|  |
| --- |
| **Comparative Study on Protein Language Models for Antimicrobial Peptide Classification**  **Elias Georgoulis**1,2,#, **Michaela Areti Zervou**3,4 and **Yannis Pantazis**1,\*  1Institute of Applied and Computational Mathematics, FORTH, Heraklion, Greece  2Department of Mathematics & Applied Mathematics, University of Crete, Heraklion, Greece  3Department of Computer Science, University of Crete, Heraklion, Greece  4Institute of Computer Science, FORTH, Heraklion, Greece  # Presenting author: Elias Georgoulis, email: math1p0001@math.uoc.gr  \* Corresponding author: Yannis Pantazis, email: pantazis@iacm.forth.gr |

**Abstract**

Deep learning has become the most powerful driving force in computational protein engineering, offering solutions to numerous challenges through both discriminative and generative neural-based models. The traditional machine learning approach involves data collection and curation, followed by model training, tuning, and evaluation. Recently, a new paradigm known as self-supervised learning has emerged, harnessing vast amounts of data and computational power to create foundation models. In this context, we compare various protein language models (PLMs) [1, 2], which are inspired by large language models. Using Antimicrobial Peptide (AMP) classification performance as an evaluation metric, our key findings are as follows: (a) Model scale is crucial, with classification performance improving as model size increases; (b) State-of-the-art results are achieved with minimal effort by using embedding representations alongside shallow classifiers; and (c) Classification performance is further enhanced through the cost-effective fine-tuning of PLMs [3].

**References**

[1] Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, Gibbs T, Feher T, Angerer C, Steinegger M, Bhowmik D, Rost B. 2022. IEEE transactions on pattern analysis machine intelligence, 44, 7112–7127

[2] Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, Dos Santos Costa A, Fazel-Zarandi M, Sercu T, Candido S, Rives A. 2023. Science 379, 1123-1130

[3] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen. 2021. ArXiv, abs/2106.09685