BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 20 2010, pages 2633–2634
doi:10.1093/bioinformatics/btq458

Gene expression

Advance Access publication August 11, 2010

RBSDesigner: software for designing synthetic ribosome binding sites that yields a desired level of protein expression

Dokyun Na† and Doheon Lee*

Department of Bio and Brain Engineering, KAIST, 373-1 335-Gwahangno, Yuseong-gu, Daejeon 305-701, South Korea

Associate Editor: Martin Bishop

ABSTRACT

Motivation: RBSDesigner predicts the translation efficiency of existing mRNA sequences and designs synthetic ribosome binding sites (RBSs) for a given coding sequence (CDS) to yield a desired level of protein expression. The program implements the mathematical model for translation initiation described in Na *et al.* (Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with a desired expression level in prokaryotes. *BMC Syst. Biol.*, 4, 71). The program additionally incorporates the effect on translation efficiency of the spacer length between a Shine–Dalgarno (SD) sequence and an AUG codon, which is crucial for the incorporation of fMet-tRNA into the ribosome. RBSDesigner provides a graphical user interface (GUI) for the convenient design of synthetic RBSs.

Availability: RBSDesigner is written in Python and Microsoft Visual Basic 6.0 and is publicly available as precompiled stand-alone software on the web (http://rbs.kaist.ac.kr).

Contact: dhlee@kaist.ac.kr

Received on June 23, 2010; revised on August 1, 2010; accepted on August 6, 2010

1 BACKGROUND

Controlling protein expression levels is a major challenge in several areas of biology. It ensures the robust operation of synthetic circuits in synthetic biology, enables high levels of production of therapeutic proteins in biotechnology, and optimizes the flux of metabolic pathways in metabolic engineering (Na *et al.*, 2010a).

Over several decades, biochemical studies have unraveled the mechanisms underlying transcription in prokaryotes, and these mechanisms have enabled the control of gene expression. However, a few nucleotide changes in the untranslated region of an mRNA can also introduce drastic changes in protein expression levels. Recent studies have shown that protein expression relies heavily on translation, specifically during initiation. In one study, changes in nucleotides around the start codon downstream of a given promoter altered protein expression levels by a factor of 250 (De Smit and Van Duin, 1990). One mechanism underlying the sensitivity of expression is that the efficiency with which RBSs recruit ribosomes for translation initiation depends on several factors, including the Shine–Dalgarno (SD) sequence and the secondary structures of nucleotides within the RBS (Makrides, 1996).

The secondary structure of a transcript is context dependent, and variations in a few nucleotides within the RBS can modify the energetics of interactions with neighboring nucleotides. Interactions between the 16S rRNA in the ribosome and a transcript can also influence the interactions at the translation initiation site. The complexity of the sequence-dependent structural interactions has led many researchers to adopt *ad hoc* experimental methods to optimize each expression problem (Zhelyabovskaya *et al.*, 2004).

Recently, we constructed a mathematical translation initiation model that combines techniques for estimating translation efficiency (Na *et al.*, 2010b). The mathematical model includes the following events in the model for translation initiation: (i) global folding and unfolding of transcribed mRNAs; (ii) regional folding and unfolding of nucleotides in an RBS; and (iii) ribosome binding mediated by SD regions and 16S rRNA. The translation efficiency is proportional to the ratio of ribosome-bound mRNA to free mRNA.

A thermodynamic model similar to the one employed here has been reported, which estimates the translation efficiency (Salis *et al.*, 2009). A web-based tool (RBSCalculator) was developed as a user front-end for this model. The predictive value of this model and our model is similar because the same key factors are considered, but the incorporation of additional factors into our model distinguishes RBSDesigner from RBSCalculator. For example, RBSDesigner can predict the translation efficiency of mRNA sequences that may potentially fold into two or more structures.

In the present work, we developed a GUI tool that facilitates the design of synthetic RBS exhibiting a desired level of protein expression. This technique may assist advances in a variety of biological fields, including synthetic biology, protein biotechnology and metabolic engineering, in which protein expression control is of key importance.

2 IMPLEMENTATION

The mathematical routines were written in Python (http://www.python.org) and compiled as an executable using py2exe (http://www.py2exe.org). This program identifies a potential SD sequence and RBS in a user-specified sequence and extracts secondary structure information for the mRNA as predicted by *UNAfold* v3.3 (Markham and Zuker, 2008). The effect of spacer length between the SD and AUG sequences on translation efficiency was considered because this spacer length influences the efficiency with which fMet-tRNAs are incorporated into the ribosome during

*To whom correspondence should be addressed.
†Present address: Department of Chemical and Biomolecular Engineering (BK21 Program), BioProcess Engineering Research Center, Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, KAIST, Daejeon 305-701, Republic of Korea.

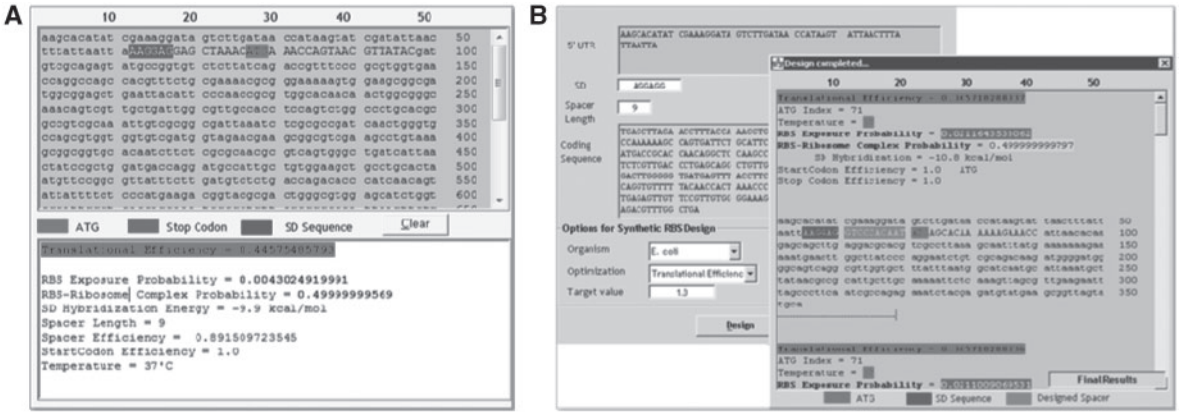

Fig. 1. Two key functionalities of RBSDesigner. (**A**) Estimation of the efficiency of translating existing mRNA: estimated properties of a user-specified mRNA sequence, such as a SD sequence, the translation efficiency, spacer effect, ribosome binding affinity, etc., are displayed. (**B**) Synthetic RBS design: candidate designed synthetic RBS sequences that exhibit the desired translation efficiency are displayed with their estimated properties.

initiation of translation elongation (Ringquist *et al*., 2006). Several sources of information were used to parameterize the translation efficiency. A GUI was constructed using Microsoft Visual Basic 6.0 to mediate the interaction of a user with the numerical expressions of the program. Thus, RBSDesigner provides a convenient way for biologists to design synthetic RBS for the control of protein expression.

3 KEY FEATURES

3.1 Translation efficiency estimation

RBSDesigner estimates the translation efficiency of a user-specified mRNA sequence. The GUI displays the predicted properties, such as the identity of a potential SD sequence, the SD–ribosome binding energy, the identity of a potential RBS sequence, the probability that the structure of the RBS is unfolded and the effect of spacer length, among other properties (Fig. 1A).

3.2 Synthetic RBS design

RBSDesigner designs synthetic RBS. More specifically, it can design the nucleotide sequence of the spacer between SD and AUG sequences such that a desired level of protein expression is achieved, as predicted by the genetic algorithm (Fig. 1B).

4 CONCLUSIONS

The availability of convenient software that facilitates the design of synthetic RBS for controlling protein expression is critical for the advancement of synthetic biology, protein biotechnology and metabolic engineering. RBSDesigner provides an intuitive and easy way to design synthetic RBS. The synthetic RBS could be used to maximize translation efficiency for the intensive production of therapeutics or industrial proteins. It may assist in balancing the levels of enzymes in a synthetic metabolic pathway to optimize the metabolic flux, or it could be used to optimize protein expression of components in constructed genetic circuits to ensure robust operation.

ACKNOWLEDGEMENTS

We would like to thank Sunjae Lee for his assistance in testing RBSDesigner and preparing this manuscript.

Funding: National Research Foundation of Korea (NRF) grants funded by the Korea Government, Ministry of Education, Science and Technology (MEST) through the BRL (Basic Research Laboratory) grant (2009-0086964); WCU (World Class University) program (R32-2009-000-10218-0).

Conflict of Interest: none declared.

REFERENCES

De Smit,M. and Van Duin,J. (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl Acad. Sci. USA*, **87**, 7668.

Makrides,S. (1996) Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol. Mol. Biol. Rev.*, **60**, 512.

Markham,N. and Zuker,M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.*, **453**, 3.

Na,D. *et al.* (2010a) Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.*, **13**, 1–8.

Na,D. *et al.* (2010b) Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with a desired expression level in prokaryotes. *BMC Syst. Biol.*, **4**, 71.

Ringquist,S. *et al.* (2006) Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Mol. Microbiol.*, **6**, 1219–1229.

Salis,H.M. *et al.* (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.

Zhelyabovskaya,O. *et al.* (2004) Artificial genetic selection for an efficient translation initiation site for expression of human RACK1 gene in Escherichia coli. *Nucleic Acids Res.*, **32**, e52.