Overall, it seems like BiGESs can be transferred from the direct selection method to a new gene pair successfully. Although, the risk of losing functionality or changing expression intensity exists and is not negligible.

Sequence analysis to identify sigma factor binding motifs.

With the intention to confirm the presence of bidirectional promoters within the BiGES sequences, we used BacPP, a sigma factor binding motif prediction software⁶⁰. BacPP is able to predict the presence of *E. coli* promoters for six sigma factors (Sigma 24, 28, 32, 38, 54 and 70) with an accuracy ranging from 83–97%. All of the confirmed 32 BiGESs generated with pSKGC were analyzed on both forward and reverse strand (see Fig. 8 and *Supplementary Figure S6*). For visualization purposes, results were translated into a color gradient with a threshold at 75% prediction. Since the analysis uses an 80 bp sliding window, 80 base pair of either *sfGFP* or *mCherry* were included to capture the impact of each nucleotide from the complete inserted sequences.

To control for the presence of internal promoters within the plasmid, genes encoding the two fusion proteins were analyzed (see *Supplementary Figure S7 & S8*). On the forward strand, we found no significant promoter prediction within *aadA* or *T2A* that could provide a transcript for *mCherry-P2A-nptII*, nor within *mCherry* that could transcribe *nptII* and could result in kanamycin resistance. However, there is an island of strong prediction within the GFP gene in the forward direction (-455 to -396 nucleotides from the start of *sfGFP*). This could lead to a transcript containing the downstream fluorescent protein sequence. The generation of an in-frame ribosome binding site before mCherry could lead to translation

of the mRNA into mCherry-P2A-nptII. It could also explain the kanamycin resistance and the low red fluorescence of the control plasmid pSKGC-Gibson. On the reverse strand, no significant predictions were identified within *nptII-P2A-mCherry*, apart from a slight increase close to the control sandwich sequence (BsaI scars) of pSKGC-Gibson. This could explain the spectinomycin resistance found for the control plasmid. However, once again within *sfGFP*, we stumbled upon a high promoter prediction potentially transcribing toward *aadA* (-458 to -531 bp from the start of *sfGFP*). Thus, spectinomycin resistance can be achieved if ribosomes bind the 5' UTR and produce an in-frame protein. Additionally, several in-frame methionine are present before the start of *aadA* and could lead to a slightly elongated but still functional spectinomycin resistance protein. These two internal potential promoters within the *sfGFP* sequence interfere with the selection of BiGES solely contained within the random sequence insert.

In order to estimate the significance of BacPP predictions for random sequences, we generated ten random sequences of 1000 bp with a python program that were subjected to promoter motif analysis (see *Supplementary Figure S9*). The predictions obtained should represent the baseline predictions of biologically meaningless sequences. For each sigma factors, the frequencies per 100 nucleotides of prediction above 75% were calculated, and a Student's t test was applied to determine a significance threshold for the predictions (see *Supplementary Table S4*). Most promoter predictions for BiGES sequences were above the significance thresholds; however, for sigma 28, 34 and 54 predictions above 75% remained rare. On the contrary, for sigma 24, 38 and 70, occurrence of high promoter predictions were substantial (see *Supplementary Figure S10* and *Table S5 & S6*). The 80 bp sliding window of the BacPP analysis allows identifying nucleotide islands of high promoter predictions for each BiGESs in both directions (see Fig. 8 and *Supplementary Figure S6*). For BiGESs generated with 50N (exemplified by BiGES_01 in Fig. 8), it appears that promoters are predicted on the reverse strand but rarely on the forward strand, except sometimes close to the start of *mCherry*. This absence may indicate that the seven 50N BiGESs analyzed do not actually contain all the 5' genetic expression elements within the 50 nucleotides and therefore can only be used in this particular genetic context and not for another gene pair. The possible internal promoter within the *sfGFP* sequence may produce a viable transcript of the downstream mCherry-T2A-nptII fusion protein; therefore, it may explain the absence of promoter prediction within the 50N sequence. For BiGESs generated with 200N, a few sequences show mediocre or very localized predictions when the others present highly conclusive promoter predictions (see Fig. 8 and *Supplementary Figure S6*). The promoter motif analysis permits to confirm or infirm the presence of gene expression elements within the BiGES sequences.

Discussion

In this study, we presented a rapid and efficient method to fabricate bidirectional promoters for gene co-expression. We build on the previous work of Lale *et al.* that used a library of random DNA sequences to provide a universal gene expression method⁵⁸. The surprisingly high proportion of random sequences driving gene expression in one direction (30–40%)⁵⁸, plus the occurrence of multiples TSSs within a

sequence, had lead us to hypothesize a reasonable probability for this event to occur also on the reverse strand. However, we obtained a lower proportion of functional BiGESs from 200N random sequences than expected (1.6% instead of 9–16%). The relatively low proportion of sequences providing BiGESs is in part related to the experimental design (antibiotic selection), but could also originate from the difficulty to generate the required DNA motifs on both strands, or a crowded transcriptional environment or RNA interference between the DNA strands. Despite the lower efficiency, we demonstrate that the gene expression method based on random sequences is applicable to bidirectional gene expression. Overall, it reinforces the idea that bidirectionality constitute an inherent feature of promoters¹⁴, that motifs are shared on opposite strands⁹ and stable bidirectional transcripts occur for a fraction of natural promoters^{9,23–25}.

The direct selection method is an efficient way to generate BiGESs libraries for various organisms. These artificial genetic parts present a wide range of expression levels as exemplified with different fluorescent intensities obtained. It is essential to answer issues related to stoichiometric relationship in enzyme co-expression and metabolic engineering^{46–50}. Additionally, it decreases the number of genetic parts and the screening required in combinatorial approaches. One drawback of the direct selection method is the potential reliance of BiGESs on the original genetic context. The first issue is that some 5' regulatory elements might not be comprised within the random sequence. As evidenced by the BiGES transfer experiment and the motif analysis, the occurrence of internal promoters and 5' UTRs within the gene pair of interest may be sufficient to express the downstream gene. Hence, some BiGESs may be functional only in the genetic context of their creation and cannot be transferred to another pair, especially for small DNA fragments such as 50N. Therefore, to generate a BiGES library, our data suggest using a longer DNA fragment, such as 200N, to include all elements necessary for bidirectional gene expression within the sequence. In addition, the gene pair used to create a BiGES library should be carefully selected to ensure the absence of internal gene expression elements such as promoter- and SD-like sequences or internal methionine codons. The second contextual issue relates to the complementarity between 5' UTRs and CDSs that form mRNA secondary structures conditioning mRNA stability and ribosome entry-sites accessibility^{52–55}. As observed with BiGES transfers, the level of gene expression can vary when BiGES are transferred to a different gene pair. The benefits of a diverse library is the ability to provide other 5' UTRs that may be more structurally appropriate for the new CDSs.

In order to overcome complementarity issues, we developed the gene-tailored approach to create specific and/or quantifiable BiGESs for any gene pair. We positioned the selection markers after the genes of interest to place the BiGESs in the intended context and we performed the selection on mature fusion proteins. The genes encoding fluorescent proteins in the plasmid pSKGC can be swapped for a gene pair of interest and select functional gene-tailored BiGESs by antibiotic resistance. However, as illustrated by the internal promoters in *sfGFP*, false selection may occur because of internal elements expressing downstream genes. Therefore, a careful selection of the genes composing the fusion proteins along with their codon usage is required to avoid expression originating internally in one of the directions.

Nevertheless, this approach demonstrates a unique way to design gene-specific BiGESs valuable for tuned enzyme co-expression and difficult to express gene pairs.

The promoter motif analysis performed with BacPP⁶⁰ allows localization of bidirectional regulatory elements. For BiGESs generated with 50N, the failure to predict a promoter on the forward strand could be consequence of the internal promoter found in the forward strand within *sfGFP*. This way, the transcript of mCherry-P2A-nptII would be created and ribosomes may find an entry point within the 5' UTR to translate the fusion protein. This exemplifies one of the genetic context dependencies evoked previously. The seven 50N-BiGESs analyzed may not contain all the genetic expression elements within the 50 nucleotides. Therefore, they cannot be utilized in another genetic context for another gene pair. For BiGESs generated with 200N, predictions are more conclusively localized within the random sequence. Interestingly, most islands of high predictions are shifted between the forward and reverse strands. Indeed, most promoters are identified closer to their respective gene (such as in BiGES_26 and BiGES_32 in Fig. 8). Although BDPs function as inseparable units because of sharing common elements^{9,24}, it seems as if sigma factors have different binding sites on each strands. It could be a result of expression optimization by avoiding RNA polymerase crowding and collision⁶¹ or deleterious antisense mRNA interference. Although some BiGESs analyzed, for example BiGES_22 and BiGES_32, possess islands of high prediction common to both strands. Within the set of analyzed sequences, we did not find conclusive evidence of overlapping – 10 box but prediction results suggest the possibility of overlapping – 35 boxes or UP elements. From the small subset of analyzed BiGES, it is not possible to draw conclusions regarding motif sharing in these systems. However, the method presented in this paper allows for generation of large libraries that can be useful in providing insights into BiGESs features.

The method presented in this paper is the first synthetic biology tool that can generate artificial sequences for bidirectional gene expression. The benefits of the method such as adaptability to the coding sequence, organism-specificity and the wide range of expression obtained with BiGES libraries are substantial assets for genetic engineering endeavors. We believe that this new tool will be helpful for enzyme co-expression, metabolic engineering and to gain insights into bidirectional promoter features.

Materials And Methods

Reagents, strains and media.

Experiments were carried out in *Escherichia coli* DH5α ordered from New England Biolabs. Oxoid growth media components were purchased from ThermoFisher Scientific. Cells were grown in LB-Lennox (10 g/L casein peptone, 5 g/L yeast extract, 5 g/L NaCl and supplemented with 15 g/L agar for LA plates) supplemented with the appropriate antibiotics (Sigma-Aldrich). All enzymes used in this study were purchased from New England Biolabs, except CloneAmp HiFi PCR premix that was ordered from Takara Bio. Primers were purchased from either Sigma-Aldrich or Eurofins Genomics (for primer list see *Supplementary Table S1*). All PCR reactions were performed with Q5 polymerase from NEB unless mentioned otherwise. PCR products and plasmid DNA purifications were performed with QiaQuick PCR

purification kit and QIAprep plasmid Miniprep kit respectively (Qiagen). Sanger DNA sequencing was performed by Eurofins Genomics.

Random DNA library generation.

Two single strand DNA fragments of 50 and 200 random nucleotides (50N/200N) were synthesized by Integrated DNA Technology (Louvain, Belgium). The ssDNA contained a fixed upstream and downstream sequence necessary for PCR and restriction cloning (see *Supplementary Figure S1*). The original 200N and 50N random sequences used by Lale et al.⁵⁸ contained an upstream 5'-TGCC-3' Bsal scar and a downstream 5'-NATG-3' Bsal scar. We wished to adapt this existing library for bidirectional gene expression, so the DNA fragment needs to match the start codons of both forward and reverse genes. The upstream Bsal scar was changed from 5'-TGCC-3' to 5'-CATN-3' (to confer 5'-NATG-3' on the reverse strand) by a two-step PCR using Q5 DNA polymerase. The first step was carried out using only the reverse primer (MFL 26) to generate the complementary single strand in excess. The ssDNA product was subsequently purified and served as template for the second PCR step with forward (MFL 167 or 247) and reverse primers (MFL 26). The two forward primers (MFL 167 and 247) contain the new Bsal scar 5'-CATN-3' and 5'-CATC-3' followed by 4N and 3N respectively (see *Supplementary Figure S2*) introducing a methionine codon to serve as start codon on the reverse strand. The two-step PCR allowed yielding a product containing in majority an upstream 5'-NATG-3' scar on the reverse strand. PCR products were purified and subsequently used for plasmid library construction (see *Supplementary Figure S3*).

Plasmid construction and selection of bidirectional regulatory elements.

Two plasmids were constructed to assert the ability of the random DNA library to provide bidirectional promoters/5' UTRs. Both plasmids were constructed on a pUC57 backbone ordered from Genscript. The plasmid contains an ampicillin resistance marker (*bla*) and a pMB1 derivative origin of replication. The direct-selection plasmid (pSK) contained spectinomycin and the kanamycin resistance genes on opposite strands (Fig. 1b). The gene encoding spectinomycin resistance (*aadA*) was amplified by PCR from the biobrick part BBa_K125520 (http://partsregistry.org/Main_Page) using primers MFL 35 and 128. The kanamycin resistance gene (*nptII*) was amplified by PCR from the biobrick part BBa_K1825005 with primers MFL 34 and 121. The backbone was amplified by PCR in two parts with primers MFL 17, 18, 32 and 33 in order to delete an internal Bsal site in the ampicillin resistance gene during the assembly. The two PCR products obtained from the backbone were treated with DpnI overnight at 37°C. All fragments were purified, mixed in equimolar ratio and assembled together by Gibson cloning (1h incubation at 50°C, total volume 20 µL)⁶². Half of the Gibson mixture was chemically transformed into competent *E. coli* DH5-α (heat shock at 42°C for 45 s) and plated on LA supplemented with 100 µg/mL ampicillin. Correct assembly was confirmed by colony PCR (primers MFL 34 & 35) performed with Taq DNA polymerase. Positive clones were grown overnight at 37°C in 5 mL LB media, their respective plasmids were purified by

Miniprep and sequenced by Sanger sequencing (primers MFL 55 & 245). The resulting plasmid, named pSK-Gibson, possesses the spectinomycin and kanamycin resistance genes facing opposite directions separated with a fixed sandwich sequence containing the BsaI sites (see *Supplementary Figure S4a*). This plasmid was used as negative control when screening for bidirectional regulatory elements with direct selection. It was plated on LA supplemented with either 25 µg/mL kanamycin or 25 µg/mL spectinomycin, or both antibiotics.

The plasmid for generating gene-specific BiGESs (pSKGC) encoded two fusion proteins, sfGFP-P2A-aadA and mcherry-T2A-nptII, in opposite directions (Fig. 1c). The *sfGFP* gene was PCR amplified from the biobrick (BBa_I746916) with primers MFL 65 & 173. The *mCherry* gene was amplified by PCR from pLT101⁵⁸ with primers MFL 174 & 175. The acceptor backbone was amplified from pSK-Gibson with primers MFL 172 & 176 (deleting the start codon of *aadA* and *nptII*) and digested with DpnI overnight at 37°C. After purification, the fragments were mixed in a 1:3 vector:insert molar ratio with 75 ng of vector and assembled by Gibson cloning⁶². As for the construction of pSK-Gibson, the Gibson mixture was chemically transformed into *E. coli* DH5-α and selected on LA supplemented with 100 µg/mL ampicillin. Colony PCR with Taq DNA polymerase (primers MFL 34 & 35) was performed on transformants to confirm correct insertion of DNA fragments. Colonies were grown overnight at 37°C in 5 mL LB media, their respective plasmids were purified and sequence-confirmed by Sanger sequencing (primers MFL 177, 245 and ION-mCh). The created plasmid, named pSKGC-Gibson, contained two encoded fusion proteins facing opposite directions separated with a fixed sandwich sequence containing the BsaI sites (see *Supplementary Figure S4b*). This plasmid was used as a negative control in the fluorescence screening for bidirectional regulatory elements. It was plated on LA supplemented with either 25 µg/mL kanamycin or 25 µg/mL spectinomycin, or both antibiotics.

In order to insert the DNA library, the pSK-Gibson and the pSKGC-Gibson plasmids were amplified by PCR using CloneAmp HiFi PCR premix with primers MFL 121 & 128 and MFL 55 & 174 respectively. The PCR products were subject to overnight digestion with DpnI at 37°C and were subsequently purified. Both PCR products were mixed with the 50N or 200N DNA library fragment in a 1:5 vector:insert molar ratio with 75 ng of vector and assembled by Golden Gate assembly⁶³; (60 cycles of 5 min at 37°C and 5 min at 16°C, followed by a final inactivation step at 65°C for 20 min, total volume 20 µL). Then, 10 µL of the Golden Gate assembly mixture were chemically transformed into competent *E. coli*. The transformants were spread on selection LA plates containing 25 µg/mL kanamycin and 25 µg/mL spectinomycin, plus on one plate containing 100 µg/mL ampicillin for assembly quality control. Colonies from the ampicillin plate and from the kanamycin/spectinomycin plate were grown overnight at 37°C in 5 mL LB media with the respective antibiotics. The plasmids from these clones were purified by Miniprep and Sanger sequenced (primer MFL 245 for pSK and MFL 177 for pSKGC) to confirm the correct insertion of either the 50N or the 200N DNA fragment. The cloning process was iterated three times to yield enough candidates for further analysis.

Bidirectional regulatory elements characterization.

Colonies isolated on 25 µg/mL kanamycin and spectinomycin for pSK and pSKGC libraries were transferred to Nunclon 96-well plates (Thermo Scientific) containing 100 µL of LB supplemented with 25 µg/mL kanamycin and 25 µg/mL spectinomycin. The pSKGC-Gibson control was grown in the first three wells of the 96-well plate in LB with 100 µg/mL ampicillin. The 96-well plates were incubated overnight at 37°C and 800 rpm in a Multitron Pro plate-shaking incubator (Infors HT).

The pSK library cells grown in 96-well plate were diluted 100 times in new 96-well plates containing LB supplemented with 25 µg/mL kanamycin and 25 µg/mL spectinomycin. Diluted cells were then replica plated using a 96-pin replicator on 15 cm large LA plates containing 25/25, 50/50, 100/100, 200/200, 400/400 µg/mL kanamycin/spectinomycin. The pSK-Gibson control plasmid was grown overnight separately in 5 mL LB supplemented with 100 µg/mL ampicillin, diluted 100 times and three spots of 2 µL were made on the gradient LA plates. Plates were incubated overnight at 37°C and pictures were taken with a Canon camera EOS M.

The pSKGC library cells were diluted four times into black 96-well plates with transparent bottom (Thermo Scientific). GFP and mCherry fluorescence was measured with an Infinite M200 Pro TECAN fluorimeter (Noax Lab AS) with excitation/emission wavelength couples of 488/525 nm and 576/610 nm respectively. Fluorescence measurements were normalized by OD₆₀₀ of the corresponding well.

Transfer of bidirectional regulatory elements.

Twenty clones generated with the pSK libraries were grown overnight at 37°C in 5 mL LB supplemented with 25 µg/mL kanamycin and 25 µg/mL spectinomycin. Their plasmids were purified and their respective BiGESs were extracted by PCR with primers MFL 314 & 315. Then, BiGESs were transferred to pSKGC by Golden Gate assembly and transformed into *E. coli* as described above. For each transfers, 10 colonies were grown into 96-well plates to monitor GFP and mCherry fluorescence as described previously.

Fluorescence microscopy.

For each clones analyzed, 1 µL of overnight culture was placed between glass slides for the microscopic analysis. An inverted microscope (Zeiss Axio Observer.Z1, 2.3.64.0) with a 20x air objective (NA 0.8) was utilized with GFP and mCherry filters to verify the presence of both fluorescent proteins. Image processing was done with Zeiss image analysis software (2.3.64.0).

Sigma factor motifs analysis.

clones with confirmed bidirectional regulatory elements from microscopic analysis were grown overnight at 37°C in 5 mL LB with 50 µg/mL kanamycin and 50 µg/mL spectinomycin. Their respective plasmids were purified by Miniprep and sequenced with primer MFL 177 (for sequences: see *Supplementary Table*

S3). Sequence analysis was performed with BacPP⁶⁰ (<http://www.bacpp.bioinfoucs.com/home>) to identify potential sigma factor binding sites. The analysis uses an 80 bp sliding window; hence, the BiGES sequences were uploaded with 80 bp from the upstream gene to allow complete analysis of all the nucleotides of the random DNA insert. The program provides predictions for promoter sequences for sigma factor 24, 28, 32, 38, 54 and 70. A color gradient was made to visualize the data. Predictions above 75% were considered as the threshold in the color visualization. The sequence encoding the fusion proteins (*GFP-P2A-aadA* and *mCherry-T2A-nptII*) were also analyzed to identify potential internal promoters. A set of 10 random sequences of 1000 nucleotides was generated with Python to serve as control (script: see *Supplementary Figure S11*). A one-tailed Student's t test was performed on the random sequence set (df = 9, IC = 95%, p = 0.05) to obtain a significance threshold for non-random occurrence of a certain motif within a 100N window (see *Supplementary Table S4*). The 95% threshold was considered as the significance limit for motif prediction of the BiGES generated with pSKGC. The distribution of predicted promoter values from BacPP were compared to the significance thresholds for each sigma factors (see *Supplementary Figure S10*).

References

1. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638–650 (2016).
2. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).
3. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208 (1999).
4. Shine, J. & Dalgarno, L. The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites (terminal labeling/stepwise degradation/protein synthesis/suppression). *Proc. Natl. Acad. Sci.* **71**, 1342–1346 (1974).
5. Butler, J. E. F. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16**, 2583–2592 (2002).
6. Dornenburg, J. E., Devita, A. M., Palumbo, M. J. & Wade, J. T. Widespread Antisense Transcription in Escherichia coli. *Am. Soc. Microbiol.* **1**, 1–4 (2010).
7. Yus, E. *et al.* Impact of genome reduction on bacterial metabolism and its regulation. *Science* (80-.). **326**, 1263–1268 (2009).
8. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250–255 (2010).
9. Warman, E. A. *et al.* Widespread divergent transcription from bacterial and archaeal promoters is a consequence of DNA-sequence symmetry. *Nat. Microbiol.* **6**, (2021).
10. Perkins, T. T. *et al.* A strand-specific RNA-seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi. *PLoS Genet.* **5**, (2009).

11. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (80-)*. **320**, 1344–1349 (2008).
12. Neil, H. *et al.* Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**, 1038–1042 (2009).
13. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci.* **4**, 5320–5325 (2006).
14. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).
15. Seila, A. C. *et al.* Divergent transcription from active promoters. *Science (80-)*. **322**, 1849–1851 (2008).
16. Kanhere, A. *et al.* Short RNAs Are Transcribed from Repressed Polycomb Target Genes and Interact with Polycomb Repressive Complex-2. *Mol. Cell* **38**, 675–688 (2010).
17. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011).
18. Li, L. *et al.* Genome-wide transcription analyses in rice using tiling microarrays. *Nat. Genet.* **38**, 124–129 (2006).
19. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
20. He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N. & Kinzle, K. W. The Antisense Transcriptomes of Human Cells. *Science (80-)*. **322**, 1855–1857 (2008).
21. Mattick, J. S. The genetic signatures of noncoding RNAs. *PLoS Genetics* vol. 5 (2009).
22. Wyers, F. *et al.* Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. *Cell* **121**, (2005).
23. Adachi, N. & Lieber, M. Bidirectional Gene Organization: A Common Architectural Feature of the Human Genome. *Cell* **109**, 807–809 (2002).
24. Trinklein, N. D. *et al.* An abundance of bidirectional promoters in the human genome. *Genome Res.* **14**, 62–66 (2004).
25. Masulis, I. S., Babaeva, Z. S., Chernyshov, S. V. & Ozoline, O. N. Visualizing the activity of *Escherichia coli* divergent promoters and probing their dependence on superhelical density using dual-colour fluorescent reporter vector. *Sci. Rep.* **5**, (2015).
26. Jangid, R. K., Kelkar, A., Muley, V. Y. & Galande, S. Bidirectional promoters exhibit characteristic chromatin modification signature associated with transcription elongation in both sense and antisense directions. *BMC Genomics* **19**, (2018).
27. Young, R. S., Kumar, Y., Bickmore, W. A. & Taylor, M. S. Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biol.* **18**, (2017).
28. Wakano, C., Byun, J. S., Di, L. J. & Gardner, K. The dual lives of bidirectional promoters. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**, 688–693 (2012).

29. Sugimoto, M., Oohashi, T. & Ninomiya, Y. The genes COL4A5 and COL4A6, coding for basement membrane collagen chains $\alpha 5(\text{IV})$ and $\alpha 6(\text{IV})$, are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. *Proc. Natl. Acad. Sci. USA* **91**, 11679–11683 (1994).
30. Momota, R. *et al.* Two genes, COL4A3 and COL4A4 coding for the human K3(IV) and K4(IV) collagen chains are arranged head-to-head on chromosome 2q36. *FEBS Lett.* **424**, 11–16 (1998).
31. Schmidt, C. *et al.* Differential effects of DNA-binding proteins on bidirectional transcription from the common promoter region of human collagen type IV genes COL4A1 and COL4A2. *Biochim. Biophys. Acta* **1174**, 1–10 (1993).
32. Albig, W., Kioschis, P., Poustka, A., Meergans, K. & Doenecke, D. Human Histone Gene Organization: Nonregular Arrangement within a Large Cluster. *Genomics* **40**, 314–322 (1997).
33. Ahn, J. & Gruen, J. R. The genomic organization of the histone clusters on human 6p21.3. *Mamm. Genome* **10**, 768–770 (1999).
34. Maxson, R., Cohn, R., Kedes, L. & Mohun, T. Expression and organization of histone genes. *Annu. Rev. Genet.* **17**, 239–277 (1983).
35. Hansen, J. J. *et al.* Genomic structure of the human mitochondrial chaperonin genes: HSP60 and HSP10 are localised head to head on chromosome 2 separated by a bidirectional promoter. *Hum. Genet.* **112**, 71–77 (2003).
36. Guarguaglini, G. *et al.* Expression of the murine RanBP1 and Htf9-c genes is regulated from a shared bidirectional promoter during cell cycle progression. *Biochem. J.* **325**, 277–286 (1997).
37. Yang, S., Sleight, S. C. & Sauro, H. M. Rationally designed bidirectional promoter improves the evolutionary stability of synthetic genetic circuits. *Nucleic Acids Res.* **41**, (2013).
38. Rendsvig, J. K. H., Workman, C. T. & Hoof, J. B. Bidirectional histone-gene promoters in *Aspergillus*: Characterization and application for multi-gene expression. *Fungal Biol. Biotechnol.* **6**, (2019).
39. Vogl, T. *et al.* Engineered bidirectional promoters enable rapid multi-gene co-expression optimization. *Nat. Commun.* **9**, (2018).
40. Öztürk, S., Ergün, B. G. & Çalık, P. Double promoter expression systems for recombinant protein production by industrial microorganisms. *Applied Microbiology and Biotechnology* vol. 101 7459–7475 (2017).
41. Xie, M., He, Y. & Gan, S. Bidirectionalization of polar promoters in plants. *Nat. Biotechnol.* **19**, 677–679 (2001).
42. Bai, J. *et al.* Comprehensive construction strategy of bidirectional green tissue-specific synthetic promoters. *Plant Biotechnol. J.* **18**, 668–678 (2020).
43. Fux, C. & Fussenegger, M. Bidirectional expression units enable streptogramin-adjustable gene expression in mammalian cells. *Biotechnol. Bioeng.* **83**, 618–625 (2003).
44. Amendola, M., Venneri, M. A., Biffi, A., Vigna, E. & Naldini, L. Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. *Nat. Biotechnol.* **23**, 108–116 (2005).

45. He, K., Rad, S. M. A. H., Poudel, A. & McLellan, A. D. Compact bidirectional promoters for dual-gene expression in a sleeping beauty transposon. *Int. J. Mol. Sci.* **21**, 1–9 (2020).
46. Pitera, D. J., Paddon, C. J., Newman, J. D. & Keasling, J. D. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* (2007) doi:10.1016/j.ymben.2006.11.002.
47. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
48. Kumaran Ajikumar, P. *et al.* Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. *Science (80-.)*. **330**, 70–74 (2010).
49. Tan, S. Z. & Prather, K. L. Dynamic pathway regulation: recent advances and methods of construction. *Current Opinion in Chemical Biology* vol. 41 28–35 (2017).
50. Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science (80-.)*. **349**, 1095–1100 (2015).
51. Tietze, L. & Lale, R. Importance of the 5' regulatory region to bacterial synthetic biology applications. *Microb. Biotechnol.* (2021) doi:10.1111/1751-7915.13868.
52. Espah Borujeni, A. & Salis, H. M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* **138**, 7016–7023 (2016).
53. Espah Borujeni, A. *et al.* Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res.* **45**, 5437–5448 (2017).
54. Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mRNAs. *Genome Biol.* **3**, 1–10 (2002).
55. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* **42**, 2646–2659 (2014).
56. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
57. Salis, H. M. The ribosome binding site calculator. in *Methods in Enzymology* vol. 498 19–42 (Academic Press Inc., 2011).
58. Lale, R. *et al.* A universal method for gene expression engineering. doi:10.1101/644989.
59. Tietze, L., Mangold, A., Hoff, M. W. & Lale, R. Identification and cross-characterisation of artificial promoters and 5' untranslated regions in *Vibrio natriegens*. *Front. Bioeng. Biotechnol. - Synth. Biol.* (2022).
60. de Avila e Silva, S., Echeverrigaray, S. & Gerhardt, G. J. L. BacPP: Bacterial promoter prediction-A tool for accurate sigma-factor specific assignment in enterobacteria. *J. Theor. Biol.* **287**, 92–99 (2011).
61. Hobson, D. J., Wei, W., Steinmetz, L. M. & Svejstrup, J. Q. RNA Polymerase II Collision Interrupts Convergent Transcription. *Mol. Cell* **48**, 365–374 (2012).

62. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
63. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, (2008).

Figures

Figure 1

a) Illustration of DNA library insertion into the receiving plasmids by Golden Gate assembly. b) Gene and plasmid maps of pSK used for direct selection of BiGESs with spectinomycin and kanamycin resistance genes. c) Gene and Plasmid maps of pSKGC used for designing gene-specific BiGESs with *sfGFP* and *mCherry* as reporters, fused to spectinomycin and kanamycin genes respectively for selection. The core elements of the pUC8 backbone are represented on the plasmid maps: ampicillin resistance gene (gray) and origin of replication (light pink).

Figure 2

Color representation of the maximum antibiotic concentration sustained for a) two 96-well plates with pSK-200N colonies and b) 48 colonies with pSK-50N. On the right-hand side, legend shows color correspondence to concentrations of spectinomycin and kanamycin in $\mu\text{g/mL}$. Orange squares designate BiGES that were later transferred to pSKGC.

Figure 3

GFP fluorescence normalized by OD₆₀₀ for the 167 selected BiGES clones (log scale). Standard deviations from biological triplicates are shown in red for each strain. Control without fluorescent protein (pSK-Gibson) is shown in position 1 (P1-A1). Asterisks represent clones analyzed under fluorescence microscope, green: confirmed GFP, black: no GFP detected, yellow: only a few BiGES-containing cells and a larger red-only fluorescent population.

Figure 4

mCherry fluorescence normalized by OD₆₀₀ for the 167 selected BiGES clones (log scale). Standard deviations from biological triplicates are shown in black for each strains. Control without fluorescent protein (pSK-Gibson) is shown in position 1 (P1-A1). All the clones visualized under microscope showed red fluorescence.

Figure 5

Examples of the different expression levels observed under fluorescence microscopy. The control pSKGC-Gibson, showing slight red fluorescence; a confirmed BiGES strain co-expressing GFP and mCherry (sample P2-2A => BiGES_21) and a sample containing two populations expressing GFP and mCherry separately (P2-5D). The fluorescence microscopy analysis of the other BiGES clones is presented in *Supplementary Figure S5*.

Figure 6

GFP and mCherry fluorescence of the 32 BiGES clones selected with pSKGC and confirmed with fluorescence microscopy. Standard deviations from biological triplicates are shown in red for GFP and in black for mCherry.

Figure 7

GFP and mCherry fluorescence of the 20 transferred BiGESs from pSK to pSKGC. Dashed lines show the clones presenting two distinct populations. Standard deviations from biological triplicates are shown in red for GFP and in black for mCherry.

Figure 8

Color representation of the sigma factor motif predictions from BacPP for three BiGESs. The six sigma factors are indicated on the side. Predictions were made for the forward and reverse strand (indicated with F and R). Sequences are displayed for each strand above the predictions. Scale is represented with a threshold was set at 75% (white) toward increasing red color for higher percentages. The length and number of each BiGES is indicated on the lefthand corner of each box. For controls and other samples see *Supplementary Figure S6*.