The Synthetic Gene Designer: a Flexible Web Platform to Explore Sequence Space of Synthetic Genes for Heterologous Expression

Gang Wu, Nabila Bashir-Bello, Stephen Freeland
Department of Biological Sciences, University of Maryland at Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250
freeland@umbc.edu

Abstract

'Codon optimization' is a general approach that seeks to improve heterologous expression when a gene is moved into a foreign genome that exhibits different codon usage from its native genome. However, it is still unclear exactly what it means to optimize codons for protein expression: diverse interpretations exist, and exploratory research continues. Thus, to facilitate flexible redesign of genes for heterologous expression, we have developed a bona fide web application: the "Synthetic Gene Designer". Compared with existing codon optimization software, the Synthetic Gene Designer provides greater flexibility in several key features of codon optimization strategy. The software, together with online tutorials and help, is available at http://www.evolvingcode.net/codon/sgd.php.

1. Introduction

Codon optimization has been widely applied to improve heterologous protein expression in the past two decades [1]. Some computer programs have been developed recently to automate the codon optimization [2][3][4]. However, there is no clear consensus on: 1) what is an appropriate reference template of codon usage; 2) what is the best algorithm for codon optimization and 3) the extent to which 'optimal' codons should be used. Thus existing software that uses a fixed model for codon optimization can only partially address current needs. To facilitate systematic and coordinated investigation of these questions, we have developed a web application: the "Synthetic Gene Designer" which facilitates flexible redesign of genes for heterologous expression.

2. Unique features of this Software

Compared with existing codon optimization software, the Synthetic Gene Designer provides <u>greater flexibility</u> in several key features of codon optimization

strategy. 1) Our software allows users to work with any of the non-standard genetic codes recognized by the NCBI. 2) By incorporating the CUTG database, our software offers a specific codon usage reference template for every organism in Genbank (7,434 organisms in total, [5]). We also provide codon usage reference templates based on highly expressed genes according to published articles for 11 organisms. Moreover, users can compile customized codon usage reference templates by submitting their own gene sequences into our program. 3) The Synthetic Gene Designer also provides flexible methods of codon adjustment. By using a scaling factor ("optimality factor"), the codon usage bias of adjusted DNA sequences can be driven to different levels of either "optimization" or "counter-optimization". Also, unwanted patterns such as high G/C repeats and restriction sites in gene sequence can be avoided automatically. Finally, users can manually edit the final sequence, codon by codon, to meet any specific considerations.

In addition to this flexibility, the Synthetic Gene Designer is the first *bona fide* web-based software dedicated to codon adjustment for heterologous expression. All other existing applications, including "DNAWorks" [2], "Codon Optimizer" [3] and "UpGene" [4] are limited to stand-alone application operating on a Microsoft Windows platform.

3. Overview of the software

The Synthetic Gene Designer is written in PHP, JavaScript and Perl. Given a gene of interest and a target genome in which it is to be expressed, the Synthetic Gene Designer comprises three major steps for gene design. First the user specifies which genetic code it is to be expressed under (Fig.1A). Next the user provides a codon usage reference template upon which the new sequence is to be modeled (Fig.1B). Then the software generates a fully-optimized nucleotide
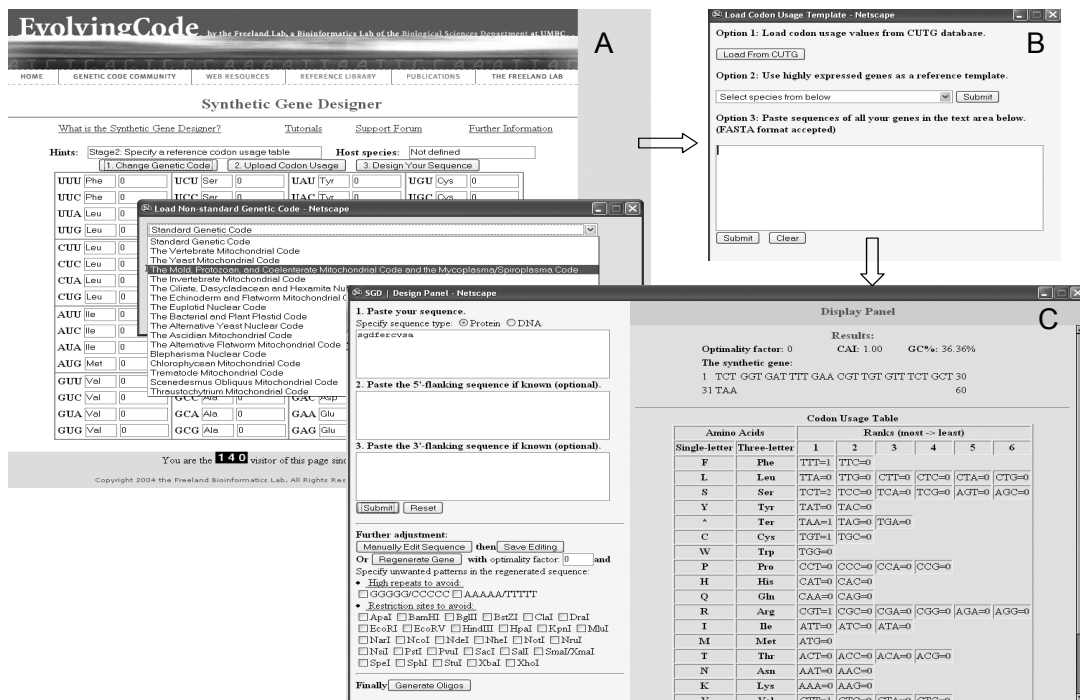
Fig. 1: Screenshots of the Synthetic Gene Designer. The software operates in 3 key steps. (A) The user specifies a genetic code for the heterologous system; (B) The user creates a template of "preferred codon usage" for the heterologous system (this may be customized, loaded from the CUTG database or from one of our 11 templates takes from the scientific literature); (C) The user inputs a coding sequence for the software to re-design. Full online tutorial is available to train new users in approx 20 minutes.

sequence according to this template (Fig.1C). The user may then further adjust codon usage, either by manual editing or by selecting a degree of (counter) optimization and other parameters (such as motifs to avoid) to automatically achieve a new sequence. Finally the user can generate a series of oligonucleotide sequences of desired lengths for assembly PCR (Detailed method descriptions are at http://www.evolvingcode.net/codon/sgd/methods.php).

4. Summary

Synthetic gene design is a technique of growing importance within modern molecular biology as the technology of synthetic gene construction becomes cheaper and more reliable. Yet the algorithmic basis for optimizing codon usage remains far from clearly understood. In this context, we present new software that couples the simplicity of platform independent web-interface with the flexibility to explore different models for codon optimization. By setting bench workers free from the constraints of "fixed optimization algorithms", this software will stimulate progress in our understanding of what exactly constitutes codon optimization.

5. References

[1] C. Gustafsson, S. Govindarajan, and J. Minshull, "Codon bias and heterologous protein expression," *Trends Biotechnol*, vol. 22, pp. 346-53, 2004.
[2] D. Hoover and J. Lubkowski, "DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis," *Nucleic Acids Res*, vol. 30, pp. e43, 2002.
[3] A. Fuglsang, "Codon optimizer: a freeware tool for codon optimization," *Protein Expr Purif*, vol. 31, pp. 247-9, 2003.
[4] W. Gao, A. Rzewski, H. Sun, P. D. Robbins, and A. Gambotto, "UpGene: Application of a web-based DNA codon optimization algorithm," *Biotechnol Prog*, vol. 20, pp. 443-8, 2004.
[5] Y. Nakamura, T. Gojobori, and T. Ikemura, "Codon usage tabulated from the international DNA sequence databases; its status 1999," *Nucleic Acids Res*, vol. 27, pp. 292, 1999.