Chapter 4

Codon Optimizing for Increased Membrane Protein Production: A Minimalist Approach

Kiavash Mirzadeh, Stephen Toddo, Morten H.H. Nørholm, and Daniel O. Daley

Abstract

Reengineering a gene with synonymous codons is a popular approach for increasing production levels of recombinant proteins. Here we present a minimalist alternative to this method, which samples synonymous codons only at the second and third positions rather than the entire coding sequence. As demonstrated with two membrane-embedded transporters in Escherichia coli, the method was more effective than optimizing the entire coding sequence. The method we present is PCR based and requires three simple steps: (1) the design of two PCR primers, one of which is degenerate; (2) the amplification of a mini-library by PCR; and (3) screening for high-expressing clones.

Key words Membrane protein, Protein expression, Codon optimization, Synonymous codon

1 Introduction

In recent years, codon optimization has emerged as a popular and often necessary approach for boosting production levels of recombinant proteins. In a nutshell it involves reengineering the gene or coding sequence (CDS) of interest with optimal synonymous codons. The codons that are deemed optimal are those whose corresponding tRNA concentration in the host cell is highest, as this is thought to favor fast decoupling by the ribosome and more efficient translation [1]. As the cellular tRNA concentrations have coevolved to mirror the frequency of usage of the corresponding codons in the genome [2, 3], this design principle may seem straightforward. However, when choosing synonymous codons, care must be taken to avoid nucleotide stretches that resemble ribosome-binding sites (RBSs), stretches that form strong mRNA structures, RNase sites, DNA recombination sites, or transcriptional terminators, as they can be detrimental for protein production [1]. The codon optimization approach is conceptually...
attractive and has gained traction, as it is increasingly cheaper to order codon-optimized CDSs from commercial vendors.

Does codon optimization increase production levels of membrane proteins? The short answer is “yes,” there are numerous examples of successful optimizations in the literature ([reviewed in [4]]). However unsuccessful optimizations are unlikely to be published, so it is hard to estimate success rates. Our own experiences as well as feedback from colleagues indicate that codon optimization of the entire CDS is often not effective. For example, when we ordered codon-optimized CDSs from two different commercial sources we did not observe significant production of two *E. coli* transporters, called AraH and NarK [5]. As an alternative we explored a minimalist codon optimization approach that samples synonymous codons immediately adjacent to the AUG start codon (i.e. +2, +3). This approach was inspired by the work of Isaksson and co-workers, who systematically sampled all 61 codons in the +2 position of an artificial gene and noted that a single synonymous codon change could affect expression levels by as much as 20-fold (note that a weak Shine Dalgarno sequence was used in these experiments) [6]. When we systematically tested synonymous codons in the +2, +3, and +4 positions of the native *araH* and *narK* CDSs we observed significant increases in production levels [5, 7]. Thus, in these two cases, single synonymous codon substitutions immediately adjacent to the AUG start codon were far more effective at increasing production than codon optimization of the entire coding sequence.

A molecular explanation for why codons adjacent to the AUG start codon influence protein production levels remains elusive. However it most likely relates to translation initiation since it is known that the 16S rRNA of the 30S ribosomal subunit must recognize the nucleotide sequence around the AUG start codon during translation initiation. Minor nucleotide changes in this region, which extends from the Shine Dalgarno sequence to the +4/+5 codons [8, 9], can decrease the efficiency of translation initiation by causing strong mRNA hairpins [10, 11]. In nature, mRNA structure around the AUG start codon has been selected against [12, 13], but in a recombinant protein production experiment this region is a composite of the vector and the 5′ end of the CDS. Thus the optimal synonymous codons for reducing mRNA structure will depend on the sequence context and will differ from CDS to CDS, and vector to vector. This fact underscores the need for sampling different combinations of synonymous codons and testing their expression level.

In this chapter we present our protocol for generating small clone libraries with all possible combinations of synonymous codons in the +2 and +3 positions. Depending on which amino acids are present at these positions of the CDS, the clone libraries can contain up to 36 different variations. Thus, a limited amount
of screening is required to identify the clones that express to the highest level. An overview of the method, which we call post-cloning optimization (PCO), is presented in Fig. 1a. It involves three simple steps: (1) the design of two overlapping primers (one of which is degenerate) for amplifying the original expression plasmid, (2) the amplification of a mini-library by PCR and then re-circularization of the plasmids by transformation into E. coli, and (3) screening for high-expressing clones. In all of our experiments, the CDS is fused to a region encoding for a -TEV-GFP-His\(_8\) tag (Fig. 1b) so that whole-cell fluorescence can be used as a proxy for protein production [14]. Protocols on how to use GFP in this manner have been presented elsewhere [15]. When we carried out PCO on the araH and narK CDSs, the mini-libraries contained 6 and 12 clone variations, respectively. And by simply screening 24 colonies we were able to identify clones with considerably increased expression. For example, araH increased by 9-fold from 1 mg/mL to 9 mg/mL (Fig. 1c), and narK by 17-fold from 1.6 mg/mL to 29 mg/mL (Fig. 1d).

The method is a minimalist approach to codon optimization that is inexpensive and simple enough to be carried out in any laboratory that has access to a PCR block. It could be implemented during cloning or as a post-cloning step as we have demonstrated here. Moreover it has a major advantage over codon optimization of the entire CDS; it does not affect the efficiency of elongation, so membrane protein folding should not be perturbed (note that there are reports that suggest elongation rate is linked to the folding of membrane proteins [16, 17]). We foresee that the method could be useful for boosting production of recombinant proteins in E. coli (both membrane and soluble). It might also be useful for tuning expression levels of genomically encoded proteins by using (randomized) oligonucleotide-based recombineering [18].

## 2 Materials

### 2.1 Components for PCR

1. Primer set for PCR amplification (see Note 1 for design principles).
2. Q5 DNA polymerase and 5x Q5 reaction buffer (New England Biolabs) (see Note 2).
3. Nucleotide triphosphates: Stock solution containing 100 mM of dATP, dTTP, dCTP, and dGTP.
4. DpnI restriction enzyme (New England Biolabs) (see Note 3).
5. Sterile H\(_2\)O.
6. 0.2 mL Soft-walled PCR tubes.
7. Thermocycler.
8. Agarose powder.
a

Step 1: Primer design
Design primers for amplification of original plasmid. Forward primer should be degenerate to allow synonymous codon changes at triplets +2 and +3. Reverse primer should overlap by approximately 15 base pairs with the forward primer to facilitate re-circularization.

b

T7 promoter

\[ \text{coding sequence} \quad \text{tev} \quad \text{gfp} \quad \text{his} \quad \text{pET28a} \]

\[ \text{XhoI} \quad \text{KpnI} \]

c

Fig. 1 A minimalistic approach to codon optimization. (a) A schematic overview of the method. (b) CDSs used in this study were cloned into a modified version of the pET28a vector and genetically fused to a region encoding a -TEV-GFP-His\textsubscript{8} tag \[14\]. This enabled us to use GFP fluorescence as a measure of expression. (c) and (d) Screening of expression levels from plasmids harboring ara\textit{H} and nar\textit{K}, where all possible combinations of synonymous codons in the +2 and +3 positions have been generated. Expression was carried out in BL21(DE3) pLysS by induction with 1.0 mM IPTG for 5 h at 25 °C. To estimate the amount of protein produced in mg/L, the whole-cell fluorescence was compared to a standard curve obtained with purified GFP. For comparison, the expression levels of the original clone and two whole-gene codon-optimized versions are indicated to the right. Parts a and b adapted from \[7\]

10. TAE buffer: Prepare 50x stock solution by mixing 242 g Tris base in 600 mL of H$_2$O. Add 57 mL of glacial acetic acid and 100 mL of 0.5 M EDTA pH 8.0, and bring the final volume to 1 l with H$_2$O. Store at room temperature.

### 2.2 Component Transformation and Re-circularization of Libraries

1. Chemically competent *E. coli* cells (see Note 4).
2. 20 g/L Luria Bertani (LB) broth in H$_2$O.
3. LB agar plates: LB broth with 15 g/L agar.
4. 1.5 mL Microfuge tubes.
5. Thermomixer for incubating microfuge tubes.
6. 50 mL Reaction tube.
7. 10 mL of LB broth supplemented with appropriate antibiotics.
8. Incubator for 50 mL reaction tube.
9. ENZA DNA mini kit (Omega bio-tek)

### 3 Methods

#### 3.1 Primer Design

1. The forward primer(s) should straddle the ATG start and should be approximately 40–50 nucleotides long (Fig. 2a). The six nucleotides downstream of the AUG start codon need to be degenerate so that different synonymous codons will be sampled. A table indicating the code used to implement degeneracy is shown in Fig. 2b. Note that in some cases it was not possible to design a single forward primer, so multiple forward primers were used and PCR products were mixed.

2. The reverse primer should match to the region upstream of the ATG start and should also be approximately 40–50 nucleotides in length (see Fig. 2a). The 5′ end of the reverse primer should match the 5′ end of the forward primer (overlapping by approximately 15 nucleotides), so that the PCR products can circularize by homologous recombination when transformed into *E. coli*.

#### 3.2 PCR Amplification of Mini-Libraries

1. Mix all reagents for PCR in a 1.5 mL microfuge tube: 71 μL of H$_2$O, 20 μL of 5× Q5 reaction buffer, 1 μL of Q5 DNA polymerase (2 U/μL), 2 μL of 50 mM dNTP mix, 1 μL of forward primer (50 pmol / μL), 1 μL of reverse primer (50 pmol / μL), 4 μL of original plasmid (4 ng/μL) (see Note 5).

2. Aliquot 20 μL of the PCR mix into five separate 0.2 mL PCR tubes.
3. Amplify the mini-library in a thermocycler using a program that consists of 95 °C for 2 min, then 30 cycles of 95 °C for 45 s, a range of temperatures from 48 to 68 °C for 45 s, and 72 °C for 6.5 min. Finish with a final elongation step at 68 °C (see Notes 6 and 7).

4. Add 10 units of DpnI to the reaction mix to digest the original plasmid.

5. To ensure that the mini-library was amplified, analyze 1 μL of the PCR product by agarose gel electrophoresis using standard protocols.

3.3 Transformation of Libraries into E. coli to Facilitate Re-circularization

1. Mix 10 μL of the PCR mix (approximately 500 ng) with 100 μL of competent MC1061 E. coli cells in a 1.5 mL reaction tube (see Note 8).

2. Incubate on ice for 30 min.

3. Heat shock for 1 min at 42 °C.

4. Incubate on ice again for 2 min.

5. To allow cells to recover, add 0.5 mL of LB broth and incubate at 37 °C with shaking for one hour.

6. Transfer cells to a 50 mL reaction tube containing 10 mL of LB broth and appropriate antibiotics, and incubate at 37 °C with shaking for 16 h (see Note 9).

7. Harvest the cells by centrifugation at 4000 × g in a bench-top centrifuge.

8. Purify the plasmids using the ENZA DNA mini kit as per the manufacturer’s instructions. This plasmid prep is a mini-library containing variants of your original plasmid, which differ only in...
the use of synonymous codons in the +2 and +3 positions, as determined by the design of your degenerate forward primer.

9. The mini-library can then be directly transformed into an expression strain such as BL21(DE3) or a derivative that has been selected or engineered for high-level production [19, 20]. To do so, take 1 μL of the mini-library, follow steps 1–5 above, and then plate out 200 μL of the culture on an LB agar plate with appropriate antibiotics.

10. Incubate at 37 °C for 16 h, and then pick colonies for expression testing (see Notes 10 and 11).

### Notes

1. The protocol requires at least one forward degenerate primer. However if codons with more than four synonymous variants are being tested, then an additional primer will be required. The protocol always requires one reverse primer.

2. While any high-fidelity polymerase can be used, we choose to use the Q5 polymerase because its error rate is so low that it is difficult to measure in a statistically significant manner (see manufacturer’s specifications). This minimizes random errors on the vector backbone and in the CDS.

3. *DpnI* comes with its own reaction buffer, which we use to dilute it to a concentration of 10 U/μL. This can then be added directly to the PCR when required, since *DpnI* also works effectively in the reaction buffer supplied for the Q5 polymerase.

4. Any *E. coli* strain that is capable of homologous recombination will suffice. We use MC1061.

5. Typically, we prepare a PCR mix of 100 μL, and then we aliquot 20 μL into five separate 0.2 mL PCR tubes. This allows us to test different annealing temperatures simultaneously by incubating the PCR mix in a thermocycler that has a gradient function.

6. To obtain maximal diversity in the library, we suggest that you test a range of annealing temperatures during the PCR and then choose the reaction that was amplified at the lowest annealing temperature.

7. Note that the extension time will depend on the size of your plasmid and the processivity of the DNA polymerase that you use. We typically allow 6.5 min for a 6.5 kb plasmid (i.e., 1 min/kb).

8. If two forward primers were required for the PCR, then mix 5 μL from each reaction.
9. This step allows propagation of the library.
10. Typically we would expect >500 colonies per plate.
11. Typically we would compare expression between 24 colonies. The type of expression testing done depends on the detection systems available.

Acknowledgments

This work was supported by a grant from the Swedish Research Council to DOD, and by the Novo Nordisk Foundation to MHHN.

Conflict of interest statement: The method considered here has been described in a patent submitted by CloneOpt AB (SE1451553-0). KM, ST, and DOD are founders and shareholders in CloneOpt AB.

References

protein overexpression and purification using GFP fusions. Nat Methods 3:303–313