

Review: Mathematical models for designing synthetic ribosome binding sites to control protein expression in prokaryotes

## INTRODUCTION

Engineering novel cellular functions in microorganisms often requires tight control over expression of both native and introduced proteins. For example, high expression levels are required for producing proteins for therapeutic and industrial applications [1]. Expression of protein systems like multidrug resistance efflux transporters may require the coordinated expression of each subunit at the appropriate ratio to produce functional systems [2]. Additionally, engineering microbes for the synthesis of chemical products may require introduction of several genes encoding the enzymes in a metabolic pathway. These genes may require a delicate balance of expression levels to circumvent bottlenecks in the synthesis pathway [2,3, 4]. As protein expression levels often directs the function of genetic systems like these, fine tuned control over protein expression is required in their design.

Protein expression is regulated at essentially every step in protein synthesis. Transcriptional level regulation is often a primary consideration when designing synthetic gene circuits through the choice and/or design of promoter systems to provide the desired transcriptional rates [5,6]. However, a number of studies have demonstrated that translation rates can also significantly influence protein expression levels. Although transcript properties such as mRNA stability and codon usage are known to influence microbial translation, initiation is often the rate-limiting step in translation [7,8]. The rate of translation initiation, and subsequently protein expression levels, is determined by the sequence of the 5' end of the mRNA transcript. Several studies have demonstrated improved expression levels of proteins *Escherichia coli* by introducing mutation in only the 5' end of the mRNA [8,9,10].

Translation initiation in prokaryotes is facilitated by the hybridization of the 3' end of the 16S rRNA with a site on the mRNA as well as the binding of the tRNA<sup>fMET</sup> to the start codon. A number of sequence-specific characteristics have been identified that can decrease translation initiation rates, and therefore result in low expression levels. The presence of secondary structure in the 5' region of the mRNA can limit translation due to sequestration of the 16S hybridization region (i.e. Shine-Dalgarno sequence) or the start codon [11]. Secondary structures

that limit translation efficiency may be present in the 5' untranslated region of the transcript as well as in the coding region downstream of the initiation codon [12]. Therefore, use of previously characterized translation initiation sequences with different proteins may not reliably produce the same translation initiation rate.

Additionally, it has been shown that the distance between the start codon and the hybridization site, called "spacing," also affects the efficiency of translation initiation. It is expected that optimal spacing is due to the physical distance between the 16S rRNA and the P-site (entry site for the initiation tRNA) in the ribosome [13,14]. Nonoptimal spacing may require deformation of the mRNA sequence to allow for hybridization of the 16S rRNA and the tRNA, thereby decreasing the rate of initiation of translation [15].

Due to the influence of the mRNA sequence on protein expression levels, several strategies have been developed for optimizing the 5' end of the transcript. Both library and computation-based methods have met with some success in providing desired expression in prokaryotes. Several studies have utilized libraries of mRNA sequences to improve protein expression levels as well as the function of synthetic genetic systems [2,3,16]. While library approaches have been used with some success, as the number of proteins in the genetic system increases the library size also increases combinatorially [17]. Thus, complex multi-protein genetic systems may require screening of hundreds of sequences to tune expression of each component [2]. Large library screens may require a considerable amount of time as well as robust, high-throughput methods to identify sequences that provide the desired functional output.

As efficient screening techniques may not be available for specific systems and the screening of large libraries become increasingly cost and time prohibitive, computational methods that design genetic components with predictive *in vivo* function become necessary alternatives. This review will describe methods for improving translational efficiency, highlighting two recently developed computational algorithms for the design of synthetic ribosome binding sites. Ribosome binding sites (RBS) in the context of this review refers to the 5' region of the mRNA transcript which flanks the start codon and influences translation initiation.

## **STRATEGIES FOR OPTIMIZING TRANSLATION INITIATION**

### **Minimization of mRNA Secondary Structure**

The presence of secondary structure in the mRNA transcript has been shown to influence translation initiation due to sequestration of the 16S rRNA hybridization sequence, region for ribosome docking, and/or the start codon [11]. To address this, algorithms for predicting RNA secondary structure have been used to minimize secondary structure in RBS sequences. Mfold is one of the most commonly used programs for predicting nucleic acid secondary structure and has been used extensively in designing RBS sequences. Mfold determines minimal free energy of predicted folded structures [18]. Several groups have demonstrated that mutations in the 5' region of the mRNA transcript that destabilize secondary structures as calculated by programs like Mfold result in improved expression levels [9,10]. This practice has primarily been used to achieve high levels of expression of one or more genes [1,19]. However, this strategy may not allow for more accurate control of translation initiation rates (and protein expression) as the presence of secondary structure is just one limiting factor in translation initiation.

### **RBS Calculator**

#### *Algorithm Summary*

As engineered genetic systems become increasingly complex, RBS design strategies that focus solely on minimizing mRNA secondary structure may not be sufficient to rationally design systems that operate as desired. Salis, et al recently developed the “RBS Calculator” which allows for both the prediction of translation initiation rates for existing RBS sequences (reverse engineering) as well as the design of synthetic ribosome binding sites with user-specified target initiation rates (forward engineering) [17]. The RBS Calculator builds upon previous methods for enhancing translation initiation that focused primarily on minimization of 5'mRNA secondary structure. In addition to secondary structure, the Calculator takes into account binding affinity for the 16S rRNA and spacing building a more complete model of translation initiation.

The RBS Calculator models the molecular interactions involved in translation initiation taking into account ribosome binding occlusion due to mRNA secondary structure as well as affinity between the transcript and the 16S rRNA. The model predicts the free energy change ( $\Delta G_{tot}$ ) during the transition from a free mRNA transcript to the mRNA subunit bound to the 30S ribosomal subunit (Figure 1).



**Figure 1: Model of translation initiation in the RBS Calculator**

The RBS Calculator determines the free energy change ( $\Delta G_{tot}$ ) in the transition from free mRNA and the 30S ribosomal subunit to the final state of a 30S bound mRNA. Adapted from ref. [17].

The overall change in free energy ( $\Delta G_{tot}$ ) is determined as follows:

$$\Delta G_{tot} = \Delta G_{mRNA:rRNA} + \Delta G_{start} + \Delta G_{spacing} - \Delta G_{standby} - \Delta G_{mRNA}$$

in which:

- 1)  $\Delta G_{mRNA:rRNA}$  = energy released upon hybridization of the 3' end of 16S rRNA with the mRNA (higher the affinity, the more negative this term)
- 2)  $\Delta G_{start}$  = energy released upon binding of the tRNA<sup>fMet</sup> with the start codon
- 3)  $\Delta G_{spacing}$  = energy penalty for suboptimal spacing (determined empirically for *E. coli*)
- 4)  $\Delta G_{standby}$  = work required to unfold secondary structure in the standby site that may occlude ribosome from docking on the transcript
- 5)  $\Delta G_{mRNA}$  = work required to unfold mRNA sequence surround start codon

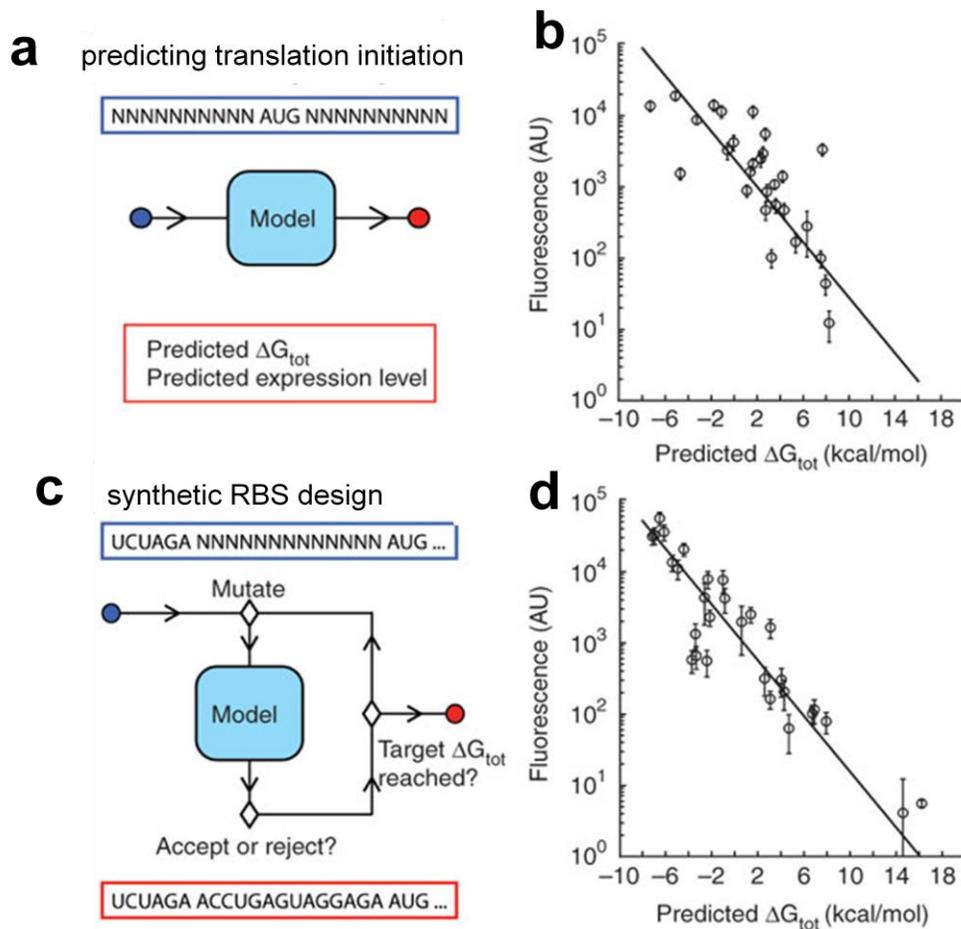
Calculations of the free energy changes for RNA hybridization interactions are determined using MFold v3.0 and NuPack algorithms. The energetic penalty for nonoptimal spacing was determined empirically for *E. coli* by fitting protein expression levels for synthetic RBS with varying spacing lengths while holding all other free energy changes constant [17].

For predicting translation initiation rates, the RBS Calculator scans the mRNA transcript for the optimal 16S rRNA hybridization sequence which balances the energy released upon hybridization with energetic penalties due to imperfect spacing. The algorithm correlates the predicted  $\Delta G_{tot}$  with a probability of the ribosome being bound to the mRNA transcript using the Boltzmann distribution. This probability is then proportional to the translation initiation rate, and subsequently protein expression levels, as shown below:

$$\text{Translation initiation rate} = r \propto \exp(-\beta \Delta G_{tot})$$

This equation allows for the prediction of the *relative* translation initiation rate. Based on the model, lower values of  $\Delta G_{tot}$  (i.e. more negative) should result in increased translation initiation rates.  $\beta$  corresponds to the apparent Boltzmann constant for the system and was empirically determined in *E. coli*.

This statistical thermodynamic model which predicts translation initiation rates was utilized with an algorithm to design synthetic ribosome sites. The user inputs the targeted translation initiation rate (on a relative scale ranging from 1 to  $10^5$ ) and the protein coding sequence. The algorithm begins with a random upstream sequence and iteratively makes nucleotide mutations and calculates the  $\Delta G_{tot}$  until the calculated initiation rate approaches the target rate. The authors demonstrated the ability of the algorithm to both predict translation initiation rates of existing RBS sequences as well as design synthetic sequences that provide a wide range of target initiation rates and therefore, expression levels (Figure 2).



**Figure 2: Use of RBS Calculator to predict initiation rates and design RBS sequences**  
 (A) Predicting translation initiation rates of a given RBS sequence. (B) Expression of red fluorescent protein (RFP) driven by existing RBSs and the corresponding  $\Delta G_{tot}$  as predicted by the RBS Calculator. (C) Designing synthetic RBS sequence: given a target initiation rate, the algorithm iteratively generates random mutations in an initial sequence and calculates the  $\Delta G_{tot}$  until the target is reached. (D) Expression of RFP under the control of RBS sequences designed by the RBS Calculator and the target  $\Delta G_{tot}$  used to design RBSs. Adapted from ref. [17].

### *Limitations of Algorithm*

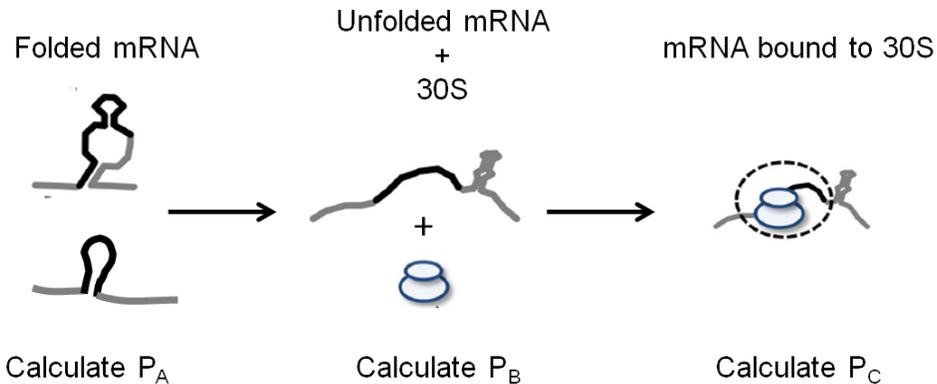
In general, previous attempts to rationally improve ribosome binding sequences have focused on maximizing expression levels [1,7]. The RBS Calculator, however, predicts sequences with translation rates (and therefore protein expression levels) over a range of five orders of magnitude. Although the RBS Calculator was demonstrated to be a relatively efficient method for predicting RBS sequences that provided desired expression levels, the algorithm is not perfect. There is a 47% probability that the output sequence will provide expression levels within two-fold of the target level. Therefore, testing several output sequences will likely be required to identify the sequence that will provide the desired expression level *in vivo* [17]. Compared with library-based approaches, the RBS Calculator can significantly reduce the number of translation initiation sites that must be screened [2].

In general, *E. coli* is the organism of choice for most microbial metabolic engineering and synthetic biology systems however, other microbial hosts have been explored for metabolic engineering and synthetic biology systems [1,19]. The RBS calculator was developed and tested in *E. coli* expression systems so it may not have the same predictive power in other microbial hosts. The RBS calculator does allow the user to specify the sequence of the 3' end of the 16S rRNA for other prokaryotic organisms thereby allowing the calculator to be used in other expression systems. However, the other parameters determined empirically for *E. coli* including energetic penalties for nonoptimal spacing and the  $\beta$  conversion factor (that translates the calculated  $\Delta G_{\text{tot}}$  to a relative translation initiation rate) are not adjusted. Gram negative bacteria, such as *E. coli*, share similarities in their translational machineries and the RBS Calculator has been implemented in two different species of gram-negative bacteria, *Klebsiella* M5AL and *Salmonella enterica* (<http://voigtlab.ucsf.edu/software/>). Gram positive bacteria are known to differ in some characteristics of their translational machinery in comparison to gram negative. For example, *Bacillus subtilis* was shown to exhibit less tolerance to suboptimal spacing and secondary structure than *E. coli*. Therefore, the energetic penalties for spacing and unfolding secondary structure may need to be weighted if the algorithm is to be implemented in *B. subtilis* [15]. In general, it seems that additional testing and parameter estimation will need to be performed to optimize the predictive power of the RBS Calculator in different organisms.

### **RBSDesigner**

#### *Algorithm Summary*

Na, et al. recently developed a mathematical model for predicting RBS translational initiation rates and applied their model to designing synthetic RBS sequences to control protein expression levels over a five order of magnitude range [20]. Although the algorithm differs slightly from the RBS Calculator, the approach is similar. The RBS Calculator sums the energetic contributions of different elements in translation initiation (16S rRNA hybridization, spacing, etc) into a final  $\Delta G_{tot}$ . This parameter is then converted into a probability for translation initiation. Na, et al. also uses a similar statistical thermodynamics approach; however, their algorithm calculates the probability of sequential stages of translation initiation based on free energy change (Figure 3). These probabilities are then applied to a system of ordinary differential equations (ODE) that model ribosome binding to unfolded mRNA at steady state. The ODE model requires several parameters including the steady state quantity of ribosomes, mRNA transcripts, and the number of ribosomes that can be on a given mRNA at any given time. These values of these parameters were estimated from the literature for *E. coli*.



**Figure 3: Summary of model for estimating translation efficiency.** Adapted from ref. [20]. Sequence of translation initiation events that are included in algorithm:

- 1) Determine all possible secondary structures of the mRNA transcript
- 2) Calculate probabilities of occupying each conformation based on their free energy ( $P_A$ )
- 3) Calculate probability of each structure unfolding thereby allowing ribosome to bind ( $P_B$ )
- 4) Calculate probability that unfolded structures will be occupied by a ribosome ( $P_C$ )

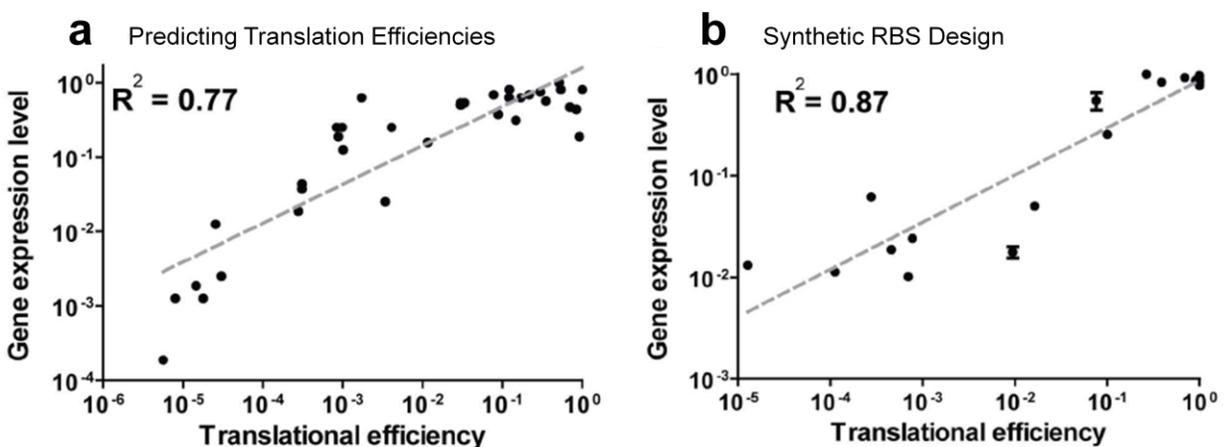
$$\text{Translation Efficiency} = \frac{\text{mRNA bound to ribosome}}{\text{Total mRNA}} = P_C$$

The final probability,  $P_C$ , is similar to the translation initiation rate used by the RBS Calculator; both are essentially a measure of the relative number of mRNA transcripts bound to ribosomes and therefore actively producing protein. The key parameters assessed in both the RBS Calculator and the Na, et al. algorithm are essentially the same: ribosome affinity and mRNA

secondary structure that may occlude ribosome binding [17,20]. Although not included in the initial design, the publically available version of the algorithm, RBSDesigner, was updated to incorporate penalties due to suboptimal spacing (similar to the RBS Calculator, although the details of the modified algorithm have not been presented) [21]. One improvement in this algorithm is that it allows multiple mRNA secondary structures while the RBS Calculator uses only a single folded mRNA transcript state [17,20].

The optimization algorithm used in RBSDesigner to design synthetic RBS sequences also differs slightly from the RBS Calculator. Both iteratively make mutations in randomized RBS sequences until the target efficiency (or initiation rates for the RBS Calculator) is reached. However, the RBSDesigner performs calculations for 100 sequences simultaneously and outputs the 10 sequences closest to the target efficiency while the calculator outputs a single sequence [20,21]. Therefore, the RBS Calculator must be run several times with the same input parameters to obtain multiple sequences. However, in personal tests of the RBS Calculator the designed RBS sequence was determined within minutes, while the RBSDesigner may be on the order of hours [21]. This time difference may be due in part due to the complexity of the algorithm (scanning 100 sequences at once as opposed to one), but also the computing power available. Currently the RBS Calculator is publically available via web server, while the RBS Designer must be downloaded onto the user computer.

The algorithm developed by Na, e al. was also demonstrated to be able to fairly accurately estimate translational efficiencies (i.e. expression levels) as well as design RBS sequences with a five-order of magnitude range of expression levels (Figure 4).



**Figure 4: RBSDesigner for predicting translation efficiencies and design RBS sequences**

(A) Estimated translation efficiencies of existing RBS sequences for the MS2 coat gene. (B) Expression levels of *luxR* under control of synthetic RBS sequences as a function of the target translation efficiency. Adapted from ref. [20].

### Limitations

As the algorithm of RBSDesigner considers the same key parameters of translation initiation as the RBS Calculator, it suffers from a number of the same limitations. Both algorithms will likely require testing of several output sequences to identify the RBS that provides the desired output function. Additionally, RBSDesigner was designed and tested in *E. coli*, therefore may not work efficiently in other microbial expression systems.

Several parameters of the RBS Calculator used to design synthetic RBS sequences were fit from empirical studies in *E. coli*. The RBSDesigner used parameters obtained from the literature including the steady state quantities of ribosomes and mRNA transcripts in *E. coli* [20]. These parameters can differ significantly even within *E. coli* between a number of experimental conditions. Therefore, the predictive power of the RBSDesigner is unclear beyond the conditions tested in this study. The RBS Calculator was demonstrated to maintain its ability to predict relative expression levels when the protein of interest was under the control of promoters of varying strength [17]. As the RBSDesigner relies on estimation of mRNA transcript quantity, conditions that perturb this quantity such as the use of different promoter strengths may reduce its predictive power. Additionally, the gene copy number (such as using plasmids with varying copy number), mRNA stability, and the physiological state of the cell will impact the parameters used in the RBSDesigner algorithm [4,17]. Additional testing is required even within *E. coli* expression systems to determine the ability of RBSDesigner to accurately predict translational efficiencies and design synthetic RBS sequences under different experimental conditions.

## FUTURE DIRECTIONS AND CONCLUSIONS

The development of the RBS Calculator and RBSDesigner demonstrate a significant step towards computational approaches to design a single component of synthetic genetic systems, ribosome binding sites. Previous work has focused primarily on maximizing protein expression through minimization of 5' secondary structure. However, engineered genetic systems may require more predictive and fine-tuned control over expression levels. These computational methods are a step towards the rational design of RBS sequences and may allow for testing of significantly fewer sequences than library-based methods.

As the development of both algorithms is fairly recent, further experimentation will need to be performed to further demonstrate the predictive power of these algorithms under different conditions such as through the expression of different proteins, under different promoters, etc. For example, due to a lack of experimental evidence it is currently unclear if these RBS design algorithms would be able to design RBS sequences in the intergenic regions of operons to tune the relative expression of the genes contained on the operon. As the transcript abundance of each gene on the operon should be equivalent, post-transcriptional level regulation is the primary method of controlling relative expression levels [2,4]. Thus, the RBS algorithms reviewed here may be a useful tool for designing the intergenic regions of synthetic operons. However, further experimentation will need to be performed since translation initiation may not be the only method of controlling expression of the genes on the operon.

Both the RBS Calculator and RBSDesigner may be primarily limited to use in an *E. coli* expression host, unless parameters are experimentally determined and validated in other organisms. In addition to organism specificity, there are several areas that may be explored for improving the predictive power of these RBS design algorithms. Several properties of the mRNA were not included in the model have been shown to influence translational efficiencies. For example, codon usage has been demonstrated to impact expression levels, potentially due to ribosome stalling and premature translation termination [1]. Computational algorithms have been developed for increasing protein expression levels based on optimization of the codon usage [22,23]. Additionally, suboptimal codon usage may cause translation elongation to be the rate limiting step in translation, thereby reducing the predictive power of the RBS algorithms which assume initiation is the rate-limiting step. Therefore, such algorithms may be used in tandem with RBS design programs to improve the predictive power of the RBS design algorithms and therefore the desired expression levels [21].

The stability of mRNA transcripts *in vivo* is also not considered in either algorithm. The degradation rate of mRNA is dependent on elements in the transcript that protect against enzymatic degradation such as the presence of RNase sites and/or stabilizing secondary structures [2,4]. Neither algorithm accounts for the presence of RNase binding sites. Additionally, secondary structure is considered at the level of occluding translation initiation, however, not as potential protection against endonuclease degradation. While the factors that control RNA susceptibility are less studied than RBS and codon usage, they are an example of

one area that may be explored to further improve the design algorithms to more accurately model translation *in vivo* [4].

One last point of consideration is that although initiation is often the rate limiting step during translation, it is only one stage in the production of proteins. Other factors such as transcription rates and protein folding efficiencies can significantly influence expression levels [19]. Regulation at each protein synthesis stage may need to be controlled to produce functioning engineered genetic systems. As engineered genetic systems grow increasingly complex, optimization of each component becomes increasingly inefficient. Therefore, models and design algorithms for various genetic elements such as ribosome binding sites may facilitate the rapid and accurate design of entire systems [5]. The RBS design algorithms presented in this review, along with models of promoters and other genetic elements, are a step towards the goal of predictive computational design of synthetic genetic systems. Although further experimental validation will likely need to be performed, computational methods for RBS design including the RBS Calculator and the RBSDesigner will likely have significant impacts in protein production technology, metabolic engineering, and synthetic genetic circuit design.

## REFERENCES

- [1] G. Hannig and S.C. Makrides, "Strategies for optimizing heterologous protein expression in *Escherichia coli*," *Trends in biotechnology*, vol. 16, 1998, pp. 54-60.
- [2] B.F. Pflieger, D.J. Pitera, C.D. Smolke, and J.D. Keasling, "Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes.," *Nature biotechnology*, vol. 24, 2006, pp. 1027-32.
- [3] E.L. Haseltine and F.H. Arnold, "Synthetic gene circuits: design with directed evolution.," *Annual review of biophysics and biomolecular structure*, vol. 36, 2007, pp. 1-19.
- [4] C.D. Smolke and J.D. Keasling, "Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon.," *Biotechnology and bioengineering*, vol. 80, 2002, pp. 762-76.
- [5] K. Clancy and C.a. Voigt, "Programming cells: towards an automated 'Genetic Compiler'," *Current Opinion in Biotechnology*, vol. 21, 2010, pp. 572-581.
- [6] J.R. Kelly, A.J. Rubin, J.H. Davis, C.M. Ajo-Franklin, J. Cumbers, M.J. Czar, K. de Mora, A.L. Gliberman, D.D. Monie, and D. Endy, "Measuring the activity of BioBrick promoters using an in vivo reference standard.," *Journal of biological engineering*, vol. 3, 2009, p. 4.
- [7] M. Welch, A. Villalobos, C. Gustafsson, and J. Minshull, "You're one in a googol: optimizing genes for protein expression.," *Journal of the Royal Society, Interface / the Royal Society*, vol. 6 Suppl 4, 2009, pp. S467-76.
- [8] G.W. Hatfield and D.A. Roth, "Optimizing scaleup yield for protein production : Computationally Optimized DNA Assembly ( CODA ) and Translation," *Biotechnology Annual Review*, vol. 13, 2007, pp. 27-42.
- [9] R. Cèbe and M. Geiser, "Rapid and easy thermodynamic optimization of the 5'-end of mRNA dramatically increases the level of wild type protein expression in *Escherichia coli*," *Protein expression and purification*, vol. 45, 2006, pp. 374-80.
- [10] B.F. Pflieger, N.J. Fawzi, and J.D. Keasling, "Optimization of DsRed production in *Escherichia coli*: effect of ribosome binding site sequestration on translation efficiency.," *Biotechnology and bioengineering*, vol. 92, 2005, pp. 553-8.
- [11] M.H. de Smit and J. van Duin, "Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, 1990, pp. 7668-72.
- [12] S.W. Seo, J. Yang, and G.Y. Jung, "Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in *Escherichia coli*," *Biotechnology and bioengineering*, vol. 104, 2009, pp. 611-6.
- [13] H. Chen, M. Bjerknes, R. Kumar, and E. Jay, "Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs.," *Nucleic acids research*, vol. 22, 1994, pp. 4953-7.
- [14] S. Ringquist, "Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site," *Molecular Microbiology*, vol. 52, 1992, pp. 687-1229.
- [15] R.L. Vellanoweth, "The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo," *Molecular Microbiology*, vol. 12, 1992, pp. 249-1114.
- [16] K. Min, M. Kim, and D. Lee, "Search for the optimal sequence of the ribosome binding site by random oligonucleotide-directed mutagenesis," *Nucleic acids research*, vol. 16, 1988, pp. 5075-88.
- [17] H.M. Salis, E.a. Mirsky, and C.a. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression.," *Nature biotechnology*, vol. 27, 2009, pp. 946-50.
- [18] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Research*, vol. 31, 2003, pp. 3406-3415.

- [19] S.C. Makrides, "Strategies for achieving high-level expression of genes in *Escherichia coli*," *Microbiological reviews*, vol. 60, 1996, pp. 512-38.
- [20] D. Na, S. Lee, and D. Lee, "Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes.," *BMC systems biology*, vol. 4, 2010, p. 71.
- [21] D. Na and D. Lee, "RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression.," *Bioinformatics (Oxford, England)*, vol. 26, 2010, pp. 2633-4.
- [22] M. Welch, S. Govindarajan, J.E. Ness, A. Villalobos, A. Gurney, J. Minshull, and C. Gustafsson, "Design parameters to control synthetic gene expression in *Escherichia coli*," *PloS one*, vol. 4, 2009, p. e7002.
- [23] P. Puigbò, E. Guzmán, A. Romeu, and S. Garcia-Vallvé, "OPTIMIZER: a web server for optimizing the codon usage of DNA sequences.," *Nucleic acids research*, vol. 35, 2007, pp. W126-31.