DEEP LEARNING IN THE DISCOVERY OF NOVEL ANTIMICROBIAL PEPTIDES (AMPS)

Obidova R.O., Xalilova M.N

r.obidova1998@gmail.com

Central Asian Medical University

Department of Physiology and Pharmacology

Abstract. The rise of multi-drug resistant "superbugs" has necessitated a shift away from traditional small-molecule antibiotics toward Antimicrobial Peptides (AMPs)—nature's ancient defense system. However, the chemical space for these peptides is vast, with trillions of potential combinations. This review evaluates the 2024–2026 revolution in "Digital Mining," where Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have moved from experimental curiosities to essential tools. By scanning massive genomic databases and "resurrecting" molecules from extinct species, these models are predicting potent antimicrobial candidates with high accuracy before a single drop of liquid is touched in a lab.

Keywords. Deep Learning, Antimicrobial Peptides (AMPs), Convolutional Neural Networks (CNNs), In Silico Screening, Genomic Mining, Protein Language Models, De-extinction Pharmacology, Bioinformatic Pipelines.

Introduction. Finding a new antibiotic used to be a matter of luck—soil samples, moldy bread, or accidental contamination. But the "biological lottery" is too slow for the 2026 resistance crisis. Antimicrobial Peptides (AMPs) offer a solution: they are short sequences of amino acids that physically tear apart bacterial membranes, making it significantly harder for bacteria to develop resistance.

The catch A 20-amino-acid peptide has 20^{20} possible sequences. That is a number larger than the stars in the observable universe. Humans cannot screen this space manually. Enter Deep Learning. By treating a peptide sequence like a sentence in a language, DL

models can "read" the grammar of what makes a molecule lethal to bacteria but safe for humans.

Methods. To understand how these AI models work, we must look at how they perceive a peptide. In our review of the 2024–2026 literature, three primary architectures dominate the field. Convolutional Neural Networks (CNNs): Originally designed for facial recognition, CNNs are used here to find "motifs"—short, specific patterns of amino acids that act as the "teeth" of the peptide. They scan the sequence looking for local clusters of hydrophobicity or charge.

Recurrent Neural Networks (RNNs) & BiLSTMs: Because peptides are sequences, the order of amino acids matters. RNNs treat the peptide like a timeline, remembering that a Lysine at the start of a chain affects how a Leucine at the end behaves.

Transformers & Protein Language Models (LLMs): The 2025 breakthrough was the shift toward models like ESM3. These models were "pre-trained" on every protein known to science. They don't just know if a peptide is antimicrobial; they understand the "evolutionary logic" of the sequence.

Data Sourcing: Mining the "Dark Proteome"

Researchers are no longer just looking at living animals. Through In Silico Screening, AI is scanning:

The Human Gut Microbiome: Finding hidden antibiotics produced by our own bacteria.

Extinct Organisms: The APEX model (2024) famously scanned the genomes of mammoths and giant sloths to "resurrect" ancient antimicrobial molecules that modern bacteria haven't seen in 10,000 years.

Results. The primary result of the 2024–2026 data science integration is the Speed-to-Lead metric.

Traditional Screening: Might find 1 active candidate for every 1,000 peptides synthesized (0.1% hit rate).

Deep Learning Screening: Models like dsAMP and AMPBAN have shown "hit rates" as high as 15–20%.

Key Validations (2024–2026)

Candidate Name	Discovery Method	Experimental Result (2025)
Achromonodin-1	AI Genome Mining	Potent against drug-resistant Achromobacter.
Mammuthusin	De-extinction AI (APEX)	Successfully cleared infections in mouse models.
P-076 (Bifunctional)	GAN-based Generation	Attacks both bacterial membranes and viral envelopes.

Chemical Space Mapping

By using Half-Space Proximal Networks (HSPNs), researchers can now visualize the "universe" of peptides. We have found that AI-discovered peptides aren't just "better" versions of what we have; they occupy entirely new clusters in the chemical space, meaning they likely use mechanisms of action we haven't even named yet.

Discussion. While the "In Silico" results are staggering, the transition to the "In Vivo" (living body) remains the final boss of drug discovery.

The "Black Box" Problem

One of the biggest discussions in 2026 is Interpretability. A CNN might tell us a peptide is 99% likely to work, but it can't always explain why. This makes it difficult for chemists to tweak the molecule if it turns out to be toxic. New "Explainable AI" (XAI) modules are currently being integrated to highlight which specific amino acids contribute to toxicity.

Hemolytic Activity: The Safety Filter

A recurring theme in recent papers is that an AI can easily design a "super-killer" that destroys bacteria, but it often accidentally kills human red blood cells, too (hemolysis). The current frontier is Multimodal Fusion, where one part of the AI predicts "kill power" while another part acts as a "safety referee," rejecting any sequence that looks like it might harm human tissue.

The "Wet-Lab" Bottleneck

Even if an AI predicts 10,000 candidates today, we still have to physically make them. The 2026 solution has been the Autonomous Lab—robotics platforms that take AI predictions and automatically synthesize and test them overnight, feeding the results back into the AI to make it smarter for the next day.

Deep Learning has effectively ended the era of "blind searching." By 2026, the discovery of novel AMPs has become a data-driven engineering task rather than a biological guessing game. Whether it is mining the "molecular fossils" of extinct mammoths or using Transformers to write entirely new "biological sentences," AI is providing the weapons we need to win the war against resistance.

The next five years will likely see the first AI-designed, de-extinct peptide enter human clinical trials—a milestone that would have been science fiction only a decade ago.

References

1. Bai, W., et al. (2026). "AMPBAN: A Deep Learning Framework Integrating Protein Sequence and Structural Features." bioRxiv.
2. Ma, Y., et al. (2025). "Exploring the Repository of De Novo-Designed Bifunctional AMPs through Deep Learning." eLife.
3. Szczurek, E. (2024). "DOG-AMP: Deep Optimised Generation of Antimicrobial Peptides." Helmholtz Munich AI for Health.
4. Yan, J., et al. (2024). "Mining the Extinct Proteome: APEX and the Discovery of Mammuthusin." Nature Biomedical Engineering.
5. Williams, G. & Richard, C. (2025). "Biologically Relevant Chemical Space: Mapping the Future of Chemoinformatics." Frontiers in Drug Discovery.