Abstract citation ID: bbaf631.027

DISCOVERY AND OPTIMIZATION OF ANTIMICROBIAL PEPTIDES FROM EXTREME ENVIRONMENTS ON GLOBAL SCALE

Zixin Kang[1,‡], *Haohong Zhang*[1,‡], *Kang Ning*[1,*]

[1]MOE Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular-imaging, Center of AI Biology, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

[*]Correspondence author

[‡]These authors contributed equally to this work

Background: Antimicrobial peptides (AMPs) inhibit microbial growth through membrane disruption or interference with intracellular processes, offering promising solutions to antimicrobial resistance (AMR). While AMPs have been extensively identified and verified from animal proteomes, reference microbial genomes and host environments, those from extreme habitats remain largely unexplored. Microbes in such extreme niches with low-level of nutrients evolve unique membrane modification and specialized metabolic pathway, representing a hidden reservoir for novel AMP discovery.

Methods: We developed Atlantis, a language model-driven framework for AMP detection and optimization from global extremophile (Fig. 1). It combined two modules: Atlantis-Prospect, which predict antimicrobial potency based on both sequence and structure information, and Atlantis-Search, which conducts iterative single-mutations with structural constraints to broaden peptides' antimicrobial activity from narrow to broad spectrum. The pretrained structure-aware language model was fine-tuned on small proteins from extremophile to capture biome-specific evolutionary information.

Results: By mining 60,461 extremophile metagenomes from nine discrete habitats across multiple geographical scales, we identified 1662 non-redundant peptides with potential antimicrobial activity from 85 bacterial and archaeal phyla, none of which match existing databases. We synthesized fourteen peptides, and two of them showed stronger antimicrobial activity then commercial AMP, LL-37, in vitro against *Escherichia coli*, *Pseudomonas aeruginosa* and *Staphylococcus aureus* (Fig. 2).

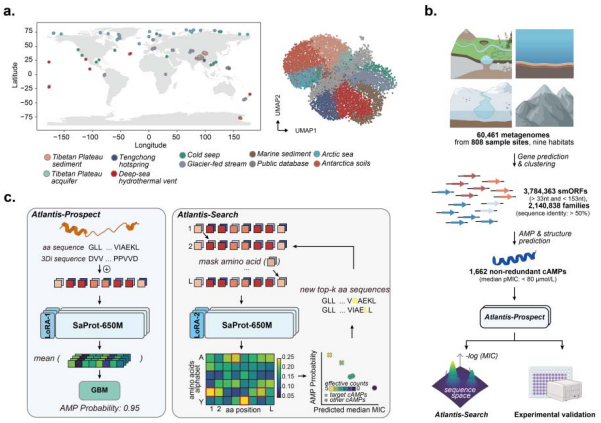

Figure 1. Overview of Atlantis. a Geographic distribution of 808 sampling sites across nine habitates: Tibetan Plateau acquifer, Tibetan Plateau sediment, Arctic sea, Antarctic soil, Glacier-fed streames, Tengchong hotspring, Deep cold seep, Deep-sea hydrothermal vent and Marine sediments. **b** To build Atlantis, (i) 60,461 metagenomes were collected. (ii) A total of 3,784,363 smORFs (33 bp ~153 bp) were predicted and clustered into 2,140,838 families at a 50% identity cutoffs. (iii) Antimicrobial activity was predicted using APEX, and peptides showing broad-spectrum activity with a median MIC ≤ 80 μmol/L were retained as candidate AMPs (cAMPs), yielding 1,662 non-redudant candidates. e Workflow of Atlantis-Prospect: (iv) To incoporate structure information, ESMFold was used to predict peptide structure, which were then tokenized into 3Di sequences with Foldseek. Combined amino acid and 3Di sequences were embedded using a fine-tuned SaProt-650M model. A Gradient Boosting Model (GBM) was then applied to estimate AMP probability. Workflow of Atlantis-Search: (v) To explore the entire fitness ladscape, residues at each position were systematically masked and predicted while preserving complete structure information. Coupled with beam search strategy, iteractive single-mutation enabled cAMPs to evolve from narrow-spectrum to broad-spectrum antimicrobial activity. The cAMPs were further validated by wet-lab experiments.

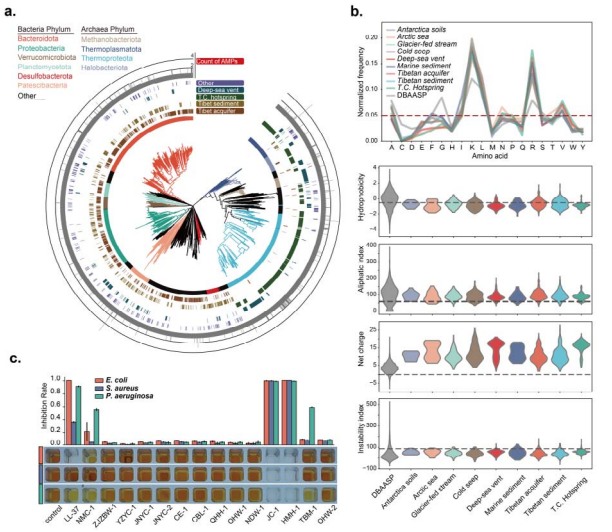

Figure 2. Extremophile AMPs identified by Atlantis. a The phylogenetic tree of extremophile metagenomes (n = 3,085) containing cAMPs, based on the placement of each MAG in the GTDB-Tk reference tree. Branch colors denote phyla, while the outer rings indicate biome type and cAMP abundance for each species. **b** Amino acid frequency and physicochemical properties of cAMPs from different extreme biomes compared with known AMPs from DBAASP database. (i) Normalized amino acid frequencies, showing an enrichment of basic residues K and R. (ii) Hydrophobicity and (iii) Aliphatic index, which influence the peptide-lipid interactions in membrane bilayers. (iv) Net charge, which influences the peptide's initial electrostatic interactions with negatively charged bacterial membranes. (v) Instability index, reflecting the peptide's stability resistance to in vivo degradation. The dashed line represents the mean value in each box. **c** The inhibitation rate of the extremophiles AMPs (64 μmol/L) against three clinically revelant pathogens. 106 bacterial cells were incubated at 37 °C. Bacterial cultures (10^6 cells) were incubated at 37 °C, and growth was assessed by optical density at 450 nm after 5 hours of treatment.