Generating antimicrobial peptides using generative adversarial networks.

Khondamir Rustamov (kh.r.rustamov@mail.ru)
Fergana State University, Fergana city, Fergana region, Republic of Uzbekistan.

Purpose.
The aim of this study is to investigate the performance of unrolled generative adversarial networks (GAN) in generating antimicrobial peptides (AMP).

Material and methods.
Dataset construction. We collected sequences of antimicrobial peptides from DBAASP [1] and APD [2] databases. Sequences containing more than 32 residues were filtered. After removing duplicates between DBAASP and APD sequences we obtained a total of 13030 sequences. Then each amino acid in sequences was encoded to a 5-bit vector.
GAN model. The structure of Generator and Discriminator models is shown in Figure 1. To escape mode, collapse we used an unrolled GAN with different unrolling steps (K=0, 3, 5), where K=0 model is a simple GAN without unrolling steps. We trained our model for 50 epochs (batch size: 1024; loss: Binary Crossentropy; optimization: Adam with learning rate 1e-4).
Estimated antimicrobial activity. We generated 5000 sequences and estimated the probability that they will feature antimicrobial activity by applying $CAMP_{R3}$ predictive models [3]. We used AlphaFold 2 [4] to predict the 3D structure of generated peptides and calculated their physicochemical features by using modlAMP 4.3 [5].
Molecular dynamics simulations. We performed the molecular dynamics (MD) simulations of the top-1 peptide with gram-negative inner membrane. Membrane-peptide system was generated in CHARMM-GUI [6]. The lipid bilayer contained POPE, POPG and TMCL1 lipids in the proportion of 70:25:5 (E. coli inner membrane). Then the system was neutralized in 0.15 mM KCl and solvated in water. MD simulation was performed in GROMACS [7]. The minimization and equilibration of the system were performed following the 6-step CHARMM-GUI protocol. The system was run for 10 ns of the production run.

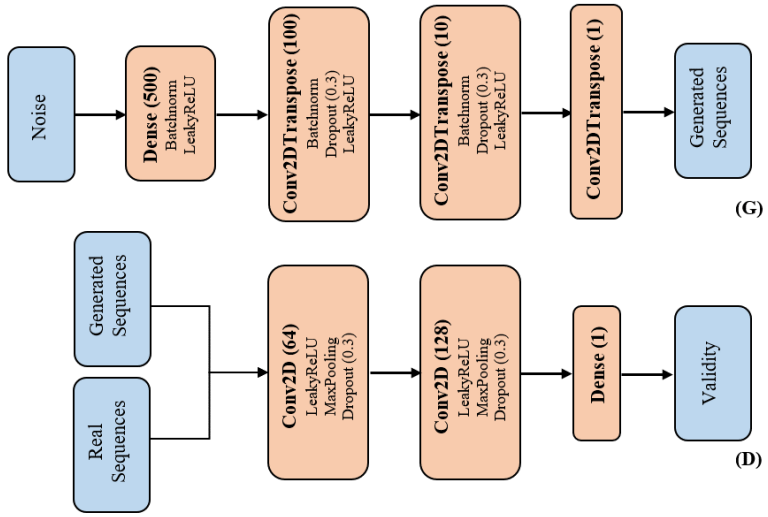

Figure 1. The architectures of Generator (G) and Discriminator (D) models.

	Unique	SVM	RF	ANN	DA
K = 0	4480	30.2901%	18.9509%	40.0223%	4.4866%
K = 3	4981	43.8206%	60.7027%	82.8756%	34.5920%
K = 5	5000	52.8293%	99.1672%	99.8932%	48.1528%

Table 1. The percentage of generated sequences being predicted as AMP. Unique is the number of unique generated sequences.

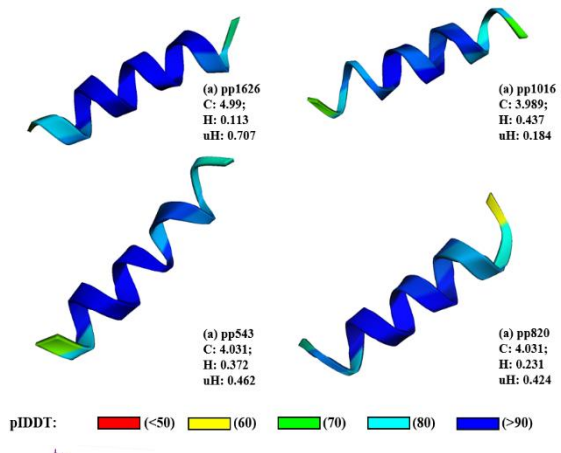

Figure 2. 3D structures and physicochemical attributes for top-4 sequences generated by K=5 model. **C, H, uH** correspond to charge, hydrophobicity and hydrophobic moment. **pIDDT** – AlphaFold model's confident score. The average physicochemical attributes for real AMPs used in this study are following: **C:** 4.78±2.84; **H:** -0.078±0.42; **uH:** 0.41±0.23

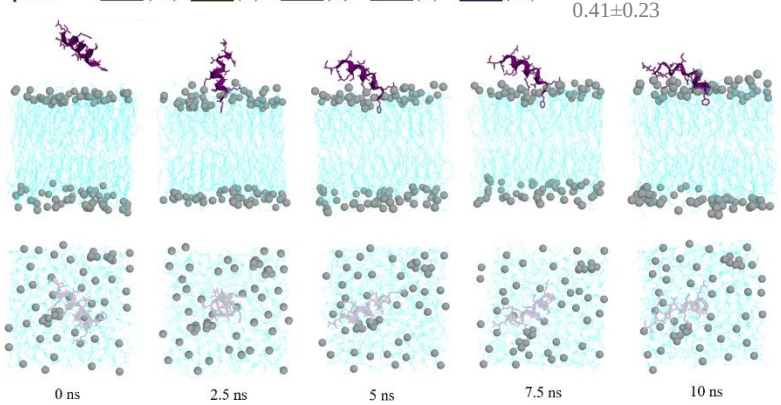

Figure 3. Representative snapshots depicting peptide-bilayer interactions (side view and top view). The membrane is shown in lines (cyan) and the peptide is shown in helical structure (purple). Phosphate atoms of the membrane are shown in gray color (only phosphate atoms of the upper leaflet are shown in the top view)

Results.
The percentage of generated sequences that were predicted as being antimicrobial for all models is demonstrated in table 1. The secondary structures of top-4 sequences generated by unrolled GAN with K=5 steps are illustrated in Figure 2. The results of MD simulation are shown in Figure 3.

Conclusion.
In this study, we estimated the performance of GANs with a different number of unrolled steps in generating antimicrobial peptides. Unrolled GANs were described as a tool for escaping mode collapse (the situation when the model generated only ranged number of similar outputs or even one output), and in our work it's obvious that a model with K=5 steps generated 100% unique sequences and greatly reduces the chance of mode collapse. However, 5 unrolling steps also improve the total score of the model in generating sequences with a high probability of being antimicrobial. MD simulations demonstrate that the top-1 peptide generated by our GAN is able to penetrate the inner membrane of gram-negative bacteria. However, 10 ns run is not enough to study peptide-membrane interactions and future MD research is required.

References.
1. Pirtskhalava, M.; Gabrielian, A.; Cruz, P.; Griggs, H. L.; Squires, R. B.; Hurt, D. E.; Grigolava, M.; Chubinidze, M.; Gogoladze, G.; Vishnepolsky, B.; et al. DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. Nucleic Acids Res. 2016, 44, D1104−D1112.
2. Wang, Z.; Wang, G. APD: the antimicrobial peptide database. Nucleic Acids Res. 2004, 32, D590−D592.
3. Waghu FH, Barai RS, Gurung P, Idicula-Thomas S. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. Nucleic Acids Res. 2015.
4. Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).
5. Müller A. T. et al. (2017) modlAMP: Python for antimicrobial peptides, Bioinformatics 33, (17), 2753-2755.
6. S. Jo, T. Kim, V.G. Iyer, and W. Im (2008) CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. J. Compute. Chem. 29:1859-1865.
7. Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. . SoftwareX 1–2:19-25.