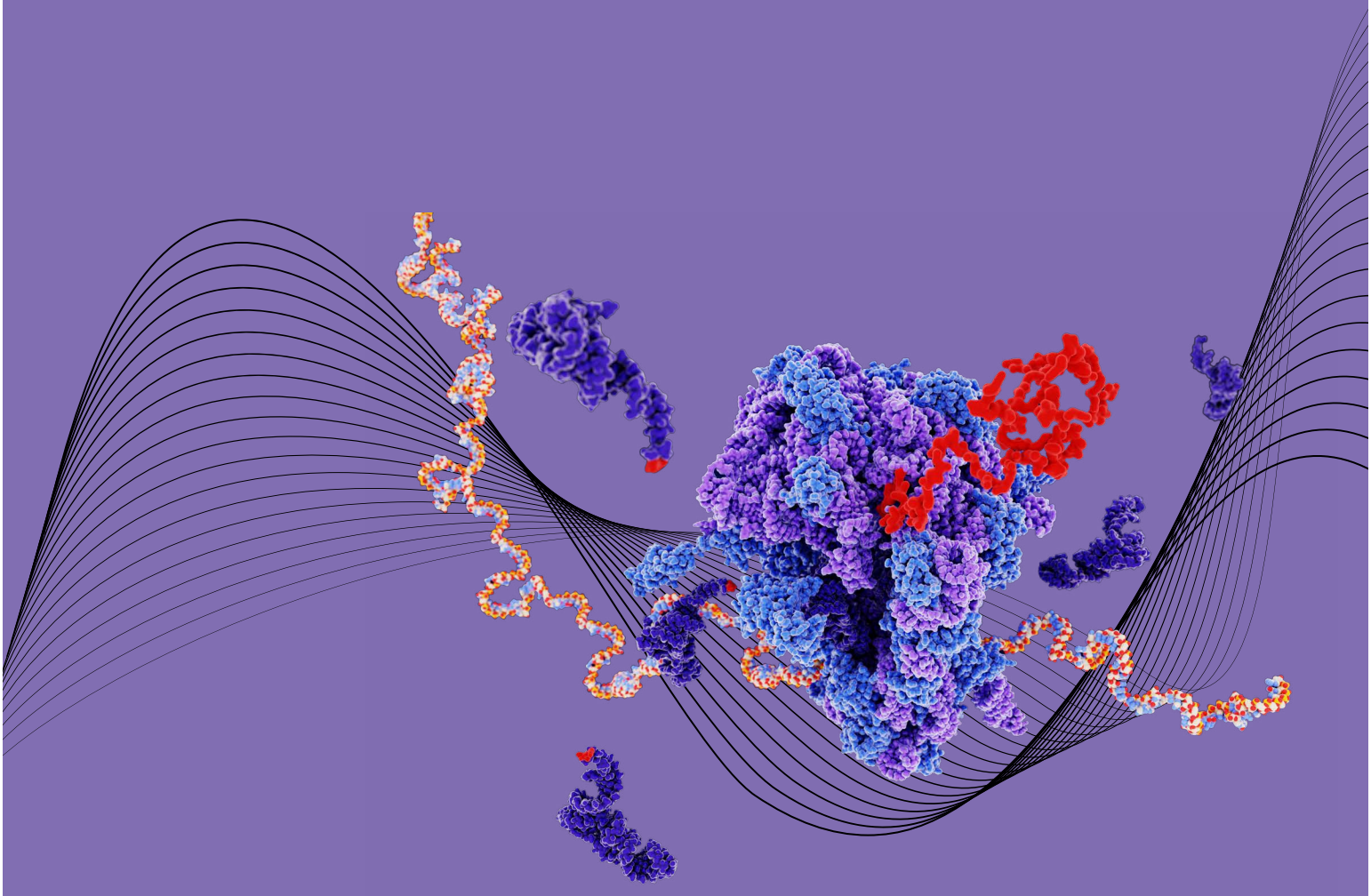
Genomics & Analytical Services

TECH NOTE

Enhancing Protein Expression by Leveraging Codon Optimization



AZENTA
LIFE SCIENCES

azenta.com

Introduction

Codon preference is among the most critical parameters affecting the expression level of foreign genes. A codon is a sequence of three nucleotides which together form a unit of genetic code in DNA or RNA that encodes a specific amino acid. The genetic code represented by a codon describes the relationship between the sequence of DNA bases in a gene and the corresponding protein sequence that it encodes. Because there are many more codons (61 codons encoding for amino acids and 3 stop codons) than there are translated amino acids (20 different translated amino acids), individual amino acids are encoded by up to six different codons but within codon families, some are used more frequently than others. For this reason, optimizing codon selection is a particularly important parameter in improving translational efficiency and gene expression. This technical note will discuss how researchers can leverage codon optimization tools in the selection of the most appropriate codon to maximize expression of target proteins within host organisms.

The Challenge

Optimizing a gene sequence so that it can express the target protein in large quantities is a common challenge experienced in gene engineering. Codon optimization is an approach that can be used to help maximize gene expression by changing similar codons based on an organism's specific codon bias.

Through codon optimization, scientists can introduce mutations throughout a gene of interest based on an organism's codon usage bias. The approach is exceptionally effective in helping to increase translational efficiency and protein expression without changing the sequence of the protein itself. Figure 1 is a visual representation of the protein expression mechanism within a given organism. Identification of the ideal codon sequence for every amino acid across an entire polypeptide, however, is a tedious process for scientists to undertake that can significantly delay research efforts.

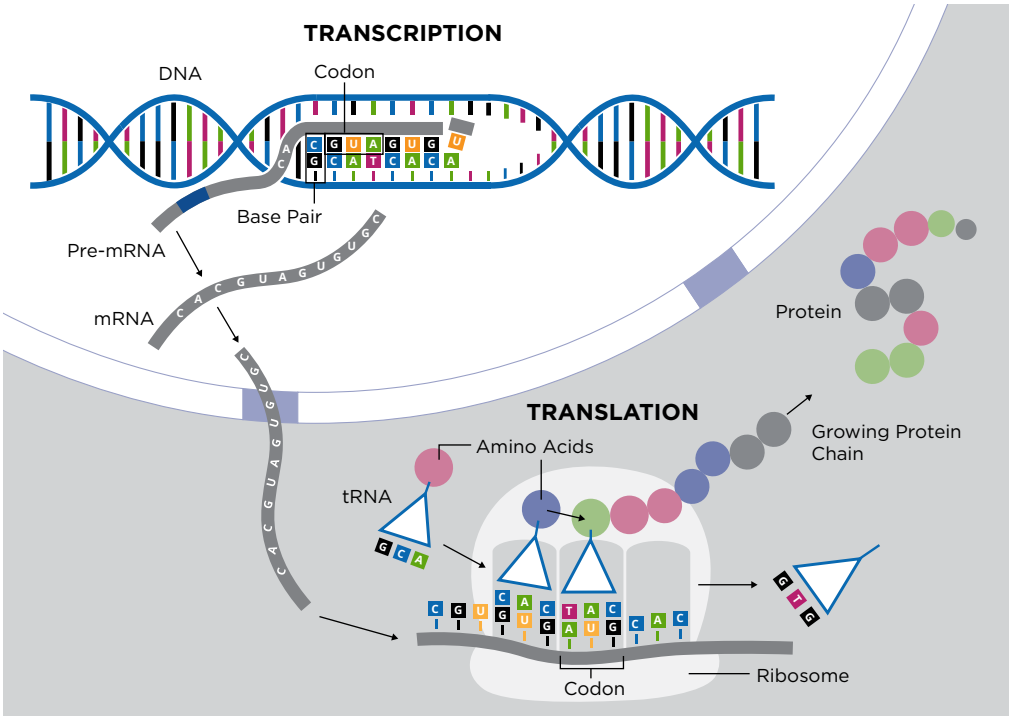


Figure 1: Schematic representation of protein expression

The Solution

The introduction of optimization algorithms, a type of artificial intelligence, has significantly streamlined the process of designing a DNA sequence to optimize protein expression in target organisms. In addition, researchers also have access to a multitude of services delivering codon-optimized genes for optimal expression of target proteins. Since 2010, Azenta Life Sciences has been supporting researchers' genetic engineering success with its free codon optimization algorithm tool that optimizes key parameters to stabilize DNA sequences and improve gene expression. The tool's algorithm is frequently improved using empirical data that addresses two critical parameters including codon usage frequency and mRNA secondary structure.

Codon Usage Frequency

Codon usage bias is recognized as a critical factor in gene expression and cellular functions that range from RNA processing to protein translation and protein folding. The process of translation can be seen in Figure 2 below. During the translation process, messenger RNA (mRNA) is decoded in a ribosome that is outside of the nucleus, to produce a specific amino acid chain or polypeptide. The polypeptide later folds into an active protein and performs its functions within the cell. Low-frequency codon usage has been cited as an obstacle to robust gene expression. By replacing low-frequency codons, Azenta's codon optimization algorithm allows researchers to effortlessly optimize gene expression within host organisms.

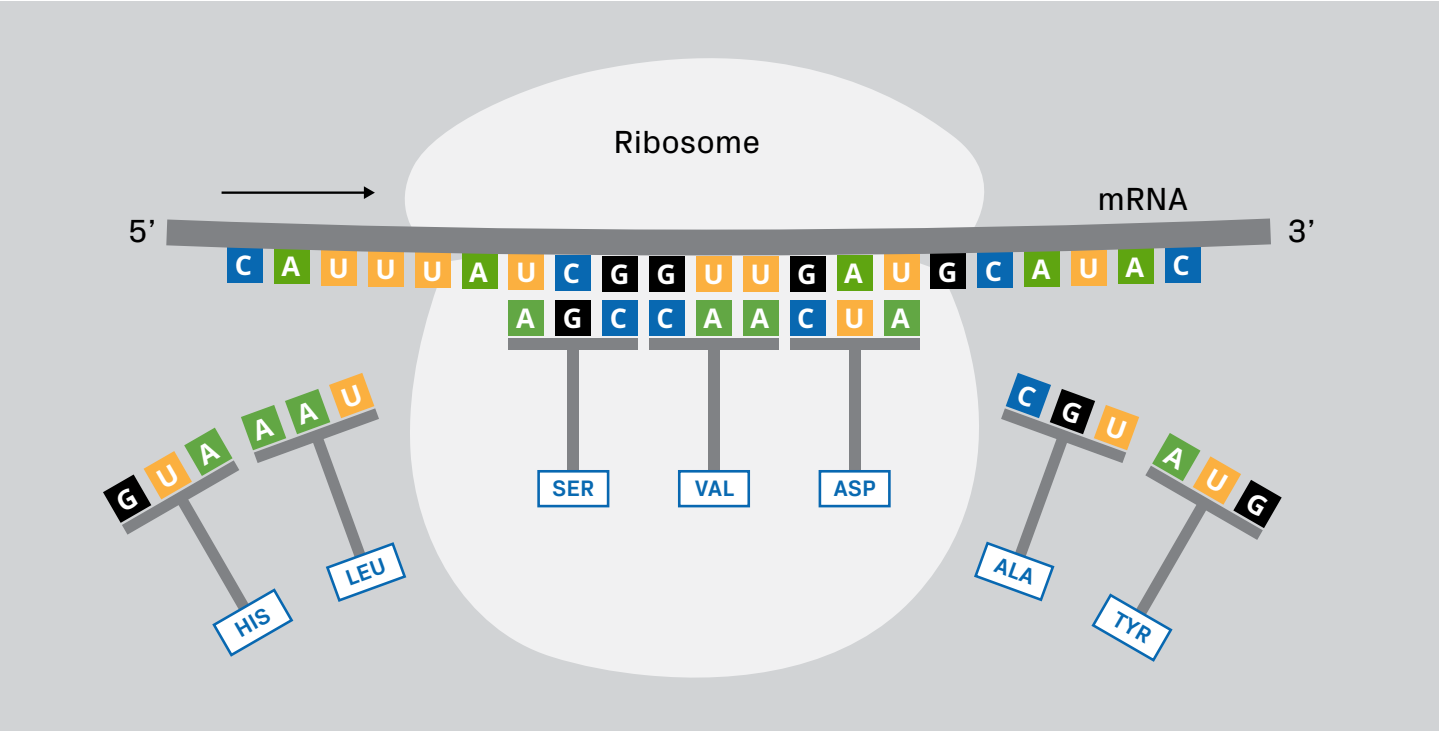


Figure 2: Process of Translation

mRNA Secondary Structure

Secondary structure of messenger RNA plays an important role in the biosynthesis of proteins. Its negative impact on translation can reduce the yield of protein by slowing or blocking the initiation and movement of ribosomes along the mRNA, becoming a major factor in the regulation of gene expression. Initiation can occur when the structured element is positioned between the Shine–Dalgarno sequence (SD) and the start codon (AUG). Despite the complicated nature of mRNA secondary structure predictions, Azenta's codon optimization algorithm avoids the appearance of inverted repeats in its systematic codon optimization. Figure 3 below illustrates how the secondary structure of mRNA can inhibit the initiation of ribosomes in specific situations.

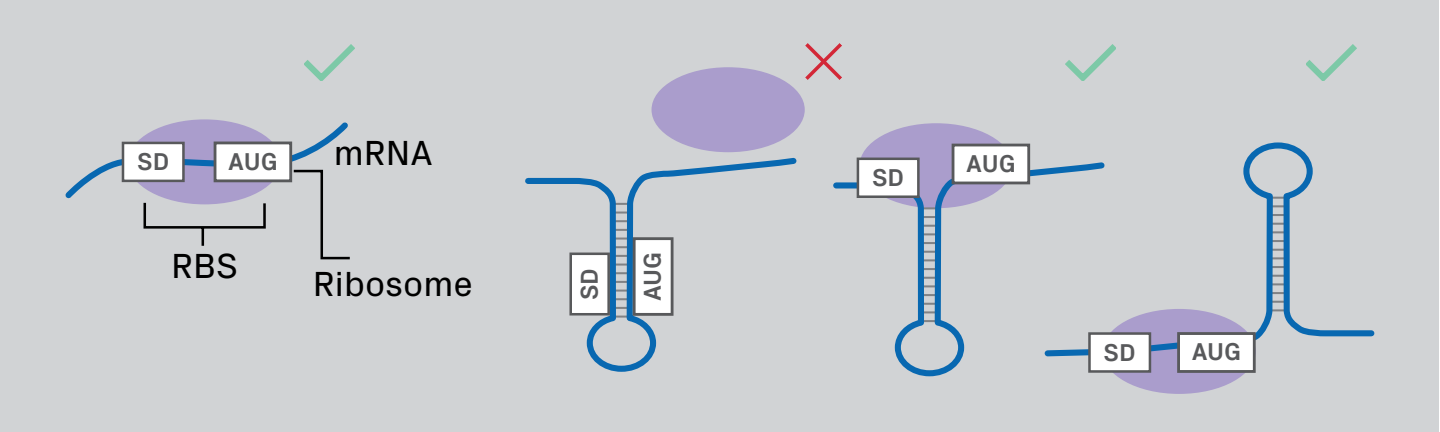


Figure 3: Secondary structure of mRNA inhibition of ribosome initiation

CpG Dinucleotide Content

CpG dinucleotides can have an impact on transcription efficiency. The Azenta codon optimization algorithm is designed to take this into account and help mRNA achieve the highest transcription efficiency. In addition to transcription efficiency, CpG dinucleotide content plays an important role in adeno-associated virus (AAV) vectors. CpG-depleted AAV vectors improve the safety and efficacy of AAV gene transfer in humans, which emphasizes the need to maximize transcription efficiency.

Other Advantages of Codon Optimization

- Guanine-cytosine (GC) nucleotide content modifies the GC content to speed up the gene synthesis process and shorten the turnaround time.
- Repeat sequences (direct repeat, inverted repeat, and dyad repeat) eliminate unnecessary repeat sequence which would be the hurdle during gene assembly.
- Restriction enzyme recognition sites remove useless sites and keeps the useful ones to ensure the precision of cloning and shorten the turnaround time.
- Customers can provide the codon usage frequency they prefer, and Azenta offers customized codon optimizations.

Application of Codon Optimization

The following case studies highlight the benefits of applying a codon-optimized approach to specific protein expression within a variety of genes. Each case study provides a visual illustration of the results obtained with and without codon optimization.

Case Study 1 – Expression of HSD17B4, DNA pol and hRad51 Genes

A comparative expression analysis of wild-type (WT) and Azenta-optimized gene sequences transformed into E. coli BL21(DE3) was performed to evaluate the protein expressions of each. Three different protein coding genes (HSD17B4, DNA pol, and hRad51) were selected for codon optimization (Figure 4). Protein expression level for WT and Azenta-optimized sequences are shown in adjacent lanes. For HSD13B4, in lane 3, the protein expression was barely visible in the wild-type DNA template. Lane 4, however, showed a significant increase of protein expression after applying codon optimization. We see a similar scenario in the study of DNA pol wherein lane 5, we noted only a slight protein expression from the wild-type DNA template with a much more prominent expression in lane 6. To further enhance the initial conclusion noted, a final study was conducted for the hRad51 gene. In this final study, we can conclude with certainty that applying Azenta's codon optimization (observed in lane 8) results in a much higher protein expression than the wild-type DNA template without codon optimization in lane 7.

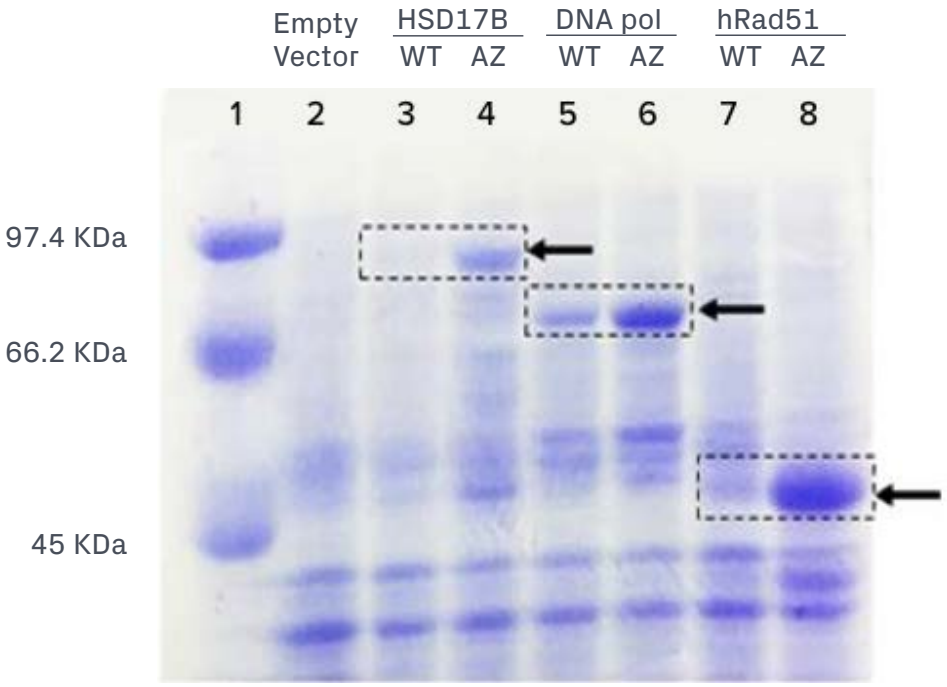
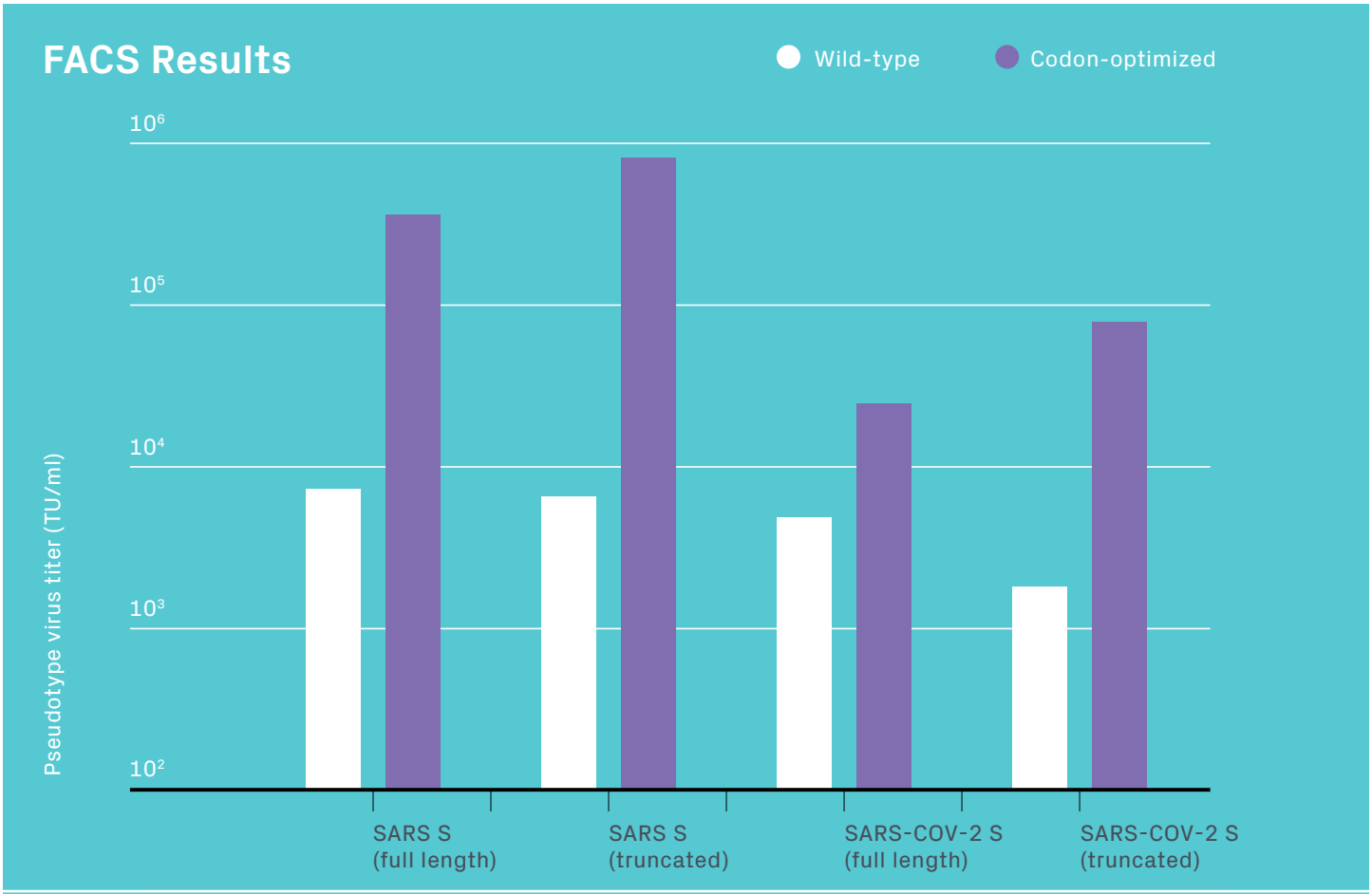


Figure 4: Comparison of protein expression between WT and Azenta-optimized sequences

Case Study 2 – Expression of SARS-COV-2 and SARS Pseudovirus

Most recently, Azenta's codon optimization algorithm was applied to the SARS-COV-2 and SARS pseudovirus in support of researchers' fight against COVID-19. The results captured in Table 1 below illustrate how application of Azenta's codon optimization algorithm significantly increased pseudovirus titer. The higher virus titer noted indicates the magnitude of increase in protein expression when the codon optimization algorithm is applied.

Figure 5 demonstrates the benefits of leveraging codon optimization using green fluorescent protein (GFP).



	Wild-type	Codon-optimized
SARS S (full length)	$7.3*10^3$	$3.7*10^5$
SARS S (truncated)	$6.5*10^3$	$8.3*10^5$
SARS-COV-2 S (full length)	$4.9*10^3$	$2.5*10^4$
SARS-COV-2 S (truncated)	$1.8*10^3$	$8.2*10^4$

Table 1: Virus titer of wild-type and optimized truncated and full-length SARS S and SARS-COV-2 S pseudovirus reflecting protein expression levels

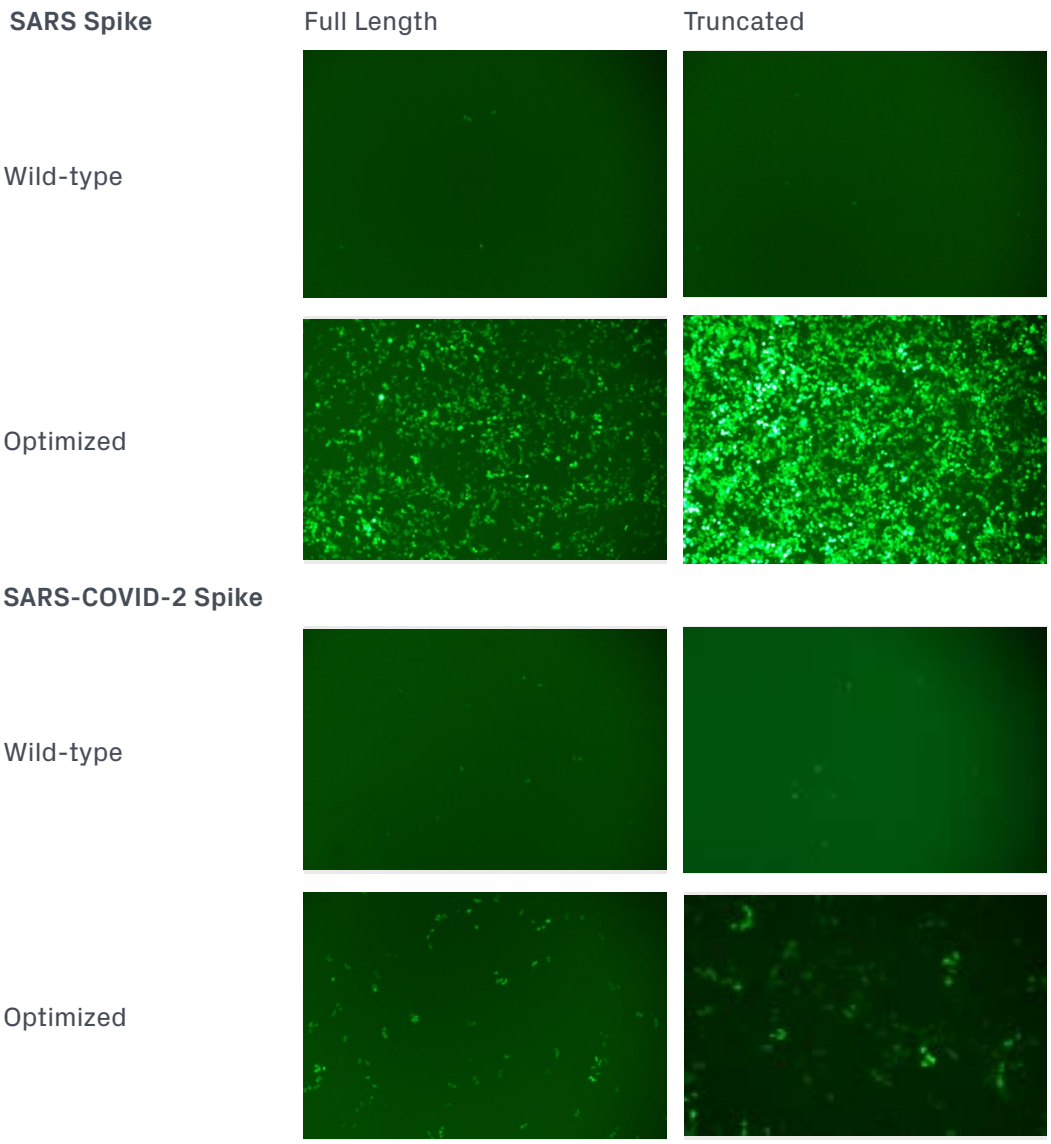


Figure 5: GFP signal of wild-type and optimized truncated and full-length SARS S and SARS-COV-2 S pseudovirus

Conclusion

Applying a codon-optimized approach in genetic engineering allows researchers to maximize gene expression by changing similar codons based on an organism's specific codon bias. This is particularly important for protein research projects where researchers cannot readily achieve high expression levels of a protein of interest.

For each protein highlighted within the case studies provided, leveraging the codon-optimized sequences generated by the codon optimization algorithm from Azenta Life Sciences resulted in significantly higher protein expression levels compared to wild-type sequences.

Gene expression and translational efficiency can be enhanced by DNA sequence optimization using multiple parameters including accommodation of codon bias of the expression host.

The Challenge	Azenta Solution
Low protein expression	Significantly improve protein expression levels by leveraging Azenta's codon optimization algorithm to analyze a wide range of data that optimizes critical parameters.
Low-frequency codon usage	Easily-identify and replace low-frequency codons with high frequency codons for all major host organisms using species-specific codon usage tables.
Difficult sequences	Normalize difficult sequences to remove unfavorable regions and reduce local GC content to desired percentages.

Referenced Publications

Plotkin, J.B. and Kudla, G. (2010). Synonymous but not the same: the causes and consequences of codon bias, Nature Reviews Genetics, 12, 32–42.

Fry L, Bastos R, Stone B, et al.(2019). Gene gun DNA immunization of cattle induces humoral and CD4 T-cell-mediated immune responses against the Theileria parva polymorphic immunodominant molecule. Vaccine, 37(12), pp.1546-1553. doi: 10.1016/j.vaccine.2019.02.009.

Li D, Fu G, Tu R, et al. (2019). High-efficiency expression and secretion of human FGF21 in Bacillus subtilis by intercalation of a mini-cistron cassette and combinatorial optimization of cell regulatory components. Microbial Cell Factories, 18(1). doi: 10.1186/s12934-019-1066-4.

Katoli P, Godbole A, Romanowski M, et al. (2018). Full-length myocilin protein is purified from mammalian cells as a dimer. Protein Expression and Purification, 147, pp.38-48. doi: 10.1016/j.pep.2018.02.008

You M, Yang Y, Zhong C, et al. (2018). Efficient mAb production in CHO cells with optimized signal peptide, codon, and UTR. Applied Microbiology and Biotechnology, 102(14), pp.5953-5964. doi: 10.1007/s00253-018-8986-5.

de Smit MH and van Duin J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proceedings of the National Academy of Sciences, 87:7668–7672.

Faust, Susan M., et al. "CpG-depleted adeno-associated virus vectors evade immune detection." The Journal of clinical investigation 123.7 (2013): 2994-3001.

AZENTA
LIFE SCIENCES

azenta.com