Codon Optimization for Therapeutic Genes to Potentially Increase Transgene Expression in Mammalian Cell Lines

Grant Proposal

Svasti Rathi

Massachusetts Academy of Math and Science at WPI

Worcester, MA

Abstract

Codon optimization is a common practice in molecular biology that changes synonymous codon usage to improve translation efficiency without changing the final protein sequence. This project investigates whether codon optimization driven by an AI algorithm developed at the Tai Lab at UMass Chan Medical School can increate RNA and protein expression of therapeutic genes in mammalian cell lines, and whether these results are consistent across various genes. Three therapeutic genes, ST3GAL5, SMN1, and A1AT, were chosen and tested on, using both optimized and non-optimized constructs. After transfections, the RNA and protein levels were assessed using quantitative PCR (qPCR), BCA protein assays, Western blotting, and sandwich ELISA techniques, the latter being used specifically for secreted proteins. Multiple experiments and replications were performed to ensure consistency between results. ST3GAL5 had no detectable RNA or protein expression within the tested conditions. SMN1 expression occurred, but the optimized constructs showed no greater expression compared to the controls. Antibody cross-reactivity also occurred, making the protein analysis more challenging. In contrast, A1AT successfully and consistently showed increased protein expression in the optimized constructs compared to the non-optimized constructs. These findings suggest that codon optimization driven by the AI algorithm can increase transgene expression in mammalian cells but is currently not effective for all genes. The results emphasize the need for gene-specific analysis and tests, along with reliable detection methods for the genes.

Keywords: codon optimization, therapeutic genes, ST3GAL5, SMN1, A1AT, gene dependency

Codon Optimization for Therapeutic Genes to Potentially Increase Transgene Expression in Mammalian Cell Lines

Codons

A codon is a string of three nucleotides, most found in mRNA, that codes for certain amino acids or certain start/stop signals during protein synthesis. Codons are the basic unit of the genetic code, and they provide the instructions on protein synthesis (Britannica Editors, 2025).

Codon Bias

Even though multiple codons can code for the same amino acid, every organism uses the codons at different rates and for different reasons. These preferences are directly related to the tRNAs, or the transfer RNAs, that are available within the organism. Transfer RNAs are RNA molecules that are necessary for protein synthesis, bringing the specific amino acids depending on the codon. The preferences are also dependent on how efficient translation is, which is once again related to the tRNAs (Quax et al., 2015). When scientists try to express a gene from one organism in another, it is possible that there will be some mistakes or mess ups in the process due to mismatches within the codon usages. It is also possible the translation will slow down, protein folding efficiency will decrease, and even general protein production and expression will lower.

Codon Optimization

Codon optimization allows for the gene's codons to be a better fit and match better to the preferred codons for the new organism by slightly changing the genetic code. Codon optimization changes a gene's DNA sequence to increase protein expression, but the protein

itself remains unchanged. This can improve the stability of the mRNA and increase the efficiency of translation within the organism. Codon optimization has become more widely used in multiple medical fields, such as biotechnology, vaccines, recombinant protein expression, and even gene therapy. Essentially, codon optimization can better the ability of cells to synthesize better and more stable versions of the target protein (Paremskaia et al., 2024).

Previous Research

In the Tai Lab in the Department of Microbiology at UMass Chan Medical School, previous research has been completed regarding reporter genes and the codon optimization technique to evaluate its effects on protein expression, with general findings being positive. However, the technique was not tested on therapeutic genes. Testing these techniques on therapeutic genes and getting positive results would allow for this to be used even more widely in gene therapy, which would be revolutionary in the medical field. I previously tested on the gene *ST3GAL5*, which codes for the GM3 Synthase, and enzyme necessary for glycolipid synthesis in the nervous system. However, there is a lack of reliable assays and antibodies to test for proper enzyme activity and effects of this gene, so the work with that gene was discontinued and new therapeutic genes were chosen to test the codon optimization.

Current Goals

In this study, we want to find if using a codon optimization algorithm on the therapeutic genes *SMN1* and *A1AT* effect the protein expression in certain cell lines. The hypothesis of this project is that using the codon optimization techniques on the *SMN1* and *A1AT* genes will improve expression of the therapeutic genes. The two new genes that will be tested on, as previously mentioned, are the *SMN1* gene and the *A1AT* gene. The *SMN1*, or the Survival Motor

Neuron 1, gene encodes for a protein that is necessary for general function of motor neurons. Loss or mutations of this gene can cause Spinal Muscular Atrophy, or SMA. SMA is a genetic disorder characterize by muscle atrophy and weakness (SMN1 gene). The *A1AT*, or Alpha-1-Antitrypsin, gene is a protease inhibitor. It is produced within the liver and gets released into the blood stream, protecting lung tissue by inhibiting certain enzymes from degrading the tissue. A loss or mutation of this gene can cause excessive protease activity, resulting in lung damage and possibly disordered such as emphysema of COPD (Alpha-1 antitrypsin deficiencys).

Section II: Specific Aims

This proposal's objective is to assess whether the use of the codon optimization algorithm increases protein expression of therapeutic genes in mammalian cell lines.

Our long-term goal is to enhance the efficiency and reliability of therapeutic gene expression for gene therapy applications by refining codon optimization strategies. The central hypothesis of this proposal is that the codon optimization algorithm will significantly increase protein expression of the therapeutic genes in mammalian cell lines compared to their wild-type sequences. The rationale is that codon usage strongly influences transcription, translation, and protein stability; optimizations of these factors can increase the performance of the transgene, thereby leading to more effective gene therapy. Work proposed here will test various optimized variants of therapeutic genes to establish which codon optimization strategies result in the highest and most reproducible levels of expression.

Specific Aim: To assess and compare the protein expression levels of codon-optimized versus wild-type versions of ST3GAL5, SMN1, and A1AT in mammalian cell lines.

The expected outcome of this work is the identification of those codon optimization designs that most effectively enhance transgene expression, while providing insights to improve both therapeutic gene developments and future optimization algorithm design.

Section III: Project Goals and Methodology

Relevance/Significance:

The goal of this project was to develop consistently high and reliable expression of therapeutic genes in mammalian cells, which is one of the key challenges for gene therapy. The present work contributes directly to the development of safer and more effective gene treatments by evaluating the impact of different codon optimization strategies on protein output for the relevant gene, namely ST3GAL5, SMN1, and A1AT. Enhancement of transgene expression strengthens therapeutic effects and allows a potential decrease in vector dosages, even enhancing long-tern gene stability. For that reason, this research will be very relevant to further development of gene therapy technologies.

Innovation:

This project tests a newly designed codon optimization algorithm at the Tai Lab that included tissue-specific optimization strategies, rather than depending on generalized or commercial optimization methods. This new algorithm diverges from the standard approaches in that it fits codon usage to the translational environment of tissues, which can increase efficiency and precision. By directly comparing the several variants of optimized coding sequences for various therapeutic genes, this research offers a new systematic study that will validate and further develop next-generational gene optimization tools.

Methodology

The experiments were performed using mammalian cell cultures and general molecular biology techniques. Mouse Neuro2a, human HEK293, Chinese Hamster Ovary (CHO), and Mouse liver (AML12) cells were used to evaluate gene expression in different cellular environments. We designed wild type and optimized constructs for ST3GAL5, SMN1, and A1AT using an AI-based codon optimization algorithm. They were then introduced into cells using transfection reagents.

We assessed the protein expression with Western blot analysis, using antibodies specific for the genes and beta-actin as a normalization control. RNA expression was evaluated using RT-PCR testing (reverse transcriptase PCR) with specific primers for the genes. We quantified secreted A1AT protein levels using Sandwich ELISA assays on the supernatants of the cell

cultures. Standard reagents, gel electrophoresis systems, imaging systems and devices, and plate readers were used throughout the study as well.

Specific Aim:

Determine whether codon optimization improves the protein expression of therapeutic genes in mammalian cells. The goal is to measure and compare wild-type and codon-optimized protein expression levels of ST3GAL5, SMN1, and A1AI using established molecular assays. This aim focuses on determining if tissue-specific and algorithm-guided codon optimization enhance expression in the tested cell lines. Out methodology includes mammalian cell lines transfected with wild-type and optimized variants of the gene, and three paths taken depending on the type of gene; extraction of cell lysate, leading to Western blots for quantification; extraction of RNA leading to RT-PCR for quantification, or extraction of cell supernatant leading to ELISA assays for quantification. All samples will be normalized with housekeeping genes, or genes that always transcribe, for accurate quantification (Joshi et al., 2022). We will test multiple optimization techniques (Brain, liver, muscle, and GenScript optimized) to determine which strategy yields the highest protein expression for each therapeutic gene. Our rationale is that codon usage directly impacts gene stability and efficiency of translation. Strategies aimed at optimizing codon usage to match that of highly expressed host genes have been shown to elevate protein expression. These findings are supported by earlier studies where a codon-optimized SMN1 transgene expressed significantly higher protein levels in vivo when compared with the wild type (Xie et al., 2024). This supports out hypothesis that codon optimization done by the AI-trained model can increase the expression levels in mammalian systems.

Justification and Feasibility. The chosen techniques of standard transfection, lysate, RNA, and supernatant extraction, and Western blots, RT-PCR, and ELISA assays are all widely used for quantifying concentrations of proteins and are highly feasible within the time scale of this project. All measurements will be normalized using housekeeping genes to ensure a sufficient reproducible quantification, as recommended in previous studies (Joshi et al., 2022). The use of multiple readouts (protein and RNA) will strengthen the design of the experiment and allowing us to ensure where the optimization is having the greatest effect. Codon optimized sequences will be created from the Ai-trained model or bought commercially, both which are available and reproducible. The experimental workflow is modular, and each gene and optimization strategy can be evaluated independently to reduce technical risks. Overall, the proposed methods are technically sound, inexpensive to perform, and can be reproducibly performed in any academic molecular biology lab.

Prior research shows that codon-optimized SMN1 resulted in high protein levels in mouse tissues compared to non—optimized constructs (Figure 1). This figure demonstrates that codon optimization can have a direct impact on protein abundance, which supports the relevance of our approach in testing similar effects in our therapeutic genes (Xie et al., 2024).

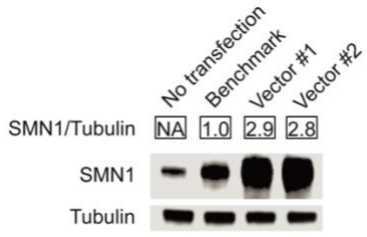

Figure 1: Increased SMN protein expression in Neuro2a cells transfected with codon-optimized human SMN1 compared to wild-type SNM1 (Xie et al., 2024)

Summary of Preliminary Data.

ST3GAL5 Expression: Western blot analysis found no detectable ST3GAL5 protein signal in transfected Neuro2a cells, despite successful transfection controls (Figure 2). RT-PCR analysis confirmed amplification in both RT and no-RT samples, indicating DNA amplification instead of true ST3GAL5 overexpression (Figure 3).

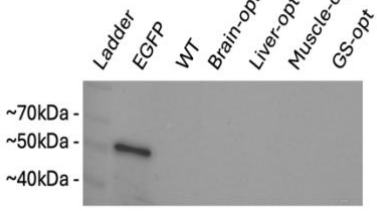

Figure 2: ST3GAL5 expression in transfected Neuro2a cells by Western Blot, showing no detectable protein expression.

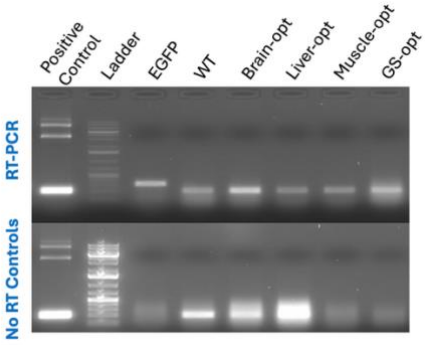

Figure 3: ST3GAL5 expression in transfected Neuro2a cells by Rt-PCR. Shows amplification in both RT and no-RT controls, suggesting DNA contamination rather than RNA-derived expression

SMN1 Expression: Western blot analysis detected SMN1 protein expression across all constructs, including wild-type and codon-optimized variants (Figure 4). Expression levels looked similar between all constructs, suggesting high basal expression in Neuro2a cells that were hiding the potential overexpression effects on the optimization.

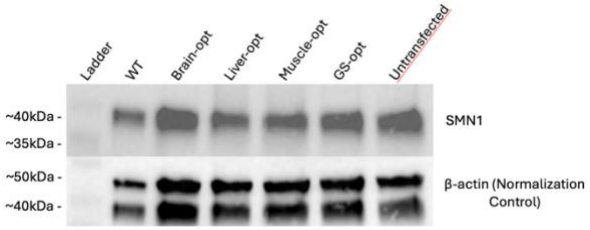

Figure 4: Western blot showing SMN1 expression in Neuro2a cells. Comparable expression levels were observed across both the wild-type and codon optimized constructs after normalization to beta-actin.

A1AT Expression: ELISA analysis indicated higher A1AT protein concentration in the cell culture supernatants from transfected cells compared to untransfected controls (Figure 5). Codon optimized constructs showed greater A1AT expression than wild-type constructs, indicating sucessful improvement of protein expression due to the optimization algorithm.

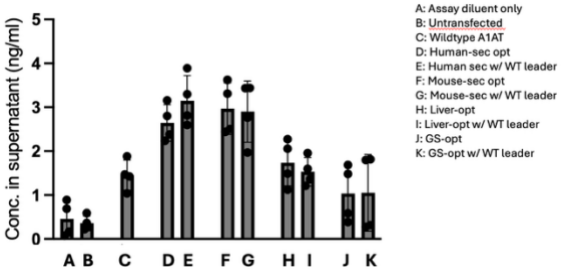

Figure 5: ELISA quantification of A1AT secretion. Codon-optimized constructs showed increased A1AT concentration in cell supernatants

Expected Outcomes The anticipated result of this aim is to determine whether codon optimization increases protein expression of therapeutic genes in mammalian cells and whether there is a particular method of optimization that works best for each gene. We expect that, compared to the wild-type sequence, codon-optimized constructs will increase protein expression to the same degree or to a greater degree depending on the gene and optimization method (Angov, 2011).

Potential Pitfalls and Alternative Strategies. One potential pitfall of our aim is that many genes may display low or undetectable protein expression in a specific cell line even after successful transfection, as observed in preliminary data for ST3GAL5. To confirm expression, we will test expression in alternative mammalian cell lines and/or use more sensitive detection methods or epitope-tagged constructs. Secondly, in the case of high endogenous expression of the gene of interest, as seen for SMN1, a modest increase in expression may be masked by the high background. We will control for this by using more sensitive quantification methods, or alternative cell lines that display lower background expression. Additionally, codon optimization may increase mRNA levels without a detectable increase in protein expression, which will be addressed by measuring mRNA-to-protein ratios to evaluate translational efficiency. Finally, the last potential pitfall of our aim is that tissue-specific optimization may not consistently show improved expression compared to generalized optimization. This would be analyzed on a gene-specific basis; however, this information would still be useful in informing future optimization strategies (Taylor et al., 2017).

Section IV: Resources/Equipment

All the major resources and equipment required for this project are common to most standard academic molecular biology laboratories. These include cell culture tools (biosafety cabinet and CO_2 incubator) for maintenance and transfection of mammalian cell lines; general analytical equipment such as plate readers, gel electrophoresis and Western blot apparatus, and PCR machines. Refrigerated centrifuges, spectrophotometers, fluorescence/chemiluminescence imaging, and basic laboratory equipment for RNA, protein, and supernatant extraction are similarly common and are used.

Section V: Ethical Considerations

Section VI: Timeline

Section VII: Appendix

Section VIII: References

Angov, E. (2011). Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnology Journal*, *6*(6), 650–659. https://doi.org/10.1002/biot.201000332

Britannica Editors. (2025, April 1). Codon. *Encyclopædia Britannica.* https://www.britannica.com/science/codon

Joshi, C. J., Ke, W., Drangowska-Way, A., O'Rourke, E. J., & Lewis, N. E. (2022). What are housekeeping genes? *PLOS Computational Biology*, *18*(7). https://doi.org/10.1371/journal.pcbi.1010295

Paremskaia, A. I., Kogan, A. A., Murashkina, A., Naumova, D. A., Satish, A., Abramov, I. S., Feoktistova, S. G., Mityaeva, O. N., Deviatkin, A. A., & Volchkov, P. Y. (2024). Codon-optimization in gene therapy: Promises, prospects and challenges. Frontiers in Bioengineering and Biotechnology, 12. https://doi.org/10.3389/fbioe.2024.1371596

Quax, T. E. F., Claassens, N. J., Söll, D., & van der Oost, J. (2015). Codon bias as a means to fine-tune gene expression. Molecular Cell, 59(2), 149–161. https://doi.org/10.1016/j.molcel.2015.05.035

Taylor, S. C., Laperriere, G., & Germain, H. (2017). Droplet digital PCR versus qPCR for gene expression analysis with low abundant targets: From variable nonsense to publication Quality Data. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-02217-x

U.S. National Library of Medicine. (n.d.-a). Alpha-1 antitrypsin deficiency: Medlineplus genetics. *MedlinePlus*. https://medlineplus.gov/genetics/condition/alpha-1-antitrypsin-deficiency/#causes

U.S. National Library of Medicine. (n.d.-b). SMN1 gene: Medlineplus genetics. *MedlinePlus.* https://medlineplus.gov/genetics/gene/smn1/

Xie, Q., Chen, X., Ma, H., Zhu, Y., Ma, Y., Jalinous, L., Cox, G. F., Weaver, F., Yang, J., Kennedy, Z., Gruntman, A., Du, A., Su, Q., He, R., Tai, P. W., Gao, G., & Xie, J. (2024). Improved gene therapy for spinal muscular atrophy in mice using codon-optimized HSMN1 transgene and HSMN1 gene-derived promotor. *EMBO Molecular Medicine*, *16*(4), 945–965. https://doi.org/10.1038/s44321-024-00037-x

Zhang, Q., Wang, J., Deng, F., Yan, Z., Xia, Y., Wang, Z., Ye, J., Deng, Y., Zhang, Z., Qiao, M., Li, R., Denduluri, S. K., Wei, Q., Zhao, L., Lu, S., Wang, X., Tang, S., Liu, H., Luu, H. H., … Jiang, L. (2015). TQPCR: A touchdown qPCR assay with significantly improved detection sensitivity and amplification efficiency of SYBR GREEN QPCR. *PLOS ONE*, *10*(7). https://doi.org/10.1371/journal.pone.0132666