Synthetic promoter design for new microbial chassis

James Gilman* and John Love*1

*Biosciences, College of Life and Environmental Sciences, The University of Exeter, Stocker Road, Exeter EX4 4QD, U.K.

Abstract

The judicious choice of promoter to drive gene expression remains one of the most important considerations for synthetic biology applications. Constitutive promoter sequences isolated from nature are often used in laboratory settings or small-scale commercial production streams, but unconventional microbial chassis for new synthetic biology applications require well-characterized, robust and orthogonal promoters. This review provides an overview of the opportunities and challenges for synthetic promoter discovery and design, including molecular methodologies, such as saturation mutagenesis of flanking regions and mutagenesis by error-prone PCR, as well as the less familiar use of computational and statistical analyses for *de novo* promoter design.

Introduction

Predictable output is a defining aspiration of synthetic biology. A number of factors affect the output from synthetic gene networks to a greater or lesser extent, including transgene copy number [1], integration into the genome or expression from plasmids [2], promoter activity [3], ribosome-binding sites [4–6], codon bias of the host [7], transcription rate and tRNA abundance [8], half-life of mRNA [9], substrate and cofactor availability [10], adjustment of enzyme kinetics [11], protein scaffolding [12] and sub-cellular localization through the use of microcompartments [13,14]. The use of RNA as a control mechanism, either through the application of riboswitches [15] or toehold switches [16] has also emerged as a powerful tool for pathway control. Each of these aspects can be investigated and improved individually, and then integrated by a model, a suite of experiments or ideally, a combination of modelling and empiricism.

Several investigations, including the now archetypal 'repressilator' [17] and the genetic toggle switch [18] have modelled promoters and generated bacteria that display patterns of gene expression consistent with mathematical predictions. However, despite these successes, when individual bacteria are investigated, strong variations in transgene expression levels become apparent, even within clonal populations [19].

Controlling transcription is often the simplest way to balance expression of a transgene or synthetic pathway, and constitutive promoters with different and predictable activation characteristics are a desirable feature of any synthetic biology toolkit. However, in practise, promoter availability tends to be restricted to relatively few sequences [20], which do not always perform as required and may not necessarily be transferrable to new microbial chassis. The fact that many promoters are characterized as merely 'weak' or 'strong' [21] highlights this issue – such definitions are hardly sufficient to allow adequate promoter selection.

A number of inducible promoter systems are available for which the concentration of inducer can, in theory, be modulated in order to achieve the desired level of protein production [22]. Although the use of inducible promoter systems has been successful in some instances, in others it can prove inadequate. Promoter hypersensitivity to the inducer [23], the cost of adding large quantities of inducer to an industrial-scale fermenter [24] or heterogeneous expression levels across a population [25] all complicate the use of inducible promoters in industrial-scale cultures. Consequently, for large-scale production applications, constitutive promoters with 'hard-wired', predictable properties are often preferred and are the focus of this review.

In this article, we review the potential and methodologies for designing and characterizing new constitutive promoter sequences with predictable outputs, including conventional PCR-based techniques, hybrid promoter engineering and the expanding use of computational analysis for *de novo* promoter design.

Characteristics of promoters for synthetic biology applications

A promoter can be broadly defined as a cis-regulatory element containing a somewhat modular suite of key motifs that control the transcription of individual ORFs or operons. In prokaryotes, the structure and organization of natural promoter motifs is relatively well understood (Figure 1A). Eukaryotic promoters are somewhat more complex than their prokaryotic counterparts (Figure 1B),

Figure 1 | Schematic representations of typical promoter sequences

(**A**) Schematic representation of a typical prokaryotic promoter sequence. The transcription start site (TSS) is shown in red. Two conserved hexamers, at approximately 10 and 35 bp upstream of the TSS [68], highlighted here in blue, serve as key binding regions for RNA polymerase [69]. No such conserved motifs have been found in the region of sequence separating the two hexamers, although a consensus length of 17 bp has been observed in some species [70]. In addition to these core promoter elements, an upstream region (highlighted here in turquoise) is present in some promoters. Typically adenine/thymine rich, these UP elements boost transcription rate through interactions with the C-terminal domain on the RNA polymerase α-subunit [71]: 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. U.S.A. **95**, 9761–9766. The UP element consensus sequence is as derived by [71]. − 10 and − 35 consensus sequences are from *E. coli* and are reproduced from [3]: Blazeck, J. and Alper, H.S. (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. Biotechnol. J. **8**, 46–58 and [72]: 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. N represents any deoxyribonucleotide. W represents adenine (A) or thymine (T). G and C represent guanine and cytosine respectively. (**B**) Schematic representation of a *S. cerevisiae* promoter sequence. The TSS is highlighted in red. Eukaryotic promoters can be broadly split into two regions, a core promoter element (shown in blue) and an upstream enhancer [3] (shown in turquoise), both of which can be modified in order to modulate expression levels. The core region provides the minimal sequence necessary for initiation of basal transcription and may contain key motifs, the most widely studied of which is the TATA box, which typically occurs 40–120 bp upstream of the TSS [73]. However, such motifs are by no means requisite for transcription initiation, as TATA boxes appear in only 20 % of *S. cerevisiae* promoter elements [74]. Diagonal lines represent the region in which TATA boxes are most common. Upstream of the core promoter, the enhancer element serves to localize transcription factors, with interactions between bound transcription factors and the transcriptional machinery serving as a determinant of promoter strength and control [56]. Transcription factor binding sites do not display uniform distribution across the enhancer element, and are represented here as solid vertical lines in arbitrary positions. The highest concentration of such binding motifs has been reported between 50–150 bp prior to the TSS [75], although they may be present as much as 500 bases upstream of the TSS [76].
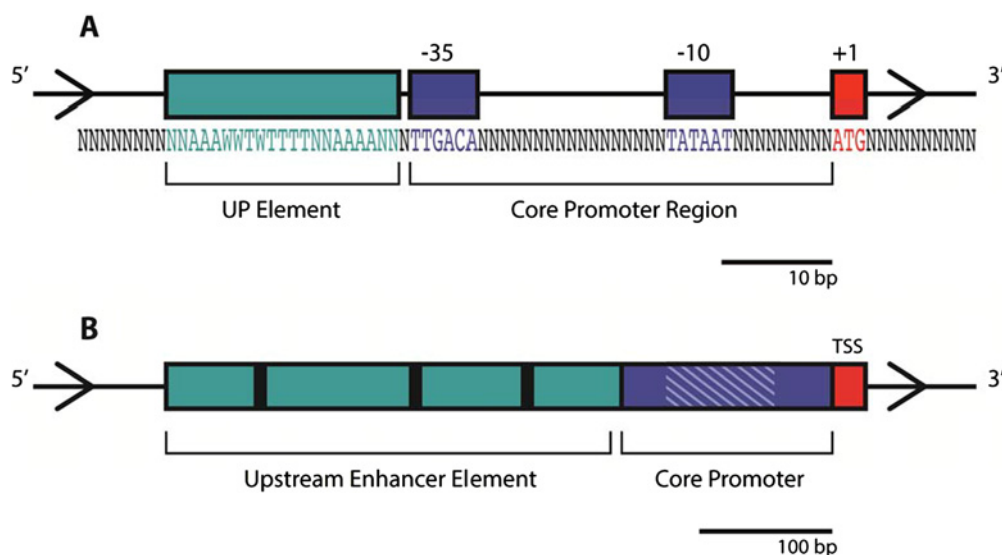
with localization of the transcriptional apparatus resulting from interactions between highly specific transcription factors, the promoter elements and co-activators [26]. The activity of promoters is typically quantified through measures of cellular mRNA or reporter proteins [27], linking the levels of promoter activity (or 'strength') to both transcription and translation. In reality, the promoter regulates only transcription but in practise, experimental constraints use protein quantification as a useful proxy for promoter activity.

From an industrial perspective, it is preferable to have a system that displays little variation, even if the overall output of that system is, on average, slightly less than that of an alternative that displays irregularities; synthetic biology aims to be boringly predictable rather than wonderfully complex. Candidate promoters for synthetic biology must therefore be well-characterized and yield consistent results, and also be insulated from the background metabolisms and molecular control systems. However, consistency is often confounded by the inherently stochastic nature of gene expression, which

subjects both promoters and any downstream proteins used in their characterization to large degrees of noise [28], as well as the all-or-nothing phenomenon [29] in inducible systems, wherein expression is typically fully induced in a subset of the population whereas the remaining cells display no expression [22,30,31].

Natural promoter sequences

The promoters available for use in synthetic systems have generally been limited to those endogenous elements isolated from model organisms, for instance, the *Escherichia coli lac* promoter and derivatives thereof [32–35] and the arabinose-inducible P_{BAD} [36–38] promoter.

Phage genomes can also be used to generate novel promoters. For example the p_L promoter, isolated from bacteriophage lambda, provides medium to high expression levels, and is tightly thermally-regulated by the cI repressor [34,39]. p_L has been successfully employed to increase yield of various proteins in *E. coli* expression systems [40–42]. Similarly, the T7 RNA polymerase-based promoter system, also initially isolated from bacteriophage, has been widely adopted [34,43].

Although natural promoters are widely used in relatively simple, laboratory applications, the relative paucity of sufficiently characterized elements makes their use in control in industrial contexts problematic. Additionally, natural promoter activity is often context-specific [3] and subject to interaction with a multitude of regulatory proteins, rendering prediction of activity levels under varying conditions non-trivial [44]. As a result of these inherent limitations, researchers have increasingly turned to libraries of synthetic promoter elements to meet their needs.

Molecular approaches for the production of synthetic promoter libraries
Saturation mutagenesis of flanking regions

A key method of forming synthetic promoter libraries (SPLs) is based on the observation that the flanking regions surrounding consensus motifs within the promoter sequence have a role in determining activity [45]. Degenerate oligonucleotides allow known consensus motifs to be maintained whereas the flanking regions are mutagenized, leading to altered promoter activity. For example saturation mutagenesis of flanking regions (SMFR) was successfully used to produce a SPL with a 400-fold activity range in *Lactococcus lactis*, with greater range being reported as a result of synthesis errors in the consensus sequences and alteration to flank length [24,45]. However, the initial approach taken to saturation mutagenesis by Jensen and Hammer [24,45] does not take into account the context-dependant nature of promoter activity. Consequently, current SPL generation uses a single PCR stage, with degenerate oligonucleotides coupled to either a full-length or truncated version of the gene that the promoter is intended to drive. This improvement allows for ectopic analysis or replacement of a wild-type promoter with a synthetic alternative, although maintaining the 5' mRNA of the target gene [23,46]. Promoter function is maintained

due to the preservation of the key consensus regions within the sequence, with altered expression levels likely being the result of minor changes in DNA confirmation within the randomized flanks [45].

SMFR has been successfully applied in a variety of prokaryotes and eukaryotes, including *Corynebacterium glutamicum* [47] and *Streptomyces coelicolor* [48], yielding robust libraries with broad expression profiles. The methodology has also shown applicability in *Saccharomyces cerevisiae*, wherein screening of an initial large library of colonies ultimately yielded 20 characterized promoters, displaying expression levels of yeast-EGFP that varied by approximately 22-fold [21]. In a separate study, a selection of constitutive promoters was initially isolated from the *S. cerevisiae* genome, and expression levels were subsequently characterized using expression profiles available from public databases. The promoter of the gene *PFΥ1* was chosen as a starting point for its robust expression profile [49]. Knowledge of *PFΥ1* structure enabled identification of a rDNA enhancer-binding protein and a poly-dT that were important for transcription initiation [50]. These regions were therefore held constant whereas a 48 bp section of the promoter core was randomized, providing a library of 36 promoter elements with a broad range of expression levels. It must be noted that none of the new sequences provided higher expression levels than the original *PFΥ1* promoter [49]. This inability to produce a synthetic promoter with higher expression levels than a natural alternative was also reported by McWhinnie and Nano [51].

Although SMFR has successfully provided many new promoters, the technique requires labour intensive cloning and an *a priori* knowledge of promoter structure in the organism of interest, something that may not be immediately available in industrially relevant microbes. Furthermore, as many libraries use composite promoter scaffolds as a starting point, establishing a definitive wild-type reference expression baseline is impossible. Definitively stating whether SMFR will improve wild-type expression capability *pre hoc*, is therefore problematic [3]. Additionally, by restricting mutagenesis to only the flanking regions, SMFR fails to take into account alterations to consensus sequences, which are known to play a significant role in modulating expression strength.

Error-prone PCR

Generating a SPL by applying error-prone PCR (epPCR) to an entire promoter sequence obviates any *a priori* knowledge of functional motif location and can potentially result in promoters with entirely new characteristics [3]. This methodology was successfully used to mutagenize a bacteriophage P_{L}-λ promoter that was subsequently placed upstream of a green fluorescent protein (GFP) coding sequence and transformed into *E. coli*, resulting in a library containing approximately 9000–12000 functional clones [52,53]. Visual screening of the colonies resulted in a subset of 200 promoters, of which 27, representing 22 discrete promoter sequences, were found to give homogeneous expression levels. Subsequently,

thorough characterization of this promoter subset resulted in a promoter library which was successfully employed to modulate levels of phosphoenolpyruvate carboxylase and lycopene production in *E. coli* [53]. epPCR for promoter production has also been employed in *C. glutamicum*, where iterative rounds of high-throughput sorting and analysis at the single-cell level ultimately yielded a library of 20 well-characterized sequences from an initial library of 10^5 mutagenized cells [54]. The technique has also been successfully applied in yeast [55].

Despite these successes, the epPCR approach to SPL production has certain limitations: a reliance on a selection of a small subset of colonies for further analysis [53,54] renders discovery of a true optimum problematic. Moreover, the extensive screening required to isolate said subset should not be underestimated; it is typical for initial libraries of hundreds or thousands of bacterial colonies to ultimately yield relatively few fully characterized promoters. Both these problems become less of an issue if visual selection of colonies is replaced by high-throughput analytical techniques such as fluorescence-activated cell sorting and/or imaging cytometry.

Hybrid promoter engineering

In addition to the two mutagenic techniques discussed above, the generation of synthetic promoters through hybridization of existing promoter elements provides an alternative strategy for promoter genesis. By combining minimal core promoter elements with various combinations of modular upstream activation sequences (UAS), Blazeck et al. [56] demonstrated that expression levels could be increased compared with a wild-type baseline in *S. cerevisiae*. A roughly linear relationship was observed between the number of UAS modules added and promoter strength, with the addition of four such elements boosting expression of a weak constitutive promoter to levels comparable with the strongest endogenous promoter [56]. Transcriptional increase was shown to depend both on the core element and UAS, but all core promoters were amenable to improvement [56].

Computational methods for synthetic promoter discovery

Although the above molecular methodologies have certainly provided new promoters of varying activities, these approaches do not represent a systematic, theoretical examination of the promoter design space. If, for arguments sake, a promoter sequence is 100 bp in length, there are 4^{100} potential promoter sequences. Therefore, although the best sequence discovered by molecular-based SPL may be sufficient for some experimental purposes, it is possible that other optima are present. *In silico* methods that are capable of deciphering the effect of individual DNA bases and motifs, or predicting promoter activity level in advance of *in vivo* characterization have, in this context, considerable potential [57]. Conventionally, the use of computational techniques in pathway design and optimization has been limited to *post hoc* data analytics [21]. However, computational modelling in biological systems design and optimization is becoming more

widespread, and a number of computational methodologies are available to facilitate the *de novo* design of synthetic promoter sequences.

Position weight matrix models

Position weight matrix (PWM) models have been widely applied for the detection of transcription factor binding sites [58,59], and have also shown some promise in predicting promoter strength. By breaking promoter sequences into constitutive motifs, PWM models were able to predict the strength of *E. coli* core promoter sequences recognized by sigma factor σ^E to a relatively high degree of accuracy [60]. The core promoter PWM was subsequently combined with a score describing the activity of upstream elements to provide a model capable of predicting the strength of entire promoter sequences [61]. In addition to this predictive power, PWM models provide increased understanding of promoter structure, something that is often limited in novel microbial chassis.

Although PWM models certainly have the potential to be applied to *de novo* sequence design, they are not without limitations. PWMs may prove inadequate for modelling in promoter families with a less conserved nature than those which interact with σ^E, as poorly conserved sequences required greater complexity within the model [60]. Application of PWMs in novel microbial chassis, where understanding of interactions between proteins and promoter sequences can be limited, may therefore be challenging.

Additionally, by assuming that the contribution of individual nucleotides to DNA-protein binding is independent and additive [61], PWMs fail to account for the effect of interactions between positions. Despite these limitations, the application of PWMs for the *pre hoc* determination of strength in certain promoter families carries great potential.

Partial least squares regression

The use of statistical modelling to quantitatively link DNA sequence to function is not a new concept [62], although as a method for the generation of synthetic promoters it remains underutilized. In a pioneering study, 25 *E. coli* promoters were analysed using a partial least squares (PLS) methodology, resulting in a statistical model that analysed the contribution of each individual nucleotide at any given position in the DNA sequence. In order to validate the model, two synthetic sequences with predicted high activity levels were synthesized. The -35, -10 and $+1$ sites were determined using the consensus sequence of the training set of 25 promoters, whereas the remainder of the synthetic sequences were determined using regression coefficients provided by the modelling process [62]. *In vivo* characterization of the synthetic promoters revealed activity levels within approximately 8% of the strength predicted by the model. Furthermore, the synthetic sequences were shown to provide higher expression levels than any of those sequences found within the training set [62].

Similar statistical methods were later applied to quantitatively link promoter structure with function for a library

of synthetic *E. coli* promoters that were generated through the randomization of flanking regions [29]. The generated model was able to predict, with reasonable accuracy, the strength of promoter sequences that had not been used in the construction of the model [29]. In further validation of this computational technique, the promoter strength predictive model was subsequently utilized to predict the strength of an endogenous *E. coli* promoter, that of the *ppc* gene [63]. Based on this information, stronger promoters were selected from the previously characterized promoter library [29] in order to fine-tune *ppc* expression levels. This knock-in approach resulted in an increase in expression levels roughly in line with the model's predictions, with a 3–4-fold increase in mRNA levels seen at flask scale [63]. Although the PLS regression doubtlessly aided in the optimization process, it was not applied, in this instance to the *de novo* design of synthetic promoter sequences.

Artificial neural networks

The linear nature of PLS modelling is a drawback when applied to the analysis of promoter sequences, confounding the effects of any interactions between bases with the main effects for each individual nucleotide position [62]. PLS models therefore may not accurately account for the complexity inherent in promoter structure, thereby increasing the probability of prediction errors and inadequate generality [64]. Indeed, many such models lack robust prediction accuracy [65], rendering their use in *de novo* sequence design challenging.

Artificial neural networks (ANNs) may provide a solution to these issues. Based upon a network of interconnected nodes designed to act as a rudimentary mimic of the brain, ANNs permit machine learning, as the order and force of connections may be altered [66]. By systematically altering node structure during the analysis of a training data set, ANN models can potentially better represent the complex, non-linear interactions occurring within a promoter sequence [64]. ANN modelling has proven successful for *de novo* promoter design [64]; using a set of synthetic promoters derived from the random mutagenesis of a wild-type *E. coli* promoter as a training set for an ANN model, strength predictions of sequences generated by *in silico* mutagenesis were used to select 16 synthetic sequences for *in vivo* verification [64]. The predicted expression levels displayed good correlation with empirical testing, suggesting that such models are indeed applicable to synthetic promoter design. Indeed, the fact that approximately 30% of *de novo* designed sequences displayed greater expression levels that the wild-type control [64] compares extremely favourably to the more traditional mutagenesis-based techniques discussed above, where much lower success rates are not uncommon.

The importance of insulation

Whichever method is applied to the generation of SPLs, promoter elements must be sufficiently insulated if they are to be efficiently used in synthetic regulatory systems. Empirical or predictive data regarding promoter strength from characterization using a reporter protein must be comparable to promoter performance when coupled to a protein of interest within a synthetic pathway; context-dependent effects should be minimal. However, achieving context dependency is non-trivial, as fluctuation in promoter activity levels may be the result of a wide array of experimental and/or genetic factors [27,53].

A possible solution to this problem is to separate core elements from their genetic context through the use of insulator sequences, such as a defined 5' mRNA sequence [67]. By using such insulators, promoter elements from a SPL can produce constant relative levels of various reporter proteins when used for both plasmid and chromosomal expression [67].

Conclusion

The ability to select a reliable promoter of known activity is of paramount importance for synthetic biology. Indeed, promoters with different and, most importantly, predictable effects on transcription may be used to regulate complex gene circuits, balance engineered metabolic pathways and exploit new chassis for industrial-scale applications. As reviewed here, a number of molecular and computational methodologies are available for the discovery and design of new constitutive promoters. Each technique has advantages and weaknesses, and a selection of one over the other will depend on the aims of specific projects. However, to date, computational approaches to promoter design remain underutilized aside from proof of principle studies in model organisms. As the applications of synthetic biology become more entrenched in the future bio-economy, which may require the development of different chassis, the application of computational modelling to promoter design can enhance and accelerate the design process and ultimately enhance our fundamental knowledge of genetic regulation in complex systems.

References

1 Ajikumar, P.K., Xiao, W.H., Tyo, K.E.J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Heng Phon, T., Pfeifer, B. and Stephanopoulos, G. (2010) Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. Science **330**, 70–74 CrossRef PubMed

2 Tyo, K.E.J, Ajikumar, P.K. and Stephanopoulos, G. (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat. Biotechnol. **27**, 760–765 CrossRef PubMed

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

4 Lin, Z., Xu, Z., Li, Y., Wang, Z., Chen, T. and Zhao, X. (2014) Metabolic engineering of *Escherichia coli* for the production of riboflavin. Microb. Cell Fact. **13**, 1–12 CrossRef PubMed

5 Ravasi, P., Peiru, S., Gramajo, H. and Menzella, H.G. (2012) Design and testing of a synthetic biology framework for genetic engineering of *Corynebacterium glutamicum*. Microb. Cell Fact. **11**, 147–158 CrossRef PubMed

6 Markley, A.L., Begemann, M.B., Clarke, R.E., Gordon, G.C. and Pfleger, B.F. (2015) Synthetic Biology toolbox for controlling gene expression in the Cyanobacterium *Synechococcus* sp. strain PCC 7002. ACS Synth. Biol. **4**, 595–603 CrossRef PubMed

7 Quax, T.E.F., Claassens, N.J., Söll, D. and van der Oost, J. (2015) Codon bias as a means to fine-tune gene expression. Mol. Cell **59**, 149–161 CrossRef PubMed

8 Angov, E. (2011) Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol. J. **6**, 650–659 CrossRef PubMed

9 Curran, K.A., Karim, A.S., Gupta, A. and Alper, H.S. (2013) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab. Eng. **19**, 88–97 CrossRef PubMed

10 Jones, J.A., Toparlak, Ö.D. and Koffas, M.A. (2015) Metabolic pathway balancing and its role in the production of biofuels and chemicals. Curr. Opin. Biotech. **33**, 52–59 CrossRef

11 Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D. and Arnold, F.H. (2005) Evolving strategies for enzyme engineering. Curr. Opin. Struct. Biol. **15**, 447–452 CrossRef PubMed

12 Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L.J. and Keasling, J.D. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotechnol. **27**, 753–759 CrossRef PubMed

13 Boyle, P.M. and Silver, P.A. (2012) Parts plus pipes: synthetic biology approaches to metabolic engineering. Metab. Eng. **14**, 223–232 CrossRef PubMed

14 Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P. and Warren, M.J. (2010) Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. Mol. Cell **38**, 305–315 CrossRef PubMed

15 Mellin, J.R. and Cossart, P. (2015) Unexpected versatility in bacterial riboswitches. Trends Genet. **31**, 150–156 CrossRef PubMed

16 Green, A.A., Silver, P.A., Collins, J.J. and Yin, P. (2014) Toehold switches: de-novo-designed regulators of gene expression. Cell **159**, 925–939 CrossRef PubMed

17 Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. Nature **403**, 335–338 CrossRef PubMed

18 Gardner, T.S., Cantor, C.R. and Collins, J.J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature **403**, 339–342 CrossRef PubMed

19 Guido, N.J., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C.R., Elston, T.C. and Collins, J.J. (2006) A bottom-up approach to gene regulation. Nature **439**, 856–860 CrossRef PubMed

20 Lu, T.K., Khalil, A.S. and Collins, J.J. (2009) Next-generation synthetic gene networks. Nat. Biotechnol. **27**, 1139–1150 CrossRef PubMed

21 Ellis, T., Wang, X. and Collins, J.J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat. Biotechnol. **27**, 465–471 CrossRef PubMed

22 Siegele, D.A. and Hu, J.C. (1997) Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. Proc. Natl. Acad. Sci. U.S.A. **94**, 8168–8172 CrossRef PubMed

23 Hammer, K., Mijakovic, I. and Jensen, P.R. (2006) Synthetic promoter libraries – tuning of gene expression. Trends Biotechnol. **24**, 53–55 CrossRef PubMed

24 Jensen, P.R. and Hammer, K. (1998) Artificial promoters for metabolic optimization. Biotechnol. Bioeng. **58**, 191–195 CrossRef PubMed

25 Khlebnikov, A., Datsenko, K.A., Skaug, T., Wanner, B.L. and Keasling, J.D. (2001) Homogeneous expression of the PBAD promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiol. **147**, 3241–3247 CrossRef

26 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 CrossRef PubMed

27 Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., de Mora, K., Glieberman, A.L., Monie, D.D. and Endy, D. (2009) Measuring the activity of BioBrick promoters using an *in vivo* reference standard. J. Biol. Eng. **3**, 4 CrossRef PubMed

28 Rudge, T.J., Brown, J.R., Federici, F., Dalchau, N., Phillips, A., Ajioka, J.W. and Haseloff, J. (2016) Characterization of intrinsic properties of promoters. ACS Synth. Biol. **5**, 89–98 CrossRef PubMed

29 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 CrossRef PubMed

30 Morgan-Kiss, R.M., Wadler, C. and Cronan, Jr, J.E. (2002) Long-term and homogeneous regulation of the *Escherichia coli* araBAD promoter by use of a lactose transporter of relaxed specificity. Proc. Natl. Acad. Sci. U.S.A. **99**, 7373–7377 CrossRef PubMed

31 Keasling, J.D. (1999) Gene-expression tools for the metabolic engineering of bacteria. Trends Biotechnol. **17**, 452–460 CrossRef PubMed

32 Makoff, A.J. and Oxer, M.D. (1991) High level heterologous expression in *E. coli* using mutant forms of the *lac* promoter. Nucleic Acids Res. **19**, 2417–2421 CrossRef PubMed

33 de Boer, H., Comstock, L.J. and Vasser, M. (1983) The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. Proc. Natl. Acad. Sci. U.S.A. **80**, 21–25 CrossRef PubMed

34 Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. **72**, 211–222 CrossRef PubMed

35 Jensen, P.R., Westerhoff, H.V. and Michelsen, O. (1993) The use of *lac*-type promoters in control analysis. Eur. J. Biochem. **211**, 181–191 CrossRef PubMed

36 Wycuff, D.R. and Matthews, K.S. (2000) Generation of an AraC-*ara*BAD promoter-regulated T7 expression system. Anal. Biochem. **277**, 67–73 CrossRef PubMed

37 Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation and high-level expression by vectors containing the Arabinose PBAD promoter. J. Bacteriol. **177**, 4121–4130 PubMed

38 Cagnon, C., Valverde, V. and Masson, J.M. (1991) A new family of sugar-inducible expression vectors for *Escherichia coli*. Protein Eng. **4**, 843–847 CrossRef PubMed

39 Valdez-Cruz, N.A., Caspeta, L., Pérez, N.O., Ramírez, O.T. and Trujillo-Roldán, M.A. (2010) Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. Microb. Cell Fact. **9**, 18 CrossRef PubMed

40 Elvin, C.M., Thompson, P.R., Argall, M.E., Hendry, P., Stamford, N.P.J., Lilley, P.E. and Dixon, N.E. (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. Gene **87**, 123–126 CrossRef PubMed

41 Mellado, R.P. and Salas, M. (1982) High level synthesis in *Escherichia coli* of the *Bacillus subtillis* phage phi 29 proteins p3 and p4 under the control of phage lambda PL promoter. Nucleic Acids Res. **10**, 5773–5784 CrossRef PubMed

42 Simons, G., Remaut, E., Allet, B., Devos, R. and Fiers, W. (1984) High-level expression of human interferon gamma in *Escherichia coli* under control of the *pL* promoter of bacteriophage lambda. Gene **28**, 55–64 CrossRef PubMed

43 Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**, 113–130 CrossRef PubMed

44 Collado-Vides, J., Magasanik, B. and Gralla, J.D. (1991) Control site location and transcriptional regulation in *Escherichia coli*. Microbiol. Rev. **55**, 371–394 PubMed

45 Jensen, P.R. and Hammer, K. (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl. Environ. Microb. **64**, 82–87

46 Solem, C. and Jensen, P.R. (2002) Modulation of gene expression made easy. Appl. Environ. Microb. **68**, 2397–2403 CrossRef

47 Rytter, J.V., Helmark, S., Chen, J., Lezyk, M.J., Solem, C. and Jensen, P.R. (2014) Synthetic promoter libraries for *Corynebacterium glutamicum*. Appl. Microbiol. Biotechnol. **98**, 2617–2623 CrossRef PubMed

48 Sohoni, S.V., Fazio, A., Workman, C.T., Mijakovic, I. and Lantz, A.E. (2014) Synthetic promoter library for modulation of Actinorhodin production in *Streptomyces coelicolor* A3(2). PLoS One **9**, e99701 CrossRef PubMed

49 Blount, B.A., Weenink, T., Vasylechko, S. and Ellis, T. (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. PLoS One **7**, e33279 CrossRef PubMed

50 Angermayr, M., Oechsner, U. and Bandlow, W. (2003) Reb1p-dependent DNA bending effects nucleosome positioning and constitutive transcription at the yeast profilin promoter. J. Biol. Chem. **278**, 17918–17926 CrossRef PubMed

51 McWhinnie, R.L. and Nano, F.E. (2013) Synthetic promoters functional in *Francisella novicida* and *Escherichia coli*. Appl. Environ. Microb. **80**, 226–234 CrossRef

52 Fischer, C.R., Alper, H., Nevoigt, E., Jensen, K.L. and Stephanopoulos, G. (2006) Response to Hammer et al.: tuning genetic control – importance of thorough promoter characterization versus generating promoter diversity. Trends Biotechnol. **24**, 55–56 CrossRef PubMed

53 Alper, H., Fischer, C., Nevoigt, E. and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. Proc. Natl. Acad. Sci. U.S.A. **102**, 12678–12683 CrossRef PubMed

54 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 CrossRef PubMed

55 Nevoigt, E., Kohnke, J., Fischer, C.R., Alper, H., Stahl, U. and Stephanopoulos, G. (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. **72**, 5266–5273 CrossRef PubMed

56 Blazeck, J., Garg, R., Reed, B. and Alper, H.S. (2012) Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. Biotechnol. Bioeng. **109**, 2884–2895 CrossRef PubMed

57 Jensen, K., Alper, H., Fischer, C. and Stephanopoulos, G. (2006) Identifying functionally important mutations from phenotypically diverse sequence data. Appl. Environ. Microb. **72**, 3696–3701 CrossRef

58 Sinha, S. (2006) On counting position weight matrix matches in a sequence, with application to discriminative motif finding. Bioinformatics **22**, e454–e463 CrossRef PubMed

59 Stromo, G.D. (2000) DNA binding sites: representation and discovery. Bioinformatics **16**, 16–23 CrossRef PubMed

60 Rhodius, V.A. and Mutalik, V.K. (2010) Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, σE. Proc. Natl. Acad. Sci. U.S.A. **107**, 2854–2859 CrossRef PubMed

61 Rhodius, V.A., Mutalik, V.K. and Gross, C.A. (2012) Predicting the strength of UP-elements and full-length *E. coli* σE promoters. Nucleic Acids Res. **40**, 2907–2924 CrossRef PubMed

62 Jonsson, J., Norberg, T., Carlsson, L., Gustafsson, C. and Wold, S. (1993) Quantitative sequence-activity models (QSAM)-tools for sequence design. Nucleic Acids Res. **12**, 733–739 CrossRef

63 De Mey, M., Maertens, J., Boogmans, S., Soetaert, W.K., Vandamme, E.J., Cunin, R. and Foulquié-Moreno, M.R. (2010) Promoter knock-in: a novel rational method for the fine tuning of genes. BMC Biotechnol. **10**, 26 CrossRef PubMed

64 Meng, H., Wang, J., Xiong, Z., Xu, F., Zhao, G. and Wang, Y. (2013) Quantitative design of regulatory elements based on high-precision strength prediction using artificial neural network. PLoS One **8**, e60288 CrossRef PubMed

65 Meng, H. and Wang, Y. (2015) *Cis*-acting regulatory elements: from random screening to quantitative design. Quant. Biol. **3**, 107–114 CrossRef

66 Buscema, P.M., Massini, G. and Maurelli, G. (2014) Artificial neural networks: an overview and their use in the analysis of the AMPHORA-3 dataset. Subst. Use Misuse **49**, 1555–1568 CrossRef PubMed

67 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 CrossRef PubMed

68 Harley, C.B. and Reynolds, R.P. (1987) Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15**, 2343–2361 CrossRef PubMed

69 Kanhere, A. and Bansal, M. (2005) Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. Nucleic Acids Res. **33**, 3165–3175 CrossRef PubMed

70 Nair, T.M. and Kulkarni, B.D. (1994) On the consensus structure within the *E. coli* promoters. Biophys. Chem. **48**, 383–393 CrossRef PubMed

71 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. U.S.A. **95**, 9761–9766 CrossRef PubMed

72 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 PubMed

73 Hampsey, M. (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol. Mol. Biol. Rev. **62**, 465–503 PubMed

74 Basehoar, A.D., Zanton, S.J. and Pugh, B.F. (2004) Identification and distinct regulation of yeast TATA box-containing genes. Cell **116**, 699–709 CrossRef PubMed

75 Hughes, J.D., Estep, P.W., Tavazoie, S. and Church, G.M. (2000) Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. J. Mol. Biol. **296**, 1205–1214 CrossRef PubMed

76 Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J. et al. (2004) Transcriptional regulatory code of a eukaryotic genome. Nature **431**, 99–104 CrossRef PubMed