### REVIEW

# Cis-acting regulatory elements: from random screening to quantitative design

#### Hailin Meng<sup>1</sup> and Yong Wang<sup>2,\*</sup>

- <sup>1</sup> Bioengineering Research Center, Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou 511458, China
- <sup>2</sup> CAS Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
- \* Correspondence: yongwang@sibs.ac.cn

Received May 31, 2015; Revised September 17, 2015; Accepted September 18, 2015

The cis-acting regulatory elements, e.g., promoters and ribosome binding sites (RBSs) with various desired properties, are building blocks widely used in synthetic biology for fine tuning gene expression. In the last decade, acquisition of a controllable regulatory element from a random library has been established and applied to control the protein expression and metabolic flux in different chassis cells. However, more rational strategies are still urgently needed to improve the efficiency and reduce the laborious screening and multifaceted characterizations. Building precise computational models that can predict the activity of regulatory elements and quantitatively design elements with desired strength have been demonstrated tremendous potentiality. Here, recent progress on construction of cisacting regulatory element library and the quantitative predicting models for design of such elements are reviewed and discussed in detail.

**Keywords:** cis-acting regulatory element; quantitative design; synthetic biology; random mutation; modeling

#### INTRODUCTION

Cis-acting regulatory elements are the DNA sequences playing regulatory role in the genome such as promoter, ribosome binding site (RBS), and terminator (Figure 1A), which can be bonded with transcription factors, RNA polymerase or ribosomes to regulate the rate of RNA transcription or protein translation. Cis-acting regulatory elements, together with structure elements like catalytic enzymes responsible for metabolic pathways [6], as well as trans-acting elements like microRNA, protein or synergetic effector (e.g., CRISPRi) [7,8], make up the basic biological parts or building blocks, which can be used for designing and constructing artificial biological systems in synthetic biology research.

Cis-acting regulatory elements can generate spatial and temporal patterns of gene expression in an organism. In synthetic biology research, it usually needs to predefine and construct a number of fine-tuning regulatory elements with various parameters or characteristics, e.g., promoters and RBSs with a wide range of strengths or activity, to

meet different requirements. For instance, when designing and assembling a new pathway for valuable compound biosynthesis, it is necessary to optimize the pathway to obtain a higher product yield. By precisely and quantitatively balancing the expression of enzymes in the pathway, it may not only eliminate the rate limiting steps, but also can effectively avoid the waste of resource by excessive expression, thus it is an effective optimization method to achieve this goal. Several strategies are available to achieve the moderate expression level of enzymes in the pathway [5,9–11]: (i) using promoters with different strength to control gene transcription rate; (ii) using RBSs with different strength to alter the rate of protein translation; (iii) using protein degradation tags to balance the protein synthesis rate and degradation rate; (iv) using terminators to influence net protein output by controlling mRNA half-life. These strategies perform precise regulation on the synthesis of pathway enzymes at various biological processes, including gene transcriptional level, protein translational level, and post-translational level. In fact, in natural biological systems, the

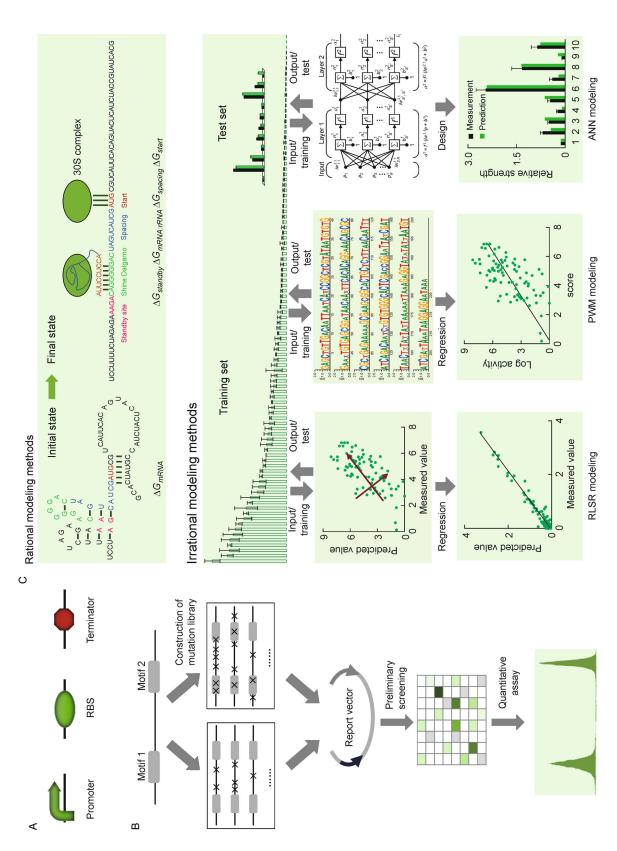


Figure 1. Methods for cis-acting regulatory element screening and design. (A) Three kinds of typical cis-acting regulatory elements: promoter, RBS, and terminator. (B) The workflow of construction of cis-acting regulatory element library based on random mutation and screening. (C) Rational and irrational modeling approaches for quantitative prediction and design the strength of cis-acting regulatory elements, including Partial Least Squares Regression (PLSR) modeling [1], Artificial Neural Network (ANN) modeling [2,3], Position weight matrix (PWM) modeling [4], and Biophysical modeling (e.g., thermodynamic modeling [5]).

quantity of different kinds of proteins is tightly regulated. This behavior acts in accordance with the law of cell economics, which helps the cell to adapt itself for environment change and survive, but usually not to meet with our requirement. Therefore, in order to achieve our specific goal (e.g., biosynthesis of valuable products), it is necessary to destroy the original regulation systems and introduce new regulation laws into the cell when designing the biological systems. This requires the quantitative and fine-tuning expression of each enzyme of the pathway or network. Hence, development of

regulatory element library with fine-tuning properties as well as the tools and databases have important significance for synthetic biology application (Tables 1 and 2).

Biological parts, including various regulatory elements, were naturally isolated, characterized and standardized to build up open access libraries for academy application, such as the Registry of Standard Biological Parts established by Massachusetts Institute of Technology (http://partsregistry.org), and The Joint BioEnergy Institute Inventory of Composable Elements (JBEI-ICEs) set up by Jay D. Keasling's group [22] (Table 2). However,

Table 1. Methods for screening and quantitative design of cis-acting regulatory elements.

	Method		Example	
Screening	From nature		Yamanishi et al. [11] evaluated the activity of 5,302 terminator regions in <i>Saccharomyces cerevisiae</i> and constructed a library of terminator.	
	By random mutation		De Mey et al. [1] used this method to design and synthesize a set of promoters with different strengths.  Siegl et al. [12] constructed and characterized a synthetic promoter library for fine-tuning the get expression in <i>Actinomycetes</i> . The relative strengths ranged from 0.02 to 3.19.  Qin et al. [13] applied this method to <i>Pichia pastoris</i> to construct a mutation library of the constitutive GAP promoter. The relative strengths ranged from 0.006 to 19.6.  Mutalik et al. [14] randomized the -10 and -35 motifs of the Ptrc promoter for design and construction of the synthetic constitutive promoter library.	
Design	Rational method	Biophysical modeling	Salis et al. [5] developed a thermodynamic model to predict the binding affinity between ribosome and RBS region, which can be used for RBS design.  Brewster et al. [15] constructed a thermodynamic model of transcription and combined with the protein-DNA binding energy function to control a targeted gene expression level in <i>E. coli</i> . Na et al. [16] proposed a mathematical model that uses mRNA sequence information to estimate translational efficiency.  Juven-Gershon et al. [17] described the design and analysis of a super core promoter (SCP1). Vilar et al. [18] provided a quantitative framework that accurately integrates sequence statistics with a biophysical model.	
	Irrational method	Partial Least Squares Regression (PLSR) modeling Position Weight	De May et al. [1] applied the PLSR method to characterize the relationship between sequences and strengths of <i>E. coli</i> promoters and established the prediction model.  Rhodius et al. [4] proposed a new method based on PWM modeling to predict the strength of	
		Matrix (PWM) modeling	$\sigma^{E}$ recognized promoters in <i>E. coli</i> .	
		Artificial Neural Network (ANN) modeling	Meng et al. [2] put forward a modeling approach based on ANN to characterize the nonlinear relationship between sequence and strength. The best training ANN model gets a high regression correlation coefficient of 0.98.	

Table 2. Tools and databases for design of cis-acting regulatory elements.

Tools/databases	Reference	Website	Description
RBSDesigner	Na et al. [19]	http://ssbio.cau.ac.kr/	It can predict the translation efficiency of existing mRNA sequences and design synthetic RBS for a given coding sequence to yield a desired level of protein expression.
PromoterCAD	Nishikata et al. [20,21]	http://promotercad.org/	Design of synthetic promoters with novel regulation functions to control gene expression in plants or mammals.
Registry of Standard Biological Parts	Massachusetts Institute of Technology	http://partsregistry.org/	A growing collection of genetic parts that can be mixed and matched to build synthetic biology devices and systems.
JBEI-ICEs	Jay D. Keasling's group [22]	http://www.jbei.org/	An open source registry platform for managing information about biological parts.

compared to the vast natural resources, only very limited biological elements have been explored. Take microorganism for example, only ~1% of the total resources can be cultured by current technology. This results in an extremely low efficiency in exploration of the natural biological elements, thus a large number of novel elements is quite difficult to obtain due to the limitation of current biotechnology. Besides, more elaborate biological elements are required for the design of complex systems, but these elements are often difficult to directly screen from nature. Therefore, it is quite necessary to perform manual construction based on the existing knowledge and make full use of the library resources of natural elements.

Two main approaches are available for manual construction of library according to the design principles: the one is based on random mutation and library screening, and the other is model driven design based on quantitative prediction models. The latter fits the goal and the concept of rational design in synthetic biology.

### CONSTRUCTION OF CIS-ACTING REGULATORY ELEMENT LIBRARY

Regulatory element library consists of element sequences with various strengths of transcriptional or translational activity. Two main approaches can be applied to construct the library: (i) random mutation based construction and screening, and (ii) isolation of elements from native organisms (Table 1).

### Random mutation based construction and screening

For certain prokaryote promoters, -35 box and -10 box are the relatively conservative regions, and the variation of the spacer region between these two boxes will alter the strength of the promoter [23,24]. Therefore, a series of promoter sequences with various strengths can be generated by changing the space region sequence according to this principle. De Mey et al. [1] degenerated oligonucleotide sequence that encodes consensus sequences for Escherichia coli promoters separated by spacers of random sequences, and designed and synthesized a set of promoters with different strength. Siegl et al. [12] constructed and characterized a synthetic promoter library based on the -10 and -35 consensus sequences of the promoter for fine-tuning the gene expression in Actinomycetes, and obtained the relative strengths ranged from 0.02 to 3.19 compared to the wild-type sequence.

Random mutations can also be introduced into element sequences via using error-prone PCR technology. Mutations may randomly occur at a certain frequency by adjusting the reaction conditions during the process of PCR amplification for target element (e.g., increasing magnesium ion concentration, adding manganese ions, changing the concentration of dNTPs, or using low fidelity of DNA polymerase). A library composed of mutated elements with various strengths can be constructed after primary screening and strength assay. Compared with the aforementioned method based on randomized primers and spacer, a wider range of mutation region beside spacer was also included in error-prone PCR, such as -10 box region and -35 box region, which provides more choices for screening. Figure 1B depicts the general workflow of element library construction based on random mutation technology: (i) selection of the regions for mutation; (ii) design of PCR primers according to the selected sequence features; (iii) performing PCR amplification with wild type promoter sequence as a template; and (iv) isolation and screening of the mutation library. For the convenience of screening, a reporter gene such as green fluorescent protein gene (gfp) can be inserted into the plasmid vector following the mutation region, by quantitatively detecting the strength of mutated sequences via measuring the fluorescence strength of downstream GFP. Afterwards, mutants with various strengths will be picked out to construct a strength-gradient element library.

Random mutation based technology for library construction has been widely used in different bacterial systems, as well as some eukaryotic systems like yeast [1,13,23–28]. For instance, Qin et al. [13] applied this method to Pichia pastoris to construct a mutation library of the constitutive GAP promoter. They selected 33 mutated sequences with a wide spread of relative strength ranged from 0.006 to 19.6 to construct the library, which provides a useful toolbox for research and engineering application of the yeast system. Besides, Mutalik et al. [14] randomized the -10 and -35 motifs of the Ptrc promoter for design and construction of the synthetic constitutive promoter library, and further developed an expression cassette architecture for genetic elements controlling transcription and translation initiation in E. coli.

#### Isolation of elements from native organisms

Cis-acting regulatory elements with various activities can also be obtained directly by isolation from existing organisms. In a typical case of terminator engineering, Yamanishi et al. [11] evaluated the activity of 5,302 terminator regions in *Saccharomyces cerevisiae*. As a result, terminator activities relative to that of the PGK1 standard terminator ranged from 0.036 to 2.52, thus the terminators with various strengths could be selected to construct a library called "Terminatome" Toolbox, besides isolation from yeast, Curran et al. [29] also

developed *de novo* design approach to obtain synthetic terminators that can be used for modulating gene expression in yeast. The best synthetic terminator resulted in 3.7-fold more fluorescent protein output and 4.4-fold increase in transcript level compared to the commonly used CYC1 terminator. These isolated or engineered terminators are useful for the development of metabolically and genetically engineered yeast system.

## QUANTITATIVE DESIGN OF CIS-ACTING REGULATORY ELEMENTS WITH DESIRED STRENGTH

Although random mutation based method is effective to construct the element library, it is still a laborious work with low efficiency, especially for the construction of different kinds of libraries for complex biological systems design. In addition, most protein elements such as enzymes and regulatory proteins lack high throughput method for library screening, and cannot quickly get the expected function of the mutant. To this end, researchers invested great efforts on discovering the complex relationship between element sequences and their corresponding strengths, and a series of quantitative prediction models have been developed based on the knowledge of the relationship. Quantitative prediction based sequence design thus becomes a trend of de novo element design for synthetic biology application. Figure 1C and Table 1 summarize the rational and irrational methods for quantitative design of cis-acting regulatory elements.

Rational design methods use the intrinsic law of interactions between protein and DNA to build biophysical models to design the sequence of DNA elements with expected parameters, such as specific strength of the promoters. However, design of such elements purely by rational method is still difficult at current stage, since it lacks sufficiently detailed information of macromolecular structures and interactions. Hence, more quantitative prediction models are built by irrational design methods, such as Partial Least Squares Regression (PLSR) modeling, Position weight matrix (PWM) modeling, and Artificial Neural Network (ANN) modeling. These modeling methods are set up based on knowledge learned from existing element sequence feature and its corresponding strength, and do not need to know much about the exact mechanism of macromolecular interactions. More often, the relationship between sequence and strength is described by empirical or semi-empirical formula to construct quantitative prediction models that can be used under certain conditions.

#### **Biophysical modeling**

A biophysical model uses mathematical formalizations of

the physical properties to simulate the behavior of biological systems. Such models can be used to predict the influence of biological and physical factors on complex systems. Take RBS design for example, a thermodynamic model was developed by Salis et al. [5] to predict the binding affinity between ribosome and RBS region based on biophysical modeling method. The designed RBSs could be used to fine-tune the translational initiation rate and expression level of downstream proteins. The correlation coefficient of model fitting is up to 0.9. This modeling approach is one of the most typical representative methods for rational design at present.

Recently, Brewster et al. [15] constructed a thermodynamic model of transcription and combined with the protein-DNA binding energy function to control a targeted gene expression level over three orders of magnitude in *E. coli*, and showed a good performance on prediction of expression level, which may provide an engineering tool for use in synthetic biology.

Na et al. [16] proposed a mathematical model that uses mRNA sequence information to estimate translational efficiency. This model effectively estimates translational efficiencies based on mRNA-folding dynamics and ribosome-binding dynamics information. It facilitates over-production or optimization of protein expression level for the construction of robust networks in synthetic biology.

Juven-Gershon et al. [17] described the design and analysis of a super core promoter (SCP1), which contains the TATA box, initiator, motif ten element and downstream promoter element, and each motif is needed for full SCP1 activity. The super core promoter is useful for the enhancement of gene expression in cells.

Vilar et al. [18] provided a quantitative framework that accurately integrates sequence statistics with a biophysical model. The model considers a decomposition of the free energy of the protein-DNA complex into different modular contributions. It can accurately predict gene expression from statistical sequence information in combination with detailed biophysical modeling of transcription regulation.

#### Partial Least Squares Regression (PLSR) modeling

PLSR is a statistical method that bears some relation to principal components regression. It is often applied to find a linear regression model by projecting the predicted variables (matrix X) and the observable variables (matrix Y) to a new space. PLSR is especially useful in quite common case where the number of descriptors (independent variables) is comparable to or greater than the number of compounds (data points) and/or there exist other factors leading to correlations between variables. De

May et al. [1] applied the PLSR method to characterize the relationship between sequences and strengths of E. coli promoters and established the prediction model, and then they found a good correlation between promoter sequence and strength. Under the guidance of model prediction and promoter knock-in technology, the authors chose the suitable strength of promoters to tune the expression of pathway genes and further optimize the metabolic pathway [30].

#### Position Weight Matrix (PWM) modeling

PWM is a commonly used representation of motifs (patterns) in biological sequences. For DNA sequences, PWM is based on the independent contribution of each base pair to the relevant function. A PWM has one row for each kind of nucleotide and one column for each position in the pattern. PWMs are often derived from a set of aligned sequences that are thought to be functionally related and they have become an important part of many software tools for computational motif discovery. PWM models are commonly used to predict the promoter regions and transcription factor binding sites since they are simple and have a predictive success comparable to that of more complex models [31]. Rhodius et al. [4] indicated that PWM models can also be used to predict the promoter strength by its sequence. They proposed a new method to predict the strength of  $\sigma^{E}$  recognized promoters in *E. coli* based on a PWM prediction model. They divided all  $60 \sigma^{E}$  promoters into several functional motifs including -35 motif, -10 motif, start motif. spacer, discriminator, and initial transcribed region, and correlated the measured strength with the sum of PWM scores of each functional motif. As a result, they found that the sum score of part of the motifs has a good correlation to the strength of the promoter and has a coefficient correlation R value ranged from 0.57 to 0.77.

#### Artificial Neural Network (ANN) modeling

Although aforementioned modeling methods have made some progress in prediction of cis-acting element strength, most of the current models still fail to achieve high precision of strength prediction. In fact, due to the extremely complex nonlinear relationship between element sequence and strength, simple regression analysis methods commonly adopted by these models do not well reflect the nonlinear relationship. This reduces the prediction accuracy for sequence design. Artificial intelligence algorithms such as ANN are available to characterize the nonlinear relationship between inputs and outputs to build prediction models with higher accuracy.

ANN simulates the structure and functional aspects of human brain neural networks. The weight of neuron

connection can be changed to be a suitable value after learning knowledge from training data set. ANN models have been widely used in various research fields of life sciences, e.g., protein structure prediction [32,33], protein stability prediction after mutations [34], RNA secondary structure prediction [35], and promoter recognition and structure analysis [36–38]. Further, our lab put forward a modeling approach based on ANN to characterize the nonlinear relationship [2] (Figure 1C). The best training ANN model gets a high regression correlation coefficient of 0.98 for both model training and test. Both the correlation coefficient of non-linear regression and the predicting accuracy are significantly improved when compared to previous works. The quantitative design sequences with desired strength were also successfully applied to improve the expression of a small peptide BmK1 and fine-tune a key enzyme gene dxs for pathway engineering of terpenoids biosynthesis in E. coli. The predicting methodology and models are competent for de novo design of fine-tuning promoter and/or RBS elements with desired properties, which are of highly significance and importance for synthetic biology applications.

### Tools and websites for cis-acting regulatory elements design

The establishment of aforementioned modeling technologies accelerates the construction of quantitative prediction models, and some useful tools and website applications have also been developed for design of cisacting regulatory elements such as promoters and RBSs (Table 2). For example, Na et al. [19] developed a tool named RBSDesigner (http://ssbio.cau.ac.kr/) which can predict the translation efficiency of existing mRNA sequences and design synthetic RBS for a given coding sequence to yield a desired level of protein expression. For promoter design, PromoterCAD is a web-based User Interface developed by Nishikata et al. [20,21] for datadriven regulatory DNA design. The web server (http:// promotercad.org) allows the design of synthetic promoters with novel regulation functions to control gene expression in plants or mammals. PromoterCAD collected high-throughput expression and motif data from Arabidopsis thaliana and Mus musculus to guide synthetic plant or mammal promoter design.

#### **PERSPECTIVES**

Building genetic elements with desired strength or activity and building computational models that can predict the activity of a genetic element has been a real challenge in gene expression area over decades. More recently immense interest in synthetic biology and its applications in engineering pathways and producing valuable chemicals in microbes has brought back the sequence-activity modeling approaches to the forefront [2]. Understanding the key sequence determining the strength of elements will be in general valuable in predicting parts in ever increasing genome sequence data.

Biological systems are dynamic, fine regulated, nonlinear complex systems that are extremely difficult to predict their behaviors. The designability of biological systems is one of the basic hypotheses in synthetic biology research. Can we really achieve this goal? How can we achieve this goal? For quite some time, synthetic biologists have made great efforts and try to answer these questions. For cis-acting regulatory element design, researchers are often obsessed with troublesome problems. One of the problem is: how to characterize the correlation between sequence feature and the strength? To solve this, they have built many quantitative prediction models using different kinds of modeling technologies. How to understand and describe the extremely complex relationship thus becomes the key step toward the element sequence design. Another problem that should be consider is the "context effect" for quantitative design of elements. Take transcriptional circuit design for instance, Lou et al. [39] found that the sequence at the junction between the input promoter and circuit can affect the input-output response of the circuit. They gave a solution to the problem by designing ribozymes to construct a library of "insulator parts", which can be used to join synthetic gene circuits and the behavior of layered circuits can be predicted by mathematical models.

Existing modeling approaches provide methodologies from different perspectives to construct quantitative prediction models and achieve encouraging progress. In the near future, quantitative characterization and standardization of regulatory elements will promote the building of prediction models for different kinds of elements and organisms. With increasing understanding of the intrinsic complex mechanisms as well as the accumulation of standardized data in literatures or databases, it would help us easily to build quantitative models with high precision to facilitate both irrational and rational design of specific cis-acting regulatory elements with desired properties.

#### **ACKNOWLEDGEMENTS**

This work was supported by the National Basic Research Program of China (973 Program, grant No. 2012CB721104), the National High Technology Research and Development Program (863 Program, grant No. 2012AA02A701), the National Natural Science Foundation of China (grant Nos. 31170101 and 31301017), and the Natural Science Foundation of Guangdong Province, China (grant No. 2015A030310317).

#### COMPLIANCE WITH ETHICS GUIDELINES

The authors Hailin Meng and Yong Wang declare they have no conflict of

interests

This article does not contain any studies with human or animal subjects performed by any of the authors.

#### REFERENCES

- 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
- 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
- Wang, J., Meng, H., Xiong, Z. and Wang, Y. (2013) Design and construction of artificial biological systems for complex natural products biosynthesis. Chinese J. Biotech. (in Chinese), 29, 1146–1160
- Rhodius, V. A. and Mutalik, V. K. (2010) Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, σ<sup>E</sup>. Proc. Natl. Acad. Sci. USA, 107, 2854–2859
- Salis, H. M., Mirsky, E. A. and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol., 27, 946–950
- Canton, B., Labno, A. and Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices. Nat. Biotechnol., 26, 787–793
- Yuan, Y., Liu, B., Xie, P., Zhang, M. Q., Li, Y., Xie, Z. and Wang, X. (2015) Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. Proc. Natl. Acad. Sci. USA, 112, 3158–3163
- Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A. and Arkin, A. P. (2012) RNA processing enables predictable programming of gene expression. Nat. Biotechnol., 30, 1002–1006
- Alper, H., Fischer, C., Nevoigt, E. and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. Proc. Natl. Acad. Sci. USA, 102, 12678–12683
- Wong, W. W., Tsai, T. Y. and Liao, J. C. (2007) Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. Mol. Syst. Biol., 3, 130
- Yamanishi, M., Ito, Y., Kintaka, R., Imamura, C., Katahira, S., Ikeuchi, A., Moriya, H. and Matsuyama, T. (2013) A genome-wide activity assessment of terminator regions in Saccharomyces cerevisiae provides a "terminatome" toolbox. ACS Synth. Biol., 2, 337–347
- Siegl, T., Tokovenko, B., Myronovskyi, M. and Luzhetskyy, A. (2013)
   Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metab. Eng., 19, 98–106
- Qin, X., Qian, J., Yao, G., Zhuang, Y., Zhang, S. and Chu, J. (2011)
   GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl. Environ. Microbiol., 77, 3600–3608
- Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., et al. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. Nat. Methods, 10, 354–360
- Brewster, R. C., Jones, D. L. and Phillips, R. (2012) Tuning promoter strength through RNA polymerase binding site design in *Escherichia* coli. PLoS Comput. Biol., 8, e1002811
- 16. Na, D., Lee, S. and Lee, D. (2010) Mathematical modeling of

- translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. BMC Syst. Biol., 4, 71
- Juven-Gershon, T., Cheng, S. and Kadonaga, J. T. (2006) Rational design of a super core promoter that enhances gene expression. Nat. Methods, 3, 917–922
- Vilar, J. M. (2010) Accurate prediction of gene expression by integration of DNA sequence statistics with detailed modeling of transcription regulation. Biophys. J., 99, 2408–2413
- Na, D. and Lee, D. (2010) RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression. Bioinformatics, 26, 2633–2634
- Nishikata, K., Cox, R. S. 3rd, Shimoyama, S., Yoshida, Y., Matsui, M., Makita, Y. and Toyoda, T. (2014) Database construction for PromoterCAD: synthetic promoter design for mammals and plants. ACS Synth. Biol., 3, 192–196
- Cox, R. S. 3rd, Nishikata, K., Shimoyama, S., Yoshida, Y., Matsui, M., Makita, Y. and Toyoda, T. (2013) PromoterCAD: Data-driven design of plant regulatory DNA. Nucleic Acids Res., 41, W569-W574
- Ham, T. S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N. J. and Keasling, J. D. (2012) Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. Nucleic Acids Res., 40, e141
- Jensen, P. R. and Hammer, K. (1998) Artificial promoters for metabolic optimization. Biotechnol. Bioeng., 58, 191–195
- Jensen, P. R. and Hammer, K. (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl. Environ. Microbiol., 64, 82–87
- Andersen, H. W., Pedersen, M. B., Hammer, K. and Jensen, P. R. (2001) Lactate dehydrogenase has no control on lactate production but has a strong negative control on formate production in *Lactococcus lactis*. Eur. J. Biochem., 268, 6379–6389
- Bakke, I., Berg, L., Aune, T. E., Brautaset, T., Sletta, H., Tøndervik, A. and Valla, S. (2009) Random mutagenesis of the PM promoter as a powerful strategy for improvement of recombinant-gene expression. Appl. Environ. Microbiol., 75, 2002–2011
- Solem, C. and Jensen, P. R. (2002) Modulation of gene expression made easy. Appl. Environ. Microbiol., 68, 2397–2403

- Solem, C., Koebmann, B. J. and Jensen, P. R. (2003) Glyceraldehyde-3phosphate dehydrogenase has no control over glycolytic flux in *Lactococcus lactis* MG1363. J. Bacteriol., 185, 1564–1571
- Curran, K. A., Morse, N. J., Markham, K. A., Wagman, A. M., Gupta,
   A. and Alper, H. S. (2015) Short synthetic terminators for improved heterologous gene expression in yeast. ACS Synth. Biol., 4, 824–832
- De Mey, M., Maertens, J., Boogmans, S., Soetaert, W. K., Vandamme, E. J., Cunin, R. and Foulquié-Moreno, M. R. (2010) Promoter knockin: a novel rational method for the fine tuning of genes. BMC Biotechnol., 10, 26
- Stormo, G. D. (2000) DNA binding sites: representation and discovery. Bioinformatics, 16, 16–23
- Kakumani, R., Devabhaktuni, V. and Ahmad, M. (2008) A two-stage neural network based technique for protein secondary structure prediction. Conf. Proc. IEEE Eng. Med. Biol. Soc., 2008, 1355–1358
- Qu, W., Sui, H., Yang, B. and Qian, W. (2011) Improving protein secondary structure prediction using a multi-modal BP method. Comput. Biol. Med., 41, 946–959
- Capriotti, E., Fariselli, P. and Casadio, R. (2004) A neural-networkbased method for predicting protein stability changes upon single point mutations. Bioinformatics, 20, i63–i68
- Koessler, D. R., Knisley, D. J., Knisley, J. and Haynes, T. (2010) A
  predictive model for secondary RNA structure using graph theory and a
  neural network. BMC Bioinformatics, 11, S21
- Wang, J., Ungar, L. H., Tseng, H. and Hannenhalli, S. (2007) MetaProm: a neural network based meta-predictor for alternative human promoter prediction. BMC Genomics, 8, 374
- Askary, A., Masoudi-Nejad, A., Sharafi, R., Mizbani, A., Parizi, S. N. and Purmasjedi, M. (2009) N4: a precise and highly sensitive promoter predictor using neural network fed by nearest neighbors. Genes Genet. Syst., 84, 425–430
- de Avila E Silva, S., Gerhardt, G. J. and Echeverrigaray, S. (2011) Rules extraction from neural networks applied to the prediction and recognition of prokaryotic promoters. Genet. Mol. Biol., 34, 353– 360
- Lou, C., Stanton, B., Chen, Y. J., Munsky, B. and Voigt, C. A. (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. Nat. Biotechnol., 30, 1137–1142