

ESM

Github: <https://github.com/facebookresearch/esm>

Models to Consider:

- ESM-2: <https://www.biorxiv.org/content/10.1101/2022.12.21.521521v1>
 - **Purpose:** General purpose
 - **Description:** ESM-2 is designed to predict structure, function, and other protein properties directly from individual protein sequences. It is the best single-sequence protein language models, outperforming predecessors across a range of structure prediction tasks.
- ESMFold: <https://www.biorxiv.org/content/10.1101/2022.12.21.521521v1>
 - **Purpose:** End-to-end single sequence 3D structure prediction.
 - **Description:** Leverages the ESM-2 language model to generate accurate structural predictions from a protein sequence without the need for multiple sequence alignments. It's particularly useful for predicting 3D structures directly and efficiently.
- ESM-MSA-1b: <https://www.biorxiv.org/content/10.1101/2021.02.12.430858v2>
 - **Purpose:** Structure inference from multiple sequence alignments.
 - **Description:** This model uses multiple sequence alignments (MSAs) to enhance the prediction of protein structures. It extracts embeddings from an MSA, enabling superior inference capabilities for structural prediction.
- ESM-1v: <https://www.biorxiv.org/content/10.1101/2021.07.09.450648v2>
 - **Purpose:** Prediction of variant effects.
 - **Description:** Specialized for zero-shot prediction of the functional effects of sequence variations. This model can predict how changes in the protein sequence (mutations) might affect protein function, without the need for retraining the model on specific mutation data.
- ESM-IF1: <https://www.biorxiv.org/content/10.1101/2022.04.10.487779v2>
 - **Purpose:** Inverse folding and sequence design for specified structures.
 - **Description:** Aimed at designing protein sequences that fold into predefined 3D structures. This model can predict sequences that are likely to fold into a given structure, or it can be used to understand the functional effects of sequence variations on a fixed backbone structure.

ESM Metagenomic Atlas:

- Link: <https://esmatlas.com>
- README: <https://github.com/facebookresearch/esm/blob/main/scripts/atlas/README.md>
- Database of 700 million metagenomic protein structures predicted by ESMFold

Original Rives Paper (ESM-1):

- **UIUC Lecture:**
 - Model was trained by passing in protein sequences with randomly *masked* characters
 - Parameters are tuned to minimize *loss function*
 - Proteins lengths in training: 100-500/600
 - Zero-Shot: Handling tasks that were *not* seen during training