TWIST BIOSCIENCE

White Paper

Benchmarking Codon Optimization Strategies for Recombinant Antibody Expression

Authors: David Wang, Rachel Chang, Esteban Toro

May 2026

Abstract

Recombinant protein expression is a fundamental tool in modern molecular biology. By expressing engineered or non-native proteins in host cells (such as yeast or bacteria), scientists can transform the cells into biofactories capable of manufacturing therapeutics, industrial chemicals, and more. However, to do so, genes coding for these proteins must be carefully designed to ensure they can be robustly expressed in the host cells. One barrier includes species-specific codon biases, wherein one synonymous codon will be favored over another. Suboptimal codon usage can greatly diminish protein expression and undermine the viability of protein production.

Many tools for codon optimization follow a rules-based approach, such that individual codons are optimized in isolation. While useful, this approach ignores the higher-order effects that are produced by codon combinations. It is therefore likely that rules-based codon optimization fails to meet the full potential output for any given gene. Here, we set out to test whether an LLM-driven codon optimization approach—one that can account for combinatorial codon effects—is capable of outperforming legacy rules-based and machine learning approaches across a panel of 32 antibody constructs.

Introduction

Recombinant protein expression is a fundamental process in molecular biology, one with far-reaching effects on society. By expressing non-native proteins—either derived from other species or through protein engineering—in favorable host cells, scientists gain the ability to manufacture therapeutic antibodies, enzymes, and myriad proteins on an industrial scale.[1-4] However, the value of this process is predicated on the assumption that desired proteins will be reliably expressed in host cells.

Many factors affect protein expression, including the way amino acids are coded in DNA.[5] Synonymous codons have divergent, species-specific effects on protein expression; such that the same protein in different species may require a different combination of codons to achieve robust expression.[6, 7, 8] For this reason, optimizing codons to match the host's preferences can improve protein expression. Many optimization algorithms are designed to follow a rules-based approach, specifically using a codon adaptation index (CAI) framework.[9, 10] Such methods typically follow a set of rules to remove rare codons, balance GC content, and eliminate undesirable sequence motifs across the encoding sequence.

While rules-based approaches can improve expression, they implicitly treat codon optimization as a local, per-codon problem. Doing so ignores higher-order features that operate across multiple codons simultaneously. For example, some organisms use patterns of rare codons with relatively few corresponding tRNA binders. This can slow translational dynamics, conferring a net positive effect on protein folding and expression.[7, 8, 11-14] Additionally, codon pair bias, local RNA secondary structure, and the nucleotide composition of the 5′ coding region can all strongly influence translation efficiency, protein folding, and overall protein yield.[8, 15] As such, CAI-driven methods can overlook higher-order features and miss essential optimizations, falling short of offering consistent, reliable yield gains.

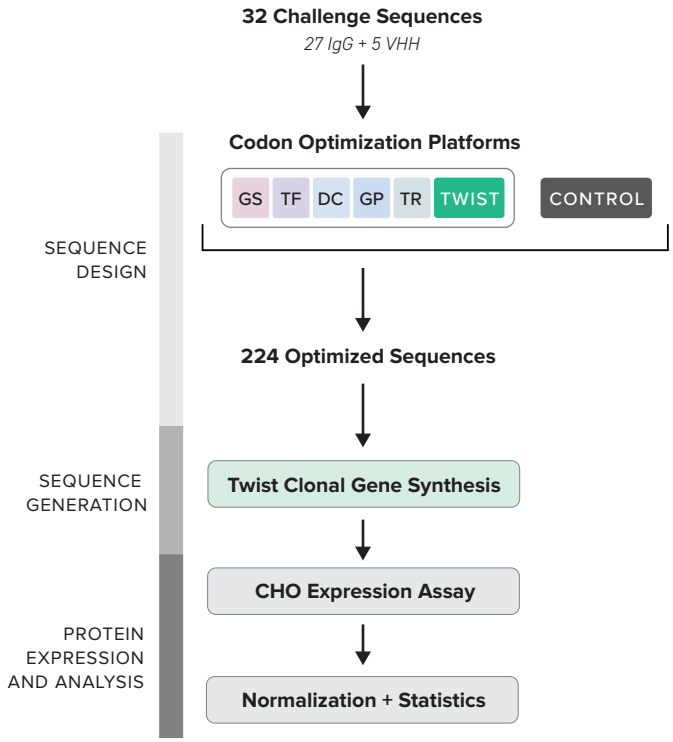
Traditional machine learning (ML) models improve on these limitations by recognizing sequence-expression patterns that may span multiple attributes. Where CAI-driven methods focus on optimizing for the most commonly used and GC-balanced codons, ML is capable of optimizing across multiple user-defined features, such as GC content, mRNA structure, and codon pair biases. While an improvement over CAI, traditional ML approaches have similar drawbacks, such as a reliance on pre-defined features, relatively narrow training objectives, and a bias

for specific feature sets. This limits their ability to capture the complex, context-dependent interactions that govern protein expression, including translation efficiency and protein folding.

Large language models (LLMs) are an emergent technology with the potential to greatly improve protein expression relative to legacy approaches.[10] These models, trained on vast genomic and transcriptomic datasets, can learn statistical representations of natural protein coding sequences, enabling the design of expression-optimized sequences based on species-specific codon preferences as well as higher-order, long-range interactions.

At Twist, we have developed an LLM-driven codon optimization tool, with the goal of improving upon our previous rules-based methods. To assess whether this algorithm offers theoretical advantages to protein expression, we conducted a benchmark study comparing six codon optimization strategies across a panel of antibody challenge sequences, collectively representing high, medium, and low expression proteins. The study integrates experimental expression measurements from CHO cell systems with sequence-level analysis across hundreds of optimized coding sequences.

Study Design



Sequence Optimization

Six commonly used codon optimization platforms were compared in this study (**Table 1**). Each was challenged to optimize the coding sequences (specifically the variable region) of 27 IgG and 5 VHH-Fc constructs for expression in Chinese Hamster Ovary (CHO) cells, representing an equal spread of high, medium, and low expression sequences. For the 32 antibodies, CHO expression data that had previously been generated on an internal CAI-based platform was used as a control. Together, this resulted in 224 codon optimized sequences for CHO-expression analysis.

To determine whether each platform improved protein expression, performance (protein yield) was normalized to the control, enabling each platform to be compared against the same baseline reference point.

Each challenge sequence used the same Kozak and leader sequences, neither of which was subject to optimization. Additionally, all IgG constant regions were excluded from optimization.

OPTIMIZATION PLATFORM	OPTIMIZATION METHOD	AVAILABILITY
Control	CAI	N/A
Platform GS	ML	Commercial
Platform TF	ML	Commercial
Platform DC	CAI	Open-source
Platform GP	CAI	Open-source
Platform TR	LLM	Open-source
Twist Codon Optimization	LLM	Open-source

Table 1: Optimization Platforms Compared

Protein Synthesis & Purification

NGS-verified Clonal DNA encoding each codon optimized sequence was produced at microprep scale using the Twist Clonal Gene Synthesis workflow. All samples were transfected in triplicate at the 1mL scale in CHO cells. Following incubation, secreted protein was purified via Protein-A chromatography resin with a 50mM pH 3 Citrate elution into 1M HEPES pH 8 for neutralization. Purified protein was quantified via absorbance at 280nm.

Expression Quantification

Expression values obtained from CHO assays were normalized relative to the control expression data. For each codon optimized sequence, expression measurements from biological replicates were averaged to obtain a single representative value. By analyzing relative expression levels, the effect size of each optimization platform could be assessed across sequences with heterogeneous baseline expression levels.

To evaluate whether individual codon optimization methods significantly improved expression relative to the control, one-sided paired Wilcoxon signed-rank tests were performed. This test was selected because it allowed each optimization method to be evaluated on the same codon optimized sequence without imposing the assumption of a normal distribution of expression differences.

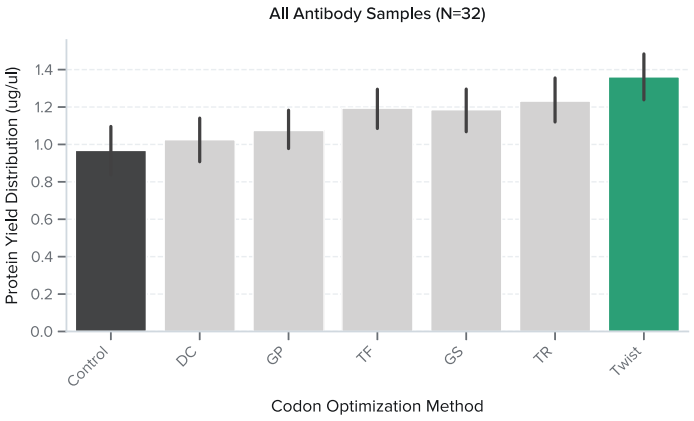


Figure 1. Mean CHO expression by optimization method. Figure shows mean antibody expression levels—as measured by protein yield (µg/µL)—with 95% confidence intervals across the 32 test antibody sequences.

Results

LLM-Driven Approaches Outperform ML & Rules-Based Methods

Both commercial ML and LLM methods provided meaningful gains over the control (**Table 2**), but the LLM-driven methods define the top performance tier.

Among the six optimization platforms, Twist Codon Optimization proved most effective, followed by Platform TR (**Figure 1**). Median gain in normalized protein yield was 27.0% and 14.7% for Twist Codon Optimization and Platform TR, respectively.

Notably, 87.5% of the 32 antibodies showed improved protein yield over the control baseline following Twist Codon Optimization, with the platform outperforming all other methods for 16 sequences (**Figures 2 and 3**). These results suggest that the gain in protein yield is not driven by rare outliers, but rather reflects the robust capability of this platform to deliver expression gains across a diverse set of test sequences. This is reinforced by statistical analyses in which the overall positive effect of Twist's Codon Optimization Platform on protein yield proved highly significant, demonstrating a paired Wilcoxon signed-rank test p-value of 2.66×10^{-6}.

It is noteworthy too that the Twist Codon Optimization platform was able to rescue protein expression for low-expression sequences (**Figure 4**), demonstrating the platform's ability to effect change for particularly challenging antibody sequences.

PLATFORM	METHOD	MEDIAN	MEDIAN GAIN VS CONTROL	% SAMPLES IMPROVED	BEST-IN-SAMPLE WINS	WILCOXON P VS CONTROL
DC	CAI	1.096	9.6%	50.0%	1	0.196
GP	CAI	1.058	5.8%	53.1%	4	0.116
GS	ML	1.109	11.0%	71.9%	1	0.024
TF	ML	1.044	4.4%	65.6%	1	0.002
TR	LLM	1.147	14.7%	81.2%	7	3.06×10^{-4}
Twist Codon Optimization	LLM	1.270	27.0%	87.5%	16	2.66×10^{-6}

Table 2. Expression benchmark summary. Median normalized protein yield reported. Best-in-sequence wins = number of sequences for which the platform was the top-performer in protein yield gain relative to control. Wilcoxon p vs control = a paired Wilcoxon signed-rank test p-value comparing the median IgG yield of each optimization method against its matched control yield across all 32 antibody samples.

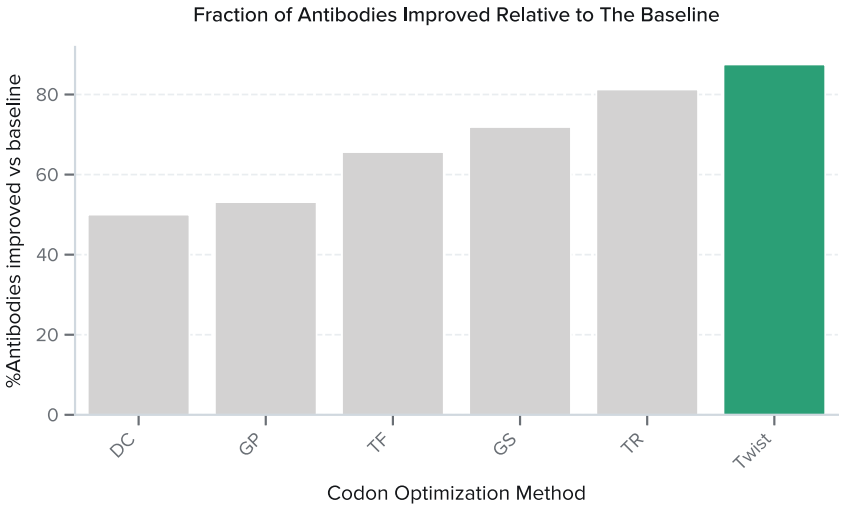


Figure 2. Fraction of test sequences improved relative to the control baseline. The Twist Codon Optimization approach improves 87.5% of the panel and Platform TR improves 81.2%, indicating that the LLM-driven methods are not just stronger on average but also more reliable across diverse optimization tasks.

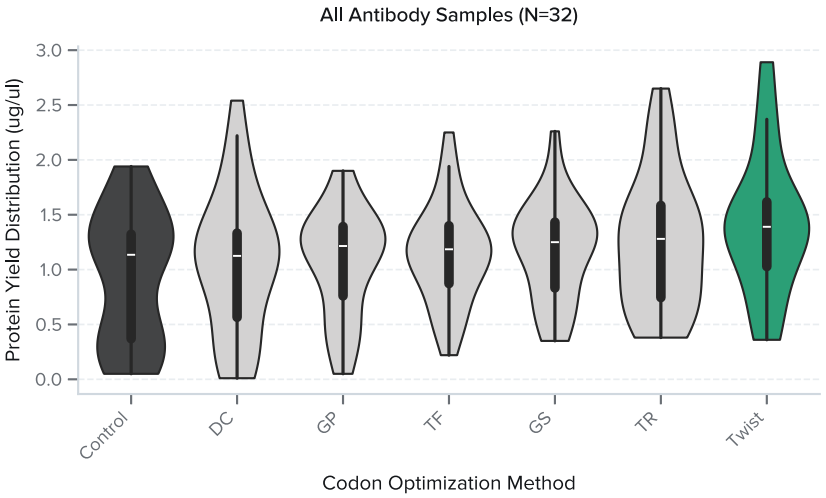


Figure 3. Per-sequence expression changes. Violin plot showing the distribution of average protein yield per platform. Twist Codon Optimization delivered superior improvements to protein expression for 16 of 32 sequences. Platform TR proved second most capable, leading for 7 sequences. Expression for two sequences were not improved by any of the six optimization methods.

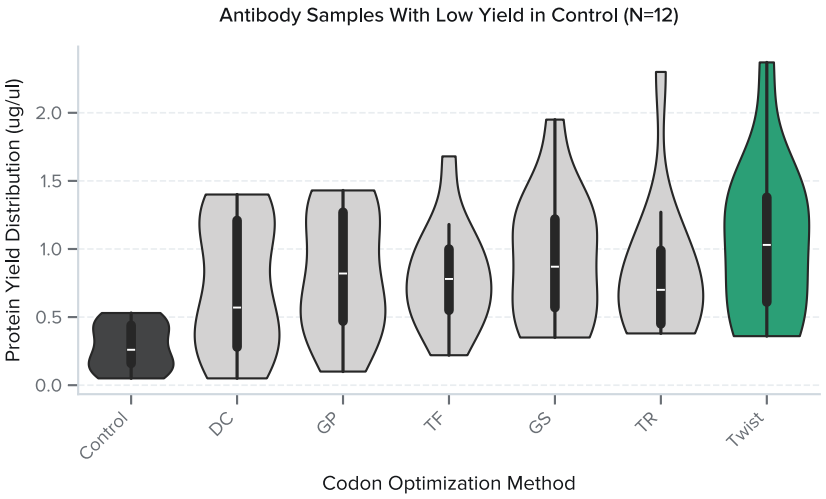


Figure 4. Changes in protein yield (g/L) for low-expression antibodies. Twist's Codon Optimization Platform outperformed all competitors with greater average gains in yield for these low-expression protein sequences.

Sequence-Level Analysis

CAI is Weakly Associated With Expression Level

We aimed to understand whether LLM-driven codon optimization methods could outperform traditional methods. Our results suggest that LLMs consistently outperform both CAI-and traditional ML approaches. This is likely due to the multifactorial nature of protein expression. Unlike traditional methods, LLMs treat each codon as a token that interacts, both locally and distally, with other tokens in the protein sequence. Doing so enables LLMs to look beyond the influence of any individual factor (such as GC content). In contrast, CAI-based and traditional ML methods tend to be biased towards specific features, thus limiting their ability to predict higher-order codon-interactions.

This is evidenced by the weak relationship (Spearman correlation of approximately ρ = 0.01) observed between sequence CAI score and protein expression (**Figure 5**). A simple proxy based on known sensitivities of early gene coding regions (first 30 codons) to GC content showed a stronger—albeit still weak—positive association with expression, with Spearman ρ = 0.22. This suggests that codon optimization platforms that focus heavily on individual parameters, be it CAI or GC content, are likely to have marginal effects on protein yield. In light of these results we recommend users utilize LLM-based optimization algorithms over CAI-based algorithms for codon optimization when looking to increase protein expression.

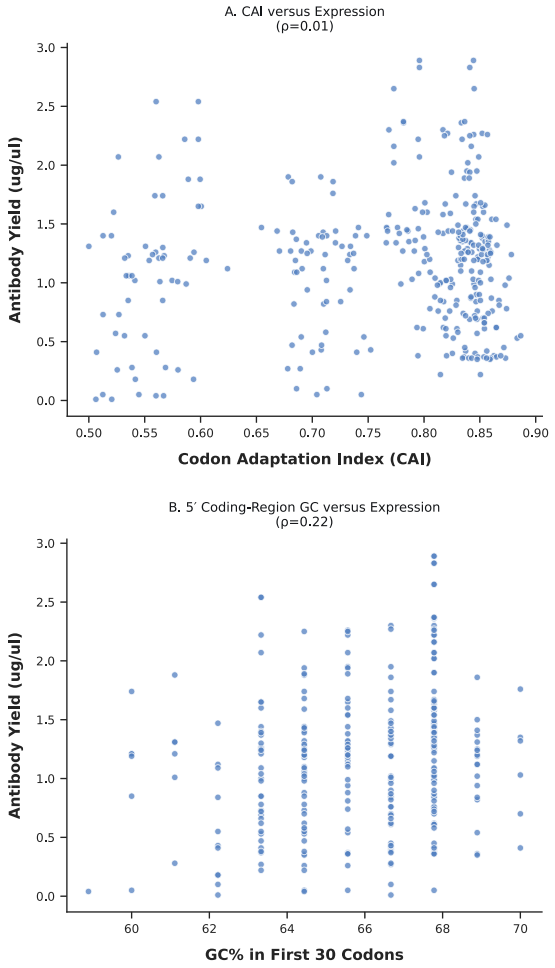


Figure 5. (A) CAI versus expression relationship and (B) 5′ coding-region GC versus expression. Each point represents an optimized sequence.

Discussion

In this study, we set out to determine if LLM-driven codon optimization could outperform rules-based and traditional ML methods by eliciting higher protein expression for 32 antibody challenge sequences.

Our results show that LLM-driven approaches are better suited to codon optimization. Among the six tested platforms, the Twist Codon Optimization platform produced greater protein expression on average and improved yield for the greatest number of sequences. Collectively, this indicates the platform is capable of driving a strong and robust improvement in protein yield.

That both LLM-driven approaches (Platform TR and Twist Codon Optimization) outperformed the rules-based and ML methods suggests that codon optimization should be treated as a global sequence design problem, rather than a local codon substitution problem. This is evidenced in the weak relationship between CAI and protein expression. Such findings are exactly what one would expect if expression depends on higher-order sequence context rather than codon frequency alone.

LLM-driven approaches likely have superior performance because they capture higher-order dependencies between codons across the full coding sequence. This allows them to preserve biologically realistic codon-pair patterns while still optimizing local sequence context and host compatibility. Between the two LLM platforms, Twist's Codon Optimization Platform proved most effective at improving overall protein yield, likely reflecting differences in LLM neuronal network architectures.

From an industrial perspective, for scientists aiming to maximize protein expression, these findings argue for treating codon optimization as a global sequence-design task. Rather than asking which codons are most frequent in a host, the more useful question is which full synonymous sequence best balances context, structure, manufacturability, and expression outcome. As such, LLM algorithms may be expected to outperform both ML and CAI-based algorithms for the optimization of protein expression.

References

1. Giles Obinna Ndochinwa, et al. “New Advances in Protein Engineering for Industrial Applications: Key Takeaways.” Open Life Sciences, vol. 19, no. 1, 1 Jan. 2024, https://doi.org/10.1515/biol-2022-0856.

2. Basher, Nosiba S. “Protein Engineering and Drug Discovery: Importance, Methodologies, Challenges, and Prospects.” Biomolecules, vol. 15, no. 11, 20 Nov. 2025, pp. 1628–1628, https://doi.org/10.3390/biom15111628.

3. Li, Chenyi, et al. “Protein Engineering for Improving and Diversifying Natural Product Biosynthesis.” Trends in Biotechnology, vol. 38, no. 7, 1 July 2020, pp. 729–744, https://doi.org/10.1016/j.tibtech.2019.12.008.

4. Miller, David C., et al. “Combining Chemistry and Protein Engineering for New-To-Nature Biocatalysis.” Nature Synthesis, vol. 1, no. 1, Jan. 2022, pp. 18–23, https://doi.org/10.1038/s44160-021-00008-x.

5. Plotkin, Joshua B., and Grzegorz Kudla. “Synonymous but Not the Same: The Causes and Consequences of Codon Bias.” Nature Reviews Genetics, vol. 12, no. 1, 23 Nov. 2010, pp. 32–42, https://doi.org/10.1038/nrg2899.

6. Moss, McKenze J., et al. “The Effects of Codon Usage on Protein Structure and Folding.” Annual Review of Biophysics, 22 Dec. 2023, https://doi.org/10.1146/annurev-biophys-030722-020555.

7. Grégory Boël, et al. “Codon Influence on Protein Expression in E. Coli Correlates with MRNA Levels.” Nature, vol. 529, no. 7586, 21 Jan. 2016, pp. 358–363, https://doi.org/10.1038/nature16509.

8. Li, Zheng-Mei, et al. “Factors Affecting the Expression of Recombinant Protein and Improvement Strategies in Chinese Hamster Ovary Cells.” Frontiers in Bioengineering and Biotechnology, vol. 10, 4 July 2022, https://doi.org/10.3389/fbioe.2022.880155.

9. Sharp, Paul M., and Wen-Hsiung Li. “The Codon Adaptation Index-a Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications.” Nucleic Acids Research, vol. 15, no. 3, 1987, pp. 1281–1295, https://doi.org/10.1093/nar/15.3.1281.

10. Demissie, Eden A, et al. “Comparative Analysis of Codon Optimization Tools: Advancing toward a Multi-Criteria Framework for Synthetic Gene Design.” Journal of Microbiology and Biotechnology, vol. 35, 10 Apr. 2025, https://doi.org/10.4014/jmb.2411.11066.

11. Novoa, Eva Maria, and Lluís Ribas de Pouplana. “Speeding with Control: Codon Usage, TRNAs, and Ribosomes.” Trends in Genetics, vol. 28, no. 11, Nov. 2012, pp. 574–581, https://doi.org/10.1016/j.tig.2012.07.006.

12. Yu, Chien-Hung, et al. “Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-Translational Protein Folding.” Molecular Cell, vol. 59, no. 5, Sept. 2015, pp. 744–754, https://doi.org/10.1016/j.molcel.2015.07.018.

13. Yang, Qian, et al. “ERF1 Mediates Codon Usage Effects on MRNA Translation Efficiency through Premature Termination at Rare Codons.” Nucleic Acids Research, vol. 47, no. 17, 14 Aug. 2019, pp. 9243–9258, https://doi.org/10.1093/nar/gkz710.

14. Kimchi-Sarfaty, C., et al. “A “Silent” Polymorphism in the MDR1 Gene Changes Substrate Specificity.” Science, vol. 315, no. 5811, 26 Jan. 2007, pp. 525–528, https://doi.org/10.1126/science.1135308.

15. Kirchner, Sebastian, et al. “Alteration of Protein Function by a Silent Polymorphism Linked to TRNA Abundance.” PLOS Biology, vol. 15, no. 5, 16 May 2017, p. e2000779, https://doi.org/10.1371/journal.pbio.2000779.

16. Gebert, Daniel, et al. “Widespread Selection for Extremely High and Low Levels of Secondary Structure in Coding Sequences across All Domains of Life.” Open Biology, vol. 9, no. 5, May 2019, https://doi.org/10.1098/rsob.190020.