

# DeepProZyme User Guide

## 1 What is DeepProZyme?

DeepProZyme is a web-based platform for protein function prediction. It supports sequence-level deep learning tools, including:

- DeepTFactor: Predicts whether a protein sequence encodes transcription factor (TF).
- DeepECtransformer: Assigns enzyme commission (EC) numbers for enzymatic function annotation.

Both tools operate directly on amino acid sequences and require only FASTA-formatted input.

Importantly, both tools support the **simultaneous submission of multiple protein sequences**, enabling efficient large-scale functional screening and high-throughput analysis.

## 2 How to run analyses with DeepProZyme tools

Job name

Mail address (optional)

Protein sequence (FASTA format)

```
>MDH_ECOLI
MKVAVLGAAGGIGQALALLLKTQLPSGSELSLYDIAPVTPGVAVDLSHIPTAVKIKGFSG
EDATPALEGADVVLISAGVARKPGMDRSDLFNVNAGIVKNLVQQVAKTCKACIGIITNP
VNTTVAIAAEVLKAGVYDKNKLFVGTLDIIRSNTFVAELKKGKQPGEVEVPVIGGHSV
TILPLLSQVPGVVSFTEQEVADLTAKRIQNAQTEVVEAKAGGGSATLSMGQAAARFGLSLVR
ALQGEQGVVECAVVEGDGQYARFFSQPLLLGKNGVEERKSIGTLSAFEQNALEGMLDTK
KDIALGEEFVNK
```

Do you want to get sequence logos ? (Currently, only one sequence is supported for logo generation.)

Yes  No

[Download an example input](#)

**Step 1. Enter your job name.**

**Step 2. (Optional) Enter your email address.**

You may provide an email address to receive updates when your analysis:

- begins,
- finishes successfully, or
- encounters an error.

**Step 3. Provide your protein sequence (FASTA format).**

You may paste the sequence directly into the input box or upload a FASTA file. A sample FASTA file is available for reference to ensure correct formatting.

**Step 4. Submit your analysis.**

Once all required fields are completed, click Submit. Your analysis will be placed in the processing queue.

- For DeepECtransformer, you may additionally request **sequence logo generation** when a single protein sequence is entered.
- The amino acid sequence length that can be analysed by DeepECtransformer is limited to **2,500 residues per protein sequence**, whereas DeepTFactor supports the analysis of protein sequences up to **1,000 residues per sequence**.

## 3 How to interpret the results

### DeepTFactor

Output file: DeepTFactor\_1.csv

Column	Meaning
Sequence ID	Identifier for the input protein sequence
Prediction	Predicted as Transcription Factor (TF) or non-TF
Score	Prediction score

### DeepECtransformer

Output file: DeepECtransformer.csv

Column	Meaning
Sequence ID	Identifier for the input protein
Prediction	Predicts EC number
Score	Prediction score. If DeepECtransformer cannot assign an EC number, the sequence is listed as unassigned, as no DIAMOND-based annotation is added in such cases.