Looking for the brightest one: fluorescent protein-based approach for identifying optimal coding sequence for recombinant protein expression in E.coli.

P-02.1-21

A. Zabolotskii[I]

[I]Lomonosov Moscow State University, Moscow, Russia

Due to the degeneracy of the genetic code, most amino acids are encoded by several codons. Different synonymous codons at the 5'-end coding sequence of mRNA (5CDS) has strong effect on protein expression. This often explained by different contribution of synonymous codons to the minimal free energy (MFE) of the 5CDS which correlates with probability of mRNA secondary structure formation. Strong secondary structure in this region interferes with ribosome binding and affects the process of translation initiation. *In silico* optimization of 5CDS can significantly increase the level of protein expression. However, this method is not always effective due to the uncertainty of the exact mechanism by which synonymous substitutions affect expression, thus it may produce numerous of non-optimal variants. An alternative approach is the generation of partially-randomized library comprising hundreds of selected synonymous variants fused to reporter gene with subsequent screening for most promising candidates according to reporter's signal intensity. For this work as a model protein, we used canine cystatin C (CCC) which is known for low expression level in *E.coli*. Thus, a library of 5CDS-partially randomized CCC fusions to superfolderGFP (CCC::sfGFP) was created. Colonies with the highest expression level of CCC::sfGFP were selected based on fluorescence intensity. Then sfGFP coding sequence was removed from the fusion and expression level of obtained non-fused CCC variants was measured. As a result, several optimized CCC sequences with an expression level exceeding the original version by ~20 times were obtained. Analysis of 5CDS from several optimal and non-optimal variants revealed that variants with highest values of MFE were the ones with the highest CCC expression levels. We suggest that this simple approach may provide efficient and inexpensive optimization method for poorly expressed proteins in prokaryotic system.