# Application of Bioinformatics Tools in the Optimization of mRNA Sequences Focused on Vaccine and Therapeutic Development

**Leo Thirso Pinto de Freitas e Souza[1]*, André Brasil Vieira Wyzykowski[1], Vinícius Pinto Costa Rocha[1]**
*[1]SENAI CIMATEC University; Salvador, Bahia, Brazil*

**This *in silico* study explored the optimization of therapeutic RNA modeling using AI (AlphaFold3, Boltz-1, Chai-1). An assembly method combined high-confidence regions to model complex sequences. Quality was evaluated using five standardized metrics (secondary structure, stability, translation, immunogenicity, and production). Simulations and bioinformatics analyses were used to assess the stability and function of the optimized structures. The objective was to develop more accurate and comparable modeling methodologies for therapeutic RNA design, while recognizing the need for subsequent experimental validation.**
**Keywords: Vaccine. Treatment. Messenger RNA. Bioinformatics.**

The optimization of mRNA sequences using artificial intelligence (AI) emerges as a promising strategy for the development of effective therapies and vaccines. Recent algorithms such as AlphaFold3, Boltz-1, and Chai-1 offer the potential to predict RNA structures, paving the way for the assembly of optimized sequences with greater stability and translational efficiency. Given the lack of standardization in evaluating these algorithms, this study proposes and applies a set of five metrics to compare the performance of AlphaFold3, Boltz-1, and Chai-1 in modeling therapeutic mRNA. Our goal is to investigate the *in silico* optimization of mRNA sequences using AI by applying this standardized evaluation framework to contribute to the development of innovative molecules, while acknowledging the need for future experimental validation.

## Materials and Methods

This research is configured as an exploratory and descriptive *in silico* study, conducted entirely in a computational environment. It does not involve the collection of primary data from humans or animals; thus, definitions of population in the sense of field research or ethics committee approval are not applicable. The research is ongoing, with no specific data collection period, utilizing the latest available versions of AI algorithms and RNA sequence databases at the time of analysis.

Initially, the Spike protein sequence of the SARS-CoV-2 virus was used as a model for structural predictions generated by the AI algorithms. The technique used for data collection and analysis involves applying bioinformatics methods and computational modeling.

Therapeutic or biotechnological RNA sequences was selected from public databases such as GenBank and Rfam, prioritizing those with potential for applications in immunotherapy or vaccine development, as identified in the scientific literature. Selected sequences was subjected to predictive structural analysis by three state-of-the-art AI algorithms: AlphaFold3, Boltz-1, and Chai-1. Each algorithm, using its machine learning models, generated multiple predictions of the three-dimensional structure of each RNA sequence, along with confidence metrics specific to each region of the model.

To address the limitations of these algorithms, particularly in regions with multiple MERs, a model assembly technique was implemented. This technique involves identifying and extracting the substructures with the highest confidence scores from each of the models generated for a given sequence. These substructures was

Received on 18 March 2025; revised 27 May 2025.
Address for correspondence: Cleide M.F. Soares. Av. Orlando Gomes, 1845, Piatã, Salvador, Bahia, Brazil. Zipcode: 41650-010. E-mail: leo.souza@aln.senaicimatec.edu.br.
Original extended abstract presented at SAPACT 2025.

J Bioeng. Tech. Health 2025;8(3):260-262
© 2025 by SENAI CIMATEC University. All rights reserved.

then computationally combined to generate an optimized consensus three-dimensional model, aiming to integrate the most robust predictions from each algorithm.

The quality of the assembled models was assessed using a set of standardized criteria proposed in this study.

The five evaluation categories are:

1. **mRNA Secondary Structure Prediction Quality**, assessed by comparing predicted structures with experimental secondary structure data from databases or literature, using concordance metrics such as sensitivity and specificity;
2. **Thermodynamic Stability of mRNA**, quantified by estimating the Gibbs free energy (ΔG) of the predicted structure using bioinformatics software such as the ViennaRNA Package;
3. **Potential Translation Efficiency**, analyzed by calculating the Codon Adaptation Index (CAI) using tools like CAIcal, evaluating the frequency of unfavorable dimers using tools like EMBOSS, and analyzing ribosome binding site (RBS) accessibility using secondary structure prediction software;
4. **Immunogenicity Potential**, assessed by searching for known immunogenic sequence motifs using databases like SIDMAP and predicting dsRNA formation potential using tools such as RNAduplex from the ViennaRNA Package; and
5. **Ease of Production and Purification**, qualitatively assessed by analyzing the structural complexity of the models (presence of knots, pseudoknots) and identifying repetitive sequences that could hinder mRNA synthesis and isolation.

The data analysis involve a quantitative comparison of the evaluation metrics among the models generated by different algorithms and those optimized by the assembly method.

Descriptive statistics was used to summarize the results and identify trends in algorithm performance and the quality of optimized models. The analysis include a discussion of the implications of the results for therapeutic mRNA design, considering the limitations of the algorithms and the challenges in producing and *in vivo* applying these molecules. If molecules with high therapeutic potential are identified through *in silico* analysis, future studies may include *in vitro* experimental validation in the SENAI CIMATEC Health Technologies Institute's laboratory.

Although this research does not involve human subjects, it rigorously adheres to best practices in scientific research, including maintaining detailed records of all methodological steps, ensuring the traceability of analyses, and presenting results transparently to enable reproducibility by other researchers.

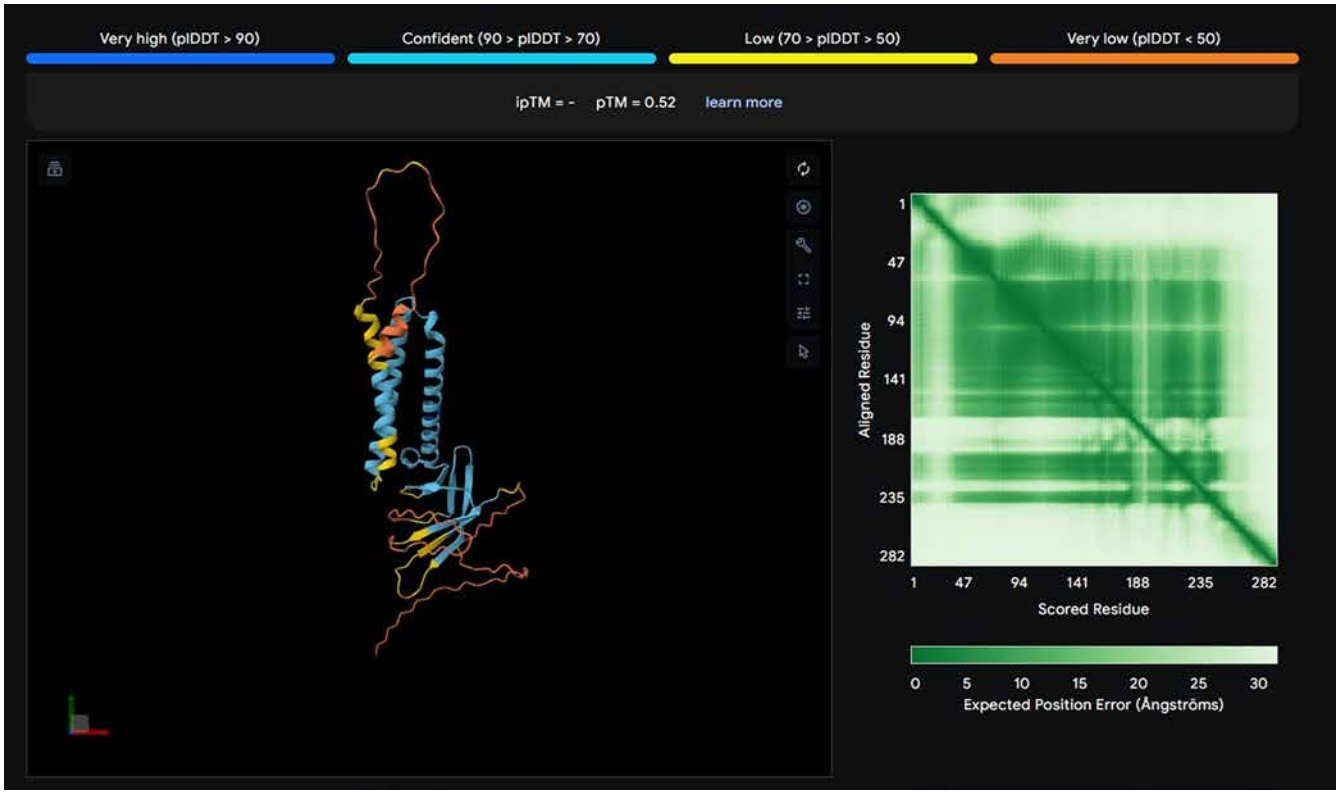
## Theoretical Framework

The versatility of mRNA drives its investigation for on-demand vaccines and therapies, with rapid and safe *in vitro* production, enabling the encoding of various antigens and modulation of immune responses [3,2]. Despite its instability and challenges in *in vivo* expression, algorithms optimize sequences by considering structure and regulation [5-8]. Our laboratory [9-11] seeks to prospect and validate optimization algorithms *in vitro*, initially using the Spike protein sequence of SARS-CoV-2. The application of these tools aims to contribute to the development of more effective vaccines and treatments, with potential for future collaborations (Figure 1).

## Conclusion

This *in silico* study proposed a framework to evaluate the optimization of therapeutic RNA by AI (AlphaFold3, Boltz-1, Chai-1). Despite limitations in clinical applicability and mRNA production, and the need for biological validation, the standardization of metrics and *in silico* identification of promising sequences represent a significant initial step for future RNA therapies.

**Figure 1.** Three-dimensional structure of the SARS-CoV-2 Spike protein sequence predicted by AlphaFold3.




Source: Data generated by AlphaFold3.

## References


1. Wolff JA, Malone RW, Williams P, Chong W, Acsadi G, Jani A, et al. Direct gene transfer into mouse muscle *in vivo*. Science. 1990;247(4949 Pt 1):1465–8.
2. Weissman D. mRNA transcript therapy. Expert Rev Vaccines. 2015 Feb;14(2):265–81.
3. Maruggi G, Zhang C, Li J, Ulmer JB, Yu D. mRNA as a Transformative Technology for Vaccine Development to Control Infectious Diseases. Mol Ther. 2019 Apr 10;27(4):757–72.
4. Machado BAS, Hodel KV, Barbosa JM, Soares MBP, Campos GB, Brustolini OJ, et al. The Importance of RNA-Based Vaccines in the Fight against COVID-19: An Overview. Vaccines (Basel). 2021 Nov;9(11):1343.
5. Gong H, Liu L, Ye Y, Fan B, Shao M, Yang J, et al. Integrated mRNA sequence optimization using deep learning. Brief Bioinform. 2023 Jan 19;24(1):bbac560.
6. Zhang H, Cao L, Gao H, Wu Y, Zhang J, Cai X, et al. Algorithm for optimized mRNA design improves stability and immunogenicity. Nature. 2023 May 2;621(7978):396–403.
7. Zarnack K, Eyras E. Artificial intelligence and machine learning in RNA biology. Brief Bioinform. 2023 Nov 1;24(6):bbad300.
8. Maharjan R, Zhao L, Yang C, Wang W, Luo Z, Liu Z, et al. Machine learning-driven optimization of mRNA-lipid nanoparticle vaccine quality with XGBoost/Bayesian method and ensemble model approaches. J Pharm Anal. 2024 May 8:100996.
9. Costa Rocha VP, Silva TP, Fialho L, Fernandes DS, Gonçalves A, Cezar N, et al. A polyvalent RNA vaccine reduces the immune imprinting phenotype in mice and induces neutralizing antibodies against omicron SARS-CoV-2. Heliyon. 2024 Apr;10(4):e25849.
10. Rocha VPC, Braga CJM, Costa Rocha VP, Silva TP, Gonçalves A, Fialho L, et al. High-Content Imaging-Based Assay for SARS-CoV-2-Neutralizing Antibodies. Vaccines (Basel). 2024 Feb 24;12(3):236.
11. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009 May;6(5):343–5.