AI-Guided Optimization of an Antimicrobial Peptide targeting *Bacillus cereus*: Enhancing Activity and Reducing Toxicity Through Machine Learning Tools

By: Shaurya Thakker

Abstract

The growing threat of antibiotic resistance has underscored the urgent need for alternative therapeutic strategies. Antimicrobial peptides (AMPs), naturally occurring molecules with broad-spectrum activity, present a promising solution but face limitations such as toxicity, instability, and delivery challenges. This study aimed to enhance the safety and efficacy of a low-performing AMP active against *Bacillus cereus* by utilizing artificial intelligence (AI) and machine learning (ML) tools. Four computational platforms - CAMPr4, ToxinPred3, AlphaFold, and ChatGPT - were used to predict antimicrobial potential, evaluate toxicity, model structure, and streamline peptide design. Starting with an AMP scoring 0.2895 in CAMPr4, sequence modifications were made to increase net charge and decrease toxicity. The final optimized peptide, named *Cereus-Black93*, achieved a predicted antimicrobial score of 0.93 and a significantly reduced toxicity score of 0.27. Structural modeling confirmed a stable alpha-helical conformation. These results demonstrate the potential of AI-driven approaches to accelerate the design of novel AMPs and pave the way for future in vitro validation and therapeutic development.

Introduction

The global rise of antibiotic resistance represents one of the most pressing public health concerns of the 21st century. While antibiotics have revolutionized medicine since their discovery, the overuse and misuse of these drugs - especially in agriculture and clinical settings - have accelerated the emergence of resistant bacterial strains. Along with this issue is the slow pace of new antibiotic development, driven in part by high research costs, diminishing returns, and increasing regulatory hurdles. Additionally, antibiotics often target specific bacterial pathways, making them susceptible to the development of resistance. In short, the age of antibiotics is facing a critical turning point [1].

In response to these challenges, scientists have been increasingly turning to alternatives - among them, antimicrobial peptides (AMPs). These small, naturally occurring molecules are part of the innate immune response in many organisms and possess broad-spectrum activity against bacteria, viruses, and fungi [2]. Unlike traditional antibiotics, AMPs often work by disrupting microbial membranes, making it more difficult for pathogens to develop resistance. However, AMPs are not without limitations: some exhibit toxicity to human cells, many degrade quickly in biological environments, and delivery across cell membranes can be problematic [3]. Given these limitations, researchers are exploring how modern technologies - especially artificial intelligence (AI) and machine learning (ML) - can be used to redesign and optimize AMP sequences to enhance their safety, stability, and effectiveness [4].

This research project aimed to use AI/ML-based tools to take a low-performing AMP and transform it into a safer and more potent therapeutic candidate. By combining multiple computational platforms, I sought to optimize a specific AMP derived from *Bacillus cereus*, a bacterium known for causing foodborne illness, and evaluate how changes to its sequence affected both antimicrobial activity and toxicity.

Background

Bacillus cereus is a Gram-positive, rod-shaped, spore-forming bacterium commonly found in soil, contaminated food, and marine environments. It is best known as a cause of foodborne illness, particularly in improperly stored or undercooked foods such as rice and pasta. Upon ingestion, it can produce two types of toxins: one that causes emetic (vomiting) symptoms and another that leads to diarrheal illness [5]. Because *B. cereus* can form heat-resistant spores, it can survive cooking processes if food is not properly handled afterward.

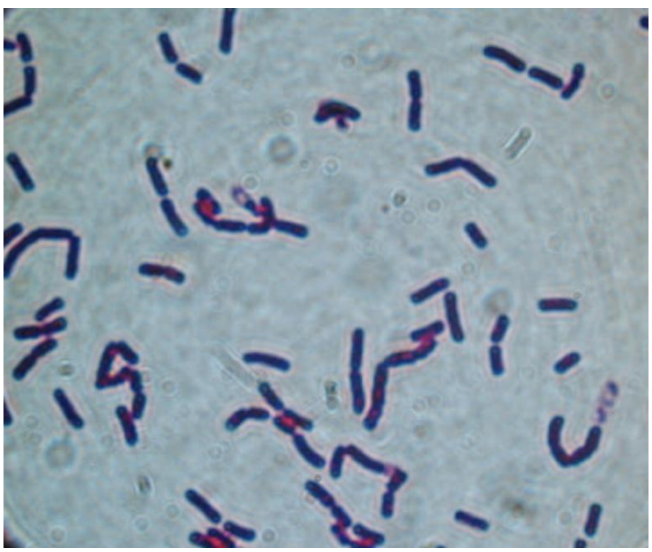

Figure 1: *Bacillus cereus* (Source: BC Centre for Disease Control)

While its pathogenicity is often considered mild, *B. cereus* can be dangerous to immunocompromised individuals and is sometimes implicated in more severe infections, including bacteremia and endophthalmitis [6]. Improving one of these could have implications for food safety and broader antimicrobial applications.

Methods

This project relied on a combination of four AI and ML-based computational tools to predict, analyze, and optimize antimicrobial peptide sequences. The first tool was CAMPr4, an online platform that uses Random Forest machine learning algorithms to assess whether a peptide is likely to function as an AMP [7]. It evaluates peptide sequences based on features such as hydrophobicity, charge, and amino acid composition, and produces a probability score ranging from 0 to 1.

To assess the potential toxicity of the peptide candidates, I used ToxinPred3, a tool that employs ML models trained on toxic and non-toxic peptide datasets [8]. It estimates how harmful a peptide might be to human cells, which is a crucial factor in determining whether a sequence is viable for therapeutic use.

Structural modeling was done using AlphaFold, a deep learning system developed by DeepMind that predicts 3D molecular structures from amino acid sequences with high accuracy [9]. The predicted structure helped validate that the redesigned peptide maintained a stable and effective conformation - specifically, an alpha-helical structure, which is common among potent AMPs.

Lastly, ChatGPT was used to convert sequences into the FASTA format required by the prediction tools, assist with reformatting results, and provide quick coding or syntax support for working with different computational platforms.

Results

The analysis began with a set of AMP sequences extracted from two *Bacillus cereus* strains from the GRAMPA database. I took them out and put them into my own. These sequences were submitted to CAMPr4, which assigned each one a predicted probability of antimicrobial activity. Most sequences scored above the threshold of 0.5, indicating a strong probability of being a functioning AMP. However, one sequence received a particularly low score of 0.2895 and was selected as the candidate for improvement.

	bacterium	sequence	strain	unit	(MIC) value	(MIC) un-logged value	type of structure	AMP probability
27863	B. cereus	AIPWIWIWWLLRKG	NCDC 0240	uM	-0.096910013	0.8	alpha helical	0.8925
33575	B. cereus	RIRFPWPWRWPWWRRVRG	NCDC 0240	uM	-0.096910013	0.8	half alpha helical half random coil	0.9615
22100	B. cereus	WLRAFRRLVRRLARGLRR	CMCC 63301	uM	-0.070581074	0.85	alpha helical	0.952
49882	B. cereus	RIRFPWPWRWPWWPRFRG	NCDC 0240	uM	0.176091259	1.5	random coil	0.8265
35309	B. cereus	AIPWIWIWRLLRKG	NCDC 0240	uM	1.096910013	12.5	alpha helical	0.9835
34850	B. cereus	WGRAFRRGVRRLARGGRR	CMCC 63301	uM	1.16613397	14.66	alpha helical	0.718
51067	B. cereus	AIPWSIWWHLLFKG	NCDC 0240	uM	1.698970004	50	half alpha helical half random coil	0.8655
27000	B. cereus	WLRAFRRLVRRLARLLRR	CMCC 63301	uM	1.725748333	53.18	alpha helical	0.9385
47932	B. cereus	WGRAFRRLVRRLARGLRR	CMCC 63301	uM	1.74647851	55.78	alpha helical	0.968
26667	B. cereus	WGRAFSRGVRRLARGGRG	CMCC 63301	uM	1.80304721	63.54	alpha helical	0.5855
17844	B. cereus	WGRAFSAGVHRLANGGNG	CMCC 63301	uM	1.84565606	70.09	alpha helical	0.352
15534	B. cereus	WGEAFSAGVHRLANGGNG	CMCC 63301	uM	1.852174904	71.15	half alpha helical half random coil	0.2895
45860	B. cereus	AIPWSIWWRLLFKG	NCDC 0240	uM	2	100	half alpha helical half random coil	0.951
24696	B. cereus	GLPWILLRWLFFRG	NCDC 0240	uM	2	100	half alpha helical half random coil	0.9405
33561	B. cereus	APKVNVNALRKGGRVIRKGLGVIGAAGTAHEVYNHVRNRNQG	CMCC 63301	uM	2.428300071	268.1020104	alpha helical	0.9635

Figure 2: Custom database with compiled strains from the GRAMPA database

To optimize this low-performing peptide, I first color-coded its amino acids by property - positive charge, negative charge, hydrophobicity, and polarity - to better understand its functional landscape. Guided by literature indicating that higher net positive charge enhances AMP activity [2], I modified the sequence by introducing more positively charged residues such as lysine and arginine. This revised sequence was then re-analyzed using CAMPr4, which returned a significantly improved probability score of 0.88.

Figure 3: Color-coded amino acids in the new *B. cereus* sequence

Next, I evaluated the safety of this modified peptide using ToxinPred3. The resulting toxicity score was 0.39, just above the suggested maximum of 0.38. Since reducing toxicity was essential, I used ToxinPred3's built-in optimization tool to propose specific sequence substitutions that would maintain antimicrobial activity while reducing predicted toxicity.

Subject	Sequence	ML Score
Seq_1	WGKAFRAGVHRLAR…	0.39

Figure 4: ToxinPred3 ML model predicting a toxicity of 0.39

Once ToxinPred3 gave the suggested changes to the amino acids inside the peptide, I picked the lowest value, which was 0.27. I put in these new mutations and then re-evaluated the peptide. CAMPr4 predicted a score of 0.93, which is higher than the previous sequence.

Seq. ID.	Class	AMP Probability
SEQUENCE_1	AMP	0.93

Figure 5: New peptide that CAMPr4 predicted to have a probability of 0.93

I named this optimized peptide *Cereus-Black93*, where "Cereus-Black" refers to the popular Harry Potter character with the bacteria name played into it, and "93" reflects its final CAMPr4 probability score.

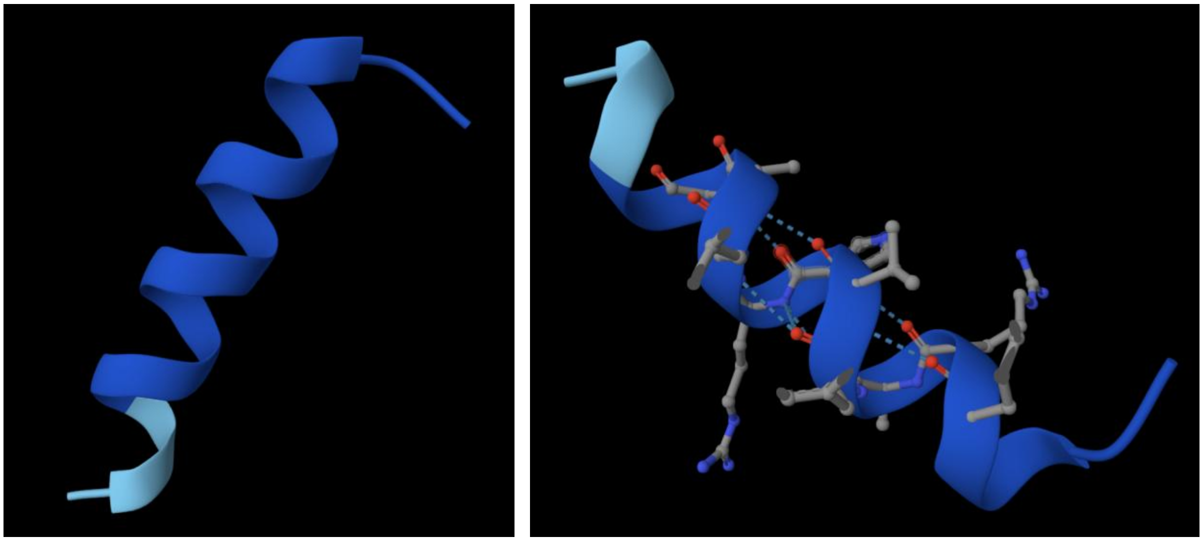

Figure 6: Images of *Cereus-Black93* using Alphafold

To confirm that the final peptide maintained a viable structure, I submitted it to AlphaFold. The resulting 3D model revealed a strong alpha-helical conformation with high confidence across most of the sequence, shown in dark blue regions on the model. This structural validation further supported the peptide's potential as an effective AMP [9].

Conclusion

This project illustrates how computational and AI methods can accelerate and enhance the process of designing functional antimicrobial peptides. What might have taken months or years of wet-lab trial and error was achieved over the course of a single summer through the integration of predictive modeling tools. The final AMP, *Cereus-Black93*, represents a safer and more potent version of its original form, demonstrating how even low-performing sequences can be meaningfully improved with the right design strategy.

However, these results are still theoretical. The next step would be to synthesize the peptide and conduct laboratory experiments to test its antimicrobial efficacy and cytotoxicity in real-world conditions. This would validate the computational predictions and provide valuable data to further improve the AI/ML models used in the design process.

In conclusion, the success of this research highlights a promising path forward - one where biology and computational science work together to develop faster, safer, and more effective treatments for the microbial challenges of the future.

References

1. CDC (2019). *Antibiotic Resistance Threats in the United States*. https://www.cdc.gov/antimicrobial-resistance/media/pdfs/2019-ar-threats-report-

508.pdf?CDC_AAref_Val=https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf

2. Talapko, J., et al. (2022). *Antimicrobial peptides - mechanisms of action, antimicrobial effects and clinical applications*. *Antibiotics*, 11(5), 700. https://doi.org/10.3390/antibiotics11050700
3. Browne, K., et al. (2020). *A New Era of Antibiotics: The Clinical Potential of Antimicrobial Peptides*. *International Journal of Molecular Sciences*, 21(22), 7047. https://www.mdpi.com/1422-0067/21/19/7047
4. Witten, J., & Witten, Z. (2022). *Deep learning regression model for antimicrobial peptide design*. *Scientific Reports*, 12, 240. https://www.biorxiv.org/content/10.1101/692681v1
5. Monique van Hoek & Barney (2017). *Harnessing the Antimicrobial Properties of Komodo Dragon Blood*. *Frontiers in Microbiology*, 8, 1577. https://doi.org/10.3389/fmicb.2017.01577
6. Boulinguez, S., & Viraben, R. (2002). *Cutaneous Bacillus cereus infection in an immunocompetent patient*. *Journal of the American Academy of Dermatology*, 47(2), 324–325.https://pubmed.ncbi.nlm.nih.gov/12140490/
7. Wlasnowolski, O., Gawde, U., & others. (2023). CAMP$_{(R4)}$: A database of natural and synthetic antimicrobial peptides. *Nucleic Acids Research, 51*(D1), D377–D383. https://doi.org/10.1093/nar/gkac933
8. Rathore, A. S., Choudhury, S., Arora, A., Tijare, P., & Raghava, G. P. S. (2024). ToxinPred 3.0: An improved method for predicting the toxicity of peptides. *Computers in Biology and Medicine*, *108926*. https://doi.org/10.1016/j.compbiomed.2024.108926
9. Jumper, J., et al. (2021). *Highly accurate protein structure prediction with AlphaFold*. *Nature*, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2