

UNDERSTANDING PROTEIN LANGUAGE THROUGH NATURAL LANGUAGE MODELS

ESM3 is an advanced language model that simulates protein evolution on a timescale equivalent to 500 million years (1). This model, based on artificial intelligence technologies, generates innovative protein sequences and structures far removed from those found in nature.

The model was used to design a variant of the green fluorescent protein (GFP) with 58% similarity to known natural proteins.

ESM3 can accelerate evolutionary discoveries by creating new and functional proteins more quickly than natural processes.

1. <https://www.science.org/doi/10.1126/science.ads0018>