Proceedings of the 6th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/2026.38100

Advances in Application of ESM3 and Multimodal Generative Protein Language Models in Intelligent Design of Antimicrobial Peptides

Shang Wang

School of Biopharmacy, China Pharmaceutical University, Nanjing, China
shangwang@stu.cpu.edu.cn

Abstract. Antimicrobial resistance (AMR) has emerged as a significant threat to global public health. Traditional methods for discovering antimicrobial peptides (AMPs) suffer from limitations such as lengthy screening cycles and high costs; however, artificial intelligence (AI) technology offers a novel pathway for the efficient discovery and design of AMPs. This paper reviews the evolution of ESM series models, the multimodal architecture and generation mechanism of ESM3, combs the application progress of deep learning, generative adversarial networks and protein language models in AMP design, and proposes a design-prediction-validation scheme based on ESM3. Existing research demonstrates that ESM3 can integrate protein sequence, structure, and functional information, exhibiting great potential for de novo protein design; the combination of deep learning and generative methods further enhances capabilities in generating candidate AMP sequences and predicting their activity. Building on this, the paper proposes an intelligent AMP design framework centered on ESM3 multimodal reasoning and deep learning-based cascade prediction, establishing a closed-loop process that encompasses candidate generation, structural validation, activity prediction, and experimental verification. Looking ahead, advancements in multimodal models and automated experimental platforms promise to further improve the efficiency and precision of AI-driven AMP design, offering new technological solutions to address the AMR crisis.

Keywords: Antimicrobial peptide, ESM3, Multimodal protein language model, De novo design, Activity prediction

1. Introduction

Antibiotic resistance (AMR) has become a major threat to global public health, and the World Health Organization has listed it as one of the top ten health threats facing mankind. Antimicrobial peptides (AMPs), as an important component of the innate immune system, have received extensive attention due to their advantages of broad-spectrum antibacterial activity and low propensity to induce resistance. However, the traditional AMP discovery methods rely on experimental screening, with long cycle and high cost, which is difficult to meet clinical needs. In recent years, the rapid

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.38100

development of artificial intelligence technology has provided new ideas for large-scale AMP mining and design [1].

Protein language models learn evolutionary laws and functional information contained in protein sequences through self-supervised pre-training on large-scale sequence data. The ESM-1b in the Evolutionary Scale Modeling (ESM) series proposed by Rives et al. [2] pioneered the application of deep language models in protein representation learning, and proved that protein sequences can be modeled like natural language. Subsequently, the ESM-2 developed by Lin et al. [3] achieved improvements in multiple dimensions: the parameter scale was increased from 650 million to 15 billion, the number of Transformer layers was increased from 33 to 48, and rotary position embedding was introduced to enhance long-sequence modeling capability. Combined with the structure prediction module ESMFold, it realized atomic-level structure prediction without multiple sequence alignment, which laid the model foundation for protein function prediction including AMPs.

In terms of generative AI, Das et al. [4] utilized deep generative models combined with molecular dynamics simulation technology to accelerate the discovery of antimicrobial peptides, which verified the feasibility of AI in de novo design of AMP sequences. Afterwards, new AMP design strategies based on diffusion models, generative adversarial networks and other methods kept emerging, demonstrating the unique advantages of generative methods in exploring the vast sequence space, but how to ensure the functionality and safety of the generated peptides is still an urgent problem to be solved.

The ESM3 proposed by Hayes et al. [5] is a multimodal generative protein language model, which can simultaneously reason about the sequence, structure and function information of proteins. It successfully generated a new fluorescent protein esmGFP after simulating five hundred million years of evolution, which proved the strong potential of generative language models in functional protein design. The deepAMP developed by Li et al. [6], as a foundation model in the AMP field, provides a unified framework for large-scale AMP prediction and generation.

This study will systematically review the application progress of ESM3 and multimodal generative protein language models in the intelligent design of antimicrobial peptides, focus on exploring their synergy potential with existing generative AI methods, and discuss the strategies for utilizing deep learning methods to evaluate AMP activity, so as to provide new ideas for coping with the antibiotic resistance crisis.

2. Review of ESM3 and multimodal generative protein language models

2.1. Development of protein language models: from ESM-1 to ESM3

The development of protein language models has experienced an evolution from single sequence modeling to multimodal joint reasoning. In 2021, ESM-1b proposed by Rives et al. [2] adopted 33-layer Transformer with 650 million parameters, and was pre-trained on about 250 million protein sequences, proving that language models can learn information related to protein structure and function from large-scale sequences. In the same year, Meier et al. [7] proved that such models can zero-shot predict the functional effects of missense mutations without additional training. ESM-2 [3] in 2023 expanded the parameter scale to 15 billion with 48 layers, trained on about 65 million non-redundant sequences in UniRef50. Combined with ESMFold, it realized atomic-level structure prediction. In 2025, ESM3 launched by Hayes et al. [5] unified sequence, structure and function tracks for modeling, marking the entry into the multimodal generation era. The key parameters of the three generations are compared in Table 1. Overall, the three generations show a systematic

DOI: 10.54254/2753-8818/2026.38100

progression in data scale, parameter volume and information dimension, moving from simple sequence statistical learning toward joint characterization of structural constraints and functional semantics, providing an increasingly rigorous methodological foundation for generative functional protein design.

Table 1. Comparison of key parameters of ESM series protein language models [2, 3, 5]

Model (Year)	Max Parameters	Training Data Scale	Modeling Information	Representative Capability
ESM-1b (2021)	650 million	UR50/Swiss-Prot ~250 million sequences	Sequence	Emergent structural representation
ESM-2 (2023)	15 billion	UniRef50 ~65 million non-redundant sequences	Sequence	ESMFold atomic-level structure prediction
ESM3 (2025)	98 billion	~2.78 billion proteins (771 billion tokens)	Sequence, Structure, Function	Multimodal controllable generation (esmGFP)

2.2. Multimodal reasoning architecture and generation mechanism of ESM3

The core innovation of ESM3 lies in encoding the sequence, three-dimensional structure and functional annotation of proteins as discrete token tracks separately, and performing joint masked modeling and reasoning through a single Transformer model [5]. Among them, structural information is compressed into structural tokens through the geometry-aware vector quantized variational autoencoder (VQ-VAE), so that the model can process three-dimensional structure in the form of pure sequence tokens. In the generation stage, ESM3 adopts an iterative sampling strategy, starting from partial prompts (such as functional sites or structural fragments) to gradually complete the entire sequence and its structure. The largest version of this model has 98 billion parameters and is trained on about 2.78 billion protein sequence and structure data (771 billion tokens), with strong controllable generation capability.

2.3. Application of ESM3 in de novo protein design: taking esmGFP as an example

The most representative application case of ESM3 is the design of a new fluorescent protein esmGFP [5]. Researchers used the key functional sites of natural green fluorescent protein (GFP) as prompts, and utilized ESM3 to generate completely new candidate sequences. The finally obtained esmGFP has only 58% sequence identity with the closest known natural fluorescent protein, yet maintains normal fluorescence emission function. This sequence difference is equivalent to the divergence distance of about five hundred million years of natural evolution, indicating that ESM3 can cross the long evolutionary path and directly explore the functional sequence space that natural evolution has not reached, which provides a concept verification for the de novo design of functional peptides such as antimicrobial peptides.

2.4. Comparison between ESM3 and other generative protein design models

Compared with existing generative methods, ESM3 has obvious advantages in modal integration capability. AMP-Diffusion proposed by Lin et al. [8] utilizes ESM-2 embeddings combined with latent diffusion model to generate antimicrobial peptides, which is superior to traditional GAN methods in diversity and activity, but can only process sequence information. Zhao et al. [9] combined the diffusion model with ESM-2 for de novo design of functional peptides, and verified

DOI: 10.54254/2753-8818/2026.38100

the activity of the generated peptides through affinity experiments, but similarly only modeled sequence information. LLAMP developed by Bae et al. [10] generates AMPs by fine-tuning large language models, relying on a large amount of annotation data. In contrast, ESM3 unifies sequence, structure and function in the same generation framework, and can impose structural constraints and functional prompts simultaneously during the generation process, which is theoretically more suitable for the design tasks of antimicrobial peptides that are sensitive to spatial conformation.

2.5. Limitations and improvement directions of ESM3 model

ESM3 has several limitations. First, the training data is mainly natural proteins, with insufficient coverage of short antimicrobial peptides and non-natural amino acid sequences, which may affect the structural accuracy of generated peptides. Second, the inference computational cost is high, with limited efficiency in large-scale screening. Third, the wet experimental verification rate of generated sequences is still low, with an obvious gap between computational design and experimental verification. Fourth, the controllability of model generation needs to be strengthened, making it difficult to precisely specify safety indicators such as hemolysis and toxicity. Future improvements include fine-tuning on AMP-specific datasets, cascade screening with activity and toxicity prediction models, and introducing reinforcement learning to optimize the generation direction.

3. Research progress of AI-assisted antimicrobial peptide design and activity prediction

3.1. Application of traditional machine learning in AMP prediction

Early computational prediction of AMPs mainly relied on traditional machine learning methods. Researchers first extracted manual features from amino acid composition, physicochemical properties (such as net charge, hydrophobicity, amphiphilicity) and sequence motifs, and then used algorithms such as support vector machine and random forest to build classification models, which were trained with public databases such as CAMP and APD. Szymczak et al. [11] systematically reviewed the application of such methods in AMP design, and pointed out that traditional machine learning methods show stable performance and strong interpretability on small and medium-sized data sets, but highly rely on manual feature engineering, are difficult to capture long-range dependencies in sequences, and have limited generalization ability when facing imbalanced data. These limitations have promoted the development of end-to-end prediction methods based on deep learning.

3.2. AMP activity prediction models based on deep learning

Deep learning methods, through architectures such as convolutional neural networks, recurrent neural networks and Transformers, can directly learn feature representations from raw sequences, significantly improving the accuracy of AMP activity prediction. Gao et al. [12] systematically evaluated the performance of BERT-like pre-trained models in AMP prediction tasks, and found that predictors based on language model embeddings are generally superior to traditional manual feature methods. Yan et al. [13] systematically reviewed the latest progress of deep learning in AMP identification and design, covering convolutional, recurrent, attention and various fusion architectures. The performance comparison of representative deep learning AMP prediction models is shown in Table 2. The deepAMP developed by Li et al. [6] further unifies prediction and

DOI: 10.54254/2753-8818/2026.38100

generation in the same foundation model, supporting large-scale virtual screening of candidate peptides.

Table 2. Performance comparison of representative deep learning AMP prediction models [12, 13]

Model/Study	Architecture	Task	Accuracy (%)	Sensitivity (%)	Specificity (%)	Other Metrics
AMPlify	BiLSTM+Attention	AMP classification	93.7	92.9	94.5	AUROC 98.4%
Veltri et al.	CNN+LSTM	AMP classification	91.0	89.9	92.1	MCC 0.82, AUROC 96.5%
Su et al.	CNN+DNN fusion	AMP classification	92.2	-	-	-
Ma et al.	CNN/LSTM/BERT ensemble	Human gut microbiome AMP detection	99.6	-	-	Precision 91.3%
sAMP-PFPDeep	VGG-16	Short AMP (<30 aa) classification	84.0	-	-	-
Witten et al.	CNN (classification+MIC regression)	AMP classification and activity regression	97.0	-	-	PCC 77.0%

In addition, Pandi et al. [14] combined deep learning prediction with cell-free protein expression system to realize a fast closed loop of "design-prediction-verification". Through five rounds of "generation-prediction-construction-test" iteration, 30 functional AMPs and 6 broad-spectrum active AMPs were finally obtained from about 500,000 virtual candidates. The design success rate was increased from 0% in the first round to 12.7% in the last round. The optimized AMPs showed strong activity against multiple ESKAPE pathogens, and did not induce significant resistance after continuous passaging for 21 days. These studies show that deep learning has become the mainstream technical route for AMP activity prediction, and provides a reliable activity evaluation tool for generative design.

3.3. Application of generative adversarial networks in de novo AMP design

Generative adversarial networks learn the sequence distribution characteristics of natural antimicrobial peptides through adversarial training. Tucs et al. [15] embedded the activity prediction signal into the GAN training loop, and the inhibition rate of the obtained candidate peptides against Escherichia coli can reach more than twice that of the traditional methods, which verified the strategic value of introducing functional constraints into the generation process. The limitation of this method is that the sequence diversity and training stability still need to be optimized.

3.4. Application of protein language models in AMP design

Protein language models, with their deep understanding of sequence evolutionary laws, are becoming an emerging force in the intelligent design of AMPs. AMP-Diffusion proposed by Lin et al. [8] combines ESM-2 embeddings with diffusion model, realizing the generation from random noise to active AMP sequences. Compared with the GAN method, the diffusion model has advantages in sequence diversity and training stability, but the activity of the generated peptides still needs experimental verification. As a new generation of multimodal model, ESM3 is expected to

DOI: 10.54254/2753-8818/2026.38100

embed structural constraints directly into the generation process, further improving the design quality.

4. Intelligent design and activity prediction of antimicrobial peptides based on ESM3

Based on the multimodal reasoning capability of ESM3 and the progress of AI-assisted AMP research, this paper proposes an AMP intelligent design framework with ESM3 multimodal reasoning as the core and combined with deep learning cascade prediction, covering four links: candidate generation, structural verification, activity prediction and experimental verification. In the candidate generation link, the antimicrobial function annotation and known AMP functional sites are used as prompts, and candidate peptide sequences are generated through ESM3 iterative masked completion sampling [5]. The structural track of ESM3 constrains key conformational characteristics such as amphiphilic helix, and the functional track guides the sequence towards known antimicrobial activity patterns, exerting selection pressure simultaneously at sequence and structural levels. In the structural verification link, ESMFold is used to rapidly predict the three-dimensional conformation of candidate peptides, evaluating their amphiphilic helix content, positive charge distribution and electrostatic complementarity with bacterial membranes [3]. Candidates with reasonable structure enter the next round of screening, while sequences with abnormal conformation or obvious aggregation tendency are eliminated, ensuring that the output molecules are physically foldable and have membrane binding capability [6]. In the activity prediction link, deepAMP is invoked for large-scale virtual screening to obtain activity priors, and then a BERT-like classifier is used for secondary fine screening, complementary to each other to improve prediction reliability [12]. This cascade strategy takes into account both throughput and accuracy, and can quickly converge to high-potential candidates in a ten-thousand-level sequence library. In the comparison link, the candidate peptides generated by ESM3 are systematically compared with the outputs of FBGAN and AMP-Diffusion at three levels: sequence novelty, structural rationality and predicted activity, to evaluate the design preferences and performance boundaries of different generation paradigms, providing candidate priority ranking for subsequent experimental verification [8, 16]. In the experimental verification link, drawing on the "design-prediction-construction-test" cycle model of Pandi et al. [14], the cell-free protein expression system is adopted to achieve rapid synthesis and activity determination, and in vivo efficacy and safety evaluation are carried out on the optimized peptides, forming a complete closed loop from computational design to experimental verification, supporting the transformation of ESM3-based AMP design towards clinical application.

5. Conclusion

This paper systematically reviews the evolution of protein language models from ESM-1b to ESM3, and reviews the research progress of AI-assisted antimicrobial peptide design. ESM3 unifies sequence, structure and function as multimodal token tracks, showing unique advantages in de novo protein design. At the same time, deep learning and generative methods have made significant progress in AMP activity prediction and sequence design. Based on this, this paper proposes an AMP intelligent design framework with ESM3 multimodal reasoning as the core and combined with deep learning cascade prediction, covering the complete closed loop of candidate generation, structural verification, activity prediction and experimental verification. Although ESM3 still faces limitations such as training corpus bias and insufficient experimental verification in AMP design scenarios, with the development of multimodal models and automated experimental platforms, AI-

DOI: 10.54254/2753-8818/2026.38100

driven AMP design is expected to provide a new approach for coping with the antibiotic resistance crisis.

References

[1] Wan, F., Wong, F., Collins, J. J., et al. (2024). Machine learning for antimicrobial peptide identification and design. *Nature Reviews Bioengineering, 2*(5), 392–407. https://doi.org/10.1038/s44222-024-00152-x

[2] Rives, A., Meier, J., Sercu, T., et al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences, 118*(15), e2016239118. https://doi.org/10.1073/pnas.2016239118

[3] Lin, Z., Akin, H., Rao, R., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science, 379*(6637), 1123–1130. https://doi.org/10.1126/science.ade2574

[4] Das, P., Sercu, T., Wadhawan, K., et al. (2021). Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nature Biomedical Engineering, 5*(6), 613–623. https://doi.org/10.1038/s41551-021-00689-x

[5] Hayes, T., Rao, R., Akin, H., et al. (2025). Simulating 500 million years of evolution with a language model. *Science, 387*(6736), 850–858. https://doi.org/10.1126/science.ads0018

[6] Li, T., Ren, X., Luo, X., et al. (2024). A foundation model identifies broad-spectrum antimicrobial peptides against drug-resistant bacterial infection. *Nature Communications, 15*, 7538. https://doi.org/10.1038/s41467-024-51933-2

[7] Meier, J., Rao, R., Verkuil, R., et al. (2021). Language models enable zero-shot prediction of the effects of missense mutations. *bioRxiv*. https://doi.org/10.1101/2021.07.09.450648

[8] Lin, Y., Chen, X., Zhang, M., et al. (2025). AMP-diffusion: Generating antimicrobial peptides with diffusion models. *Cell Biomaterials, 1*(9), 100183. https://doi.org/10.1016/j.celbio.2025.100183

[9] Zhao, L., He, Q., Song, H., et al. (2024). Protein A-like peptide design based on diffusion and ESM2 models. *Molecules, 29*(20), 4965. https://doi.org/10.3390/molecules29204965

[10] Bae, D., Kim, M., Seo, J., et al. (2025). AI-guided discovery and optimization of antimicrobial peptides through species-aware language model. *Briefings in Bioinformatics, 26*(4), bbaf343. https://doi.org/10.1093/bib/bbaf343

[11] Szymczak, P., & Szczurek, E. (2023). Artificial intelligence-driven antimicrobial peptide discovery. *Current Opinion in Structural Biology, 83*, 102733. https://doi.org/10.1016/j.sbi.2023.102733

[12] Gao, W., Zhao, J., Gui, J., et al. (2024). Comprehensive assessment of BERT-based methods for predicting antimicrobial peptides. *Journal of Chemical Information and Modeling, 64*(19), 7772–7785. https://doi.org/10.1021/acs.jcim.4c00507

[13] Yan, J., Cai, J., Zhang, B., et al. (2022). Recent progress in the discovery and design of antimicrobial peptides using traditional machine learning and deep learning. *Antibiotics, 11*(10), 1451. https://doi.org/10.3390/antibiotics11101451

[14] Pandi, A., Koch, H., Lassek, M., et al. (2023). Cell-free biosynthesis combined with deep learning accelerates de novo development of antimicrobial peptides. *Nature Communications, 14*, 7197. https://doi.org/10.1038/s41467-023-42434-9

[15] Tucs, A., Tran, D. P., Yumoto, A., et al. (2020). Generating ampicillin-level antimicrobial peptides with activity-aware generative adversarial networks. *ACS Omega, 5*(36), 22847–22851. https://doi.org/10.1021/acsomega.0c02088

[16] Zervou, M. A., Doutsi, E., Pantazis, Y., et al. (2024). De novo antimicrobial peptide design with feedback generative adversarial networks. *International Journal of Molecular Sciences, 25*(10), 5506. https://doi.org/10.3390/ijms25105506