

DeepRES: Deep learning enables reaction-based comprehensive enzyme screening

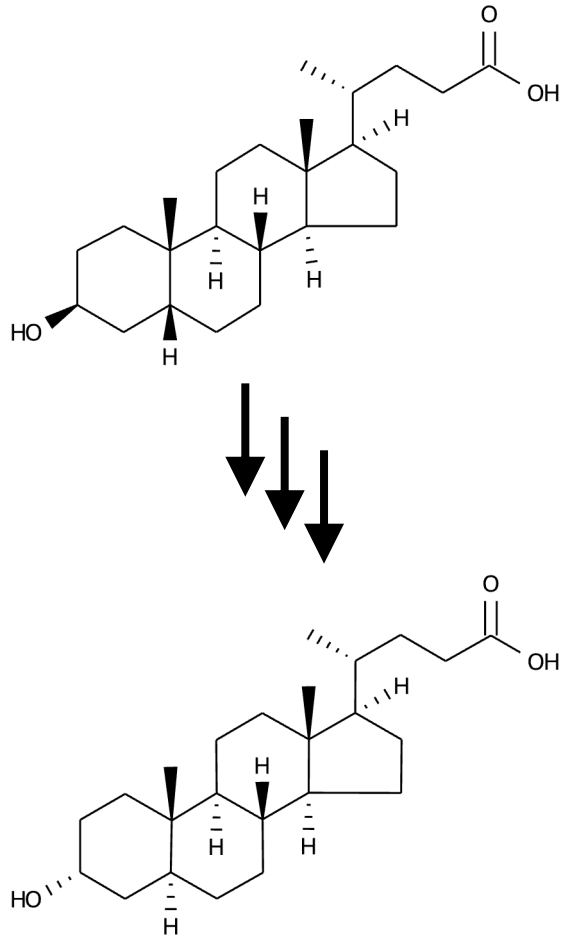
第5回 バイオ生成 AI 研究会
東京科学大学 山田研究室 博士1年
廣田 佳亮

Table of contents

- 1. Introduction**
2. Development of DeepRES
3. Exploration of orphan enzyme candidates
4. Conclusion

Metabolic pathway

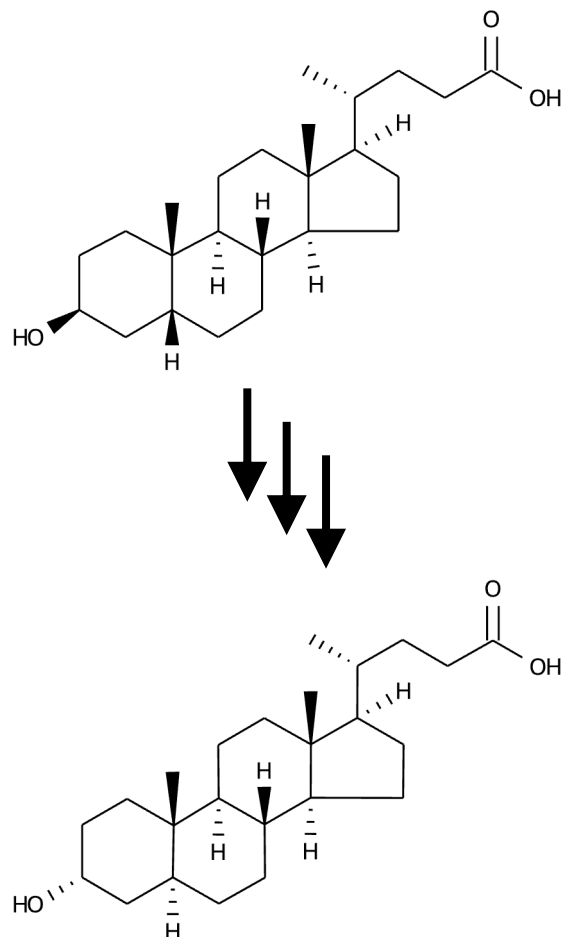
AlloLCA biosynthesis



- Metabolic pathway is a series of biochemical reactions
- Metabolic pathway DB (ex. KEGG ^[1]) helps representation and analysis of biological systems

Orphan enzymes in pathway databases

AlloLCA biosynthesis

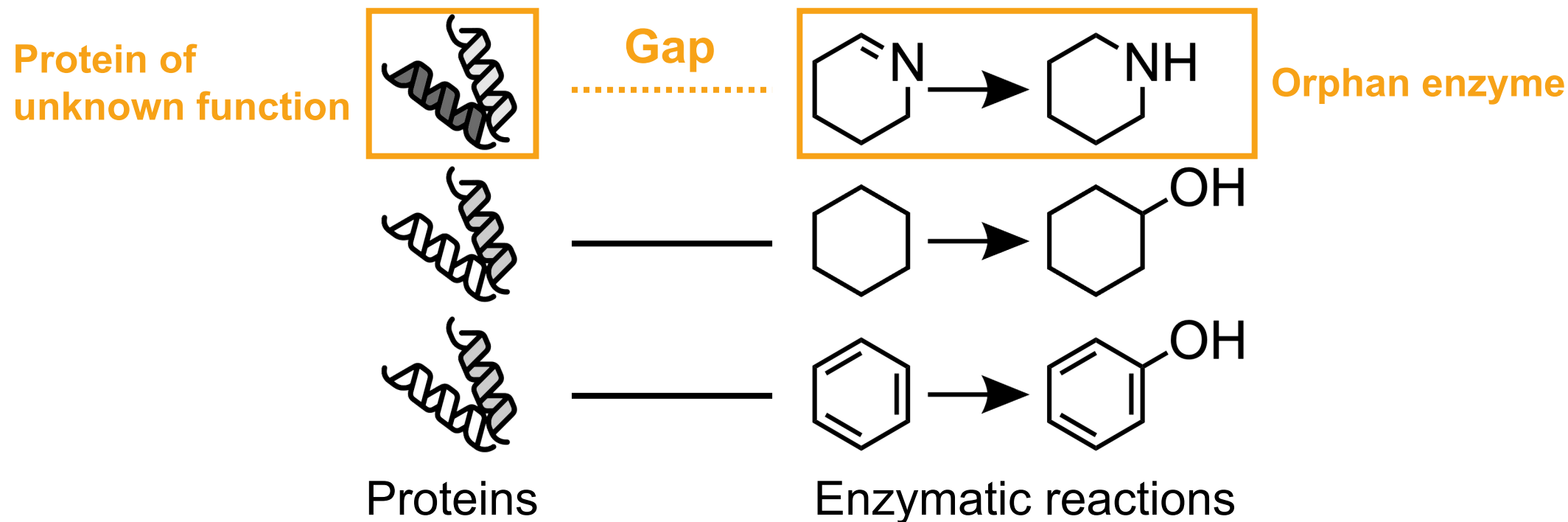


- Metabolic pathway is a series of biochemical reactions
- Metabolic pathway DB (ex. KEGG ^[1]) helps representation and analysis of biological systems
- **20-50%** of the enzymatic reactions in pathway DBs lack sequence information ^[2,3]
= Orphan enzymes ^[4]

[1] Kanehisa M et al. *Nucleic Acids Res.* 2025 [2] Shiroma H et al. *Brief. Bioinform.* 2024

[3] Hirota K et al. *Bioinformatics* 2025 [4] Lespinet O et al. *Science* 2005

Gap between proteins and enzymatic reactions



Due to this gap between known proteins and enzymatic reactions, **various proteins of unknown function are expected to be orphan enzymes.**

Existing computational tools cannot bridge this gap



Exploration space: Proteins of unknown function Orphan reactions

Enzyme Classifier
(ex. CLEAN ^[1])



Enzyme Predictor
(ex. E-zyme2.0 ^[2])



Proposed Method

DeepRES

Objectives >> link proteins of unknown function with orphan enzymes

Deep learning enables Reaction-based Enzyme Screening

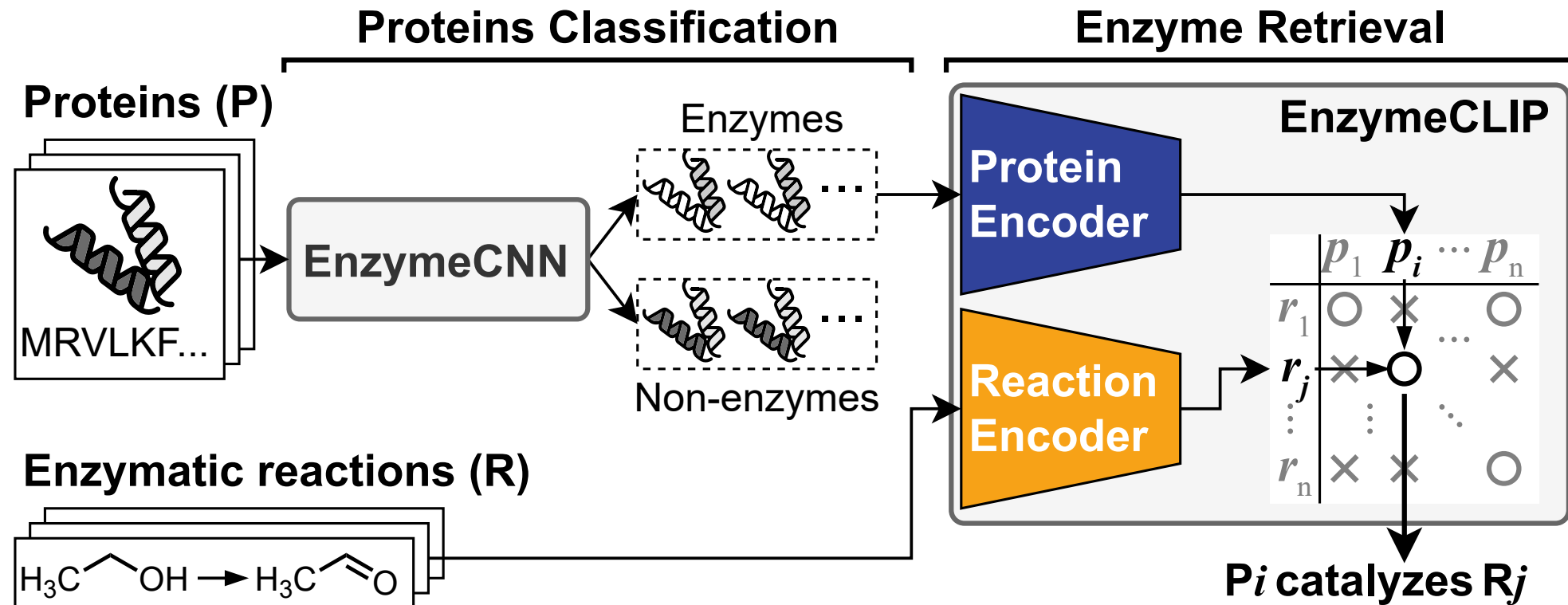


Table of contents

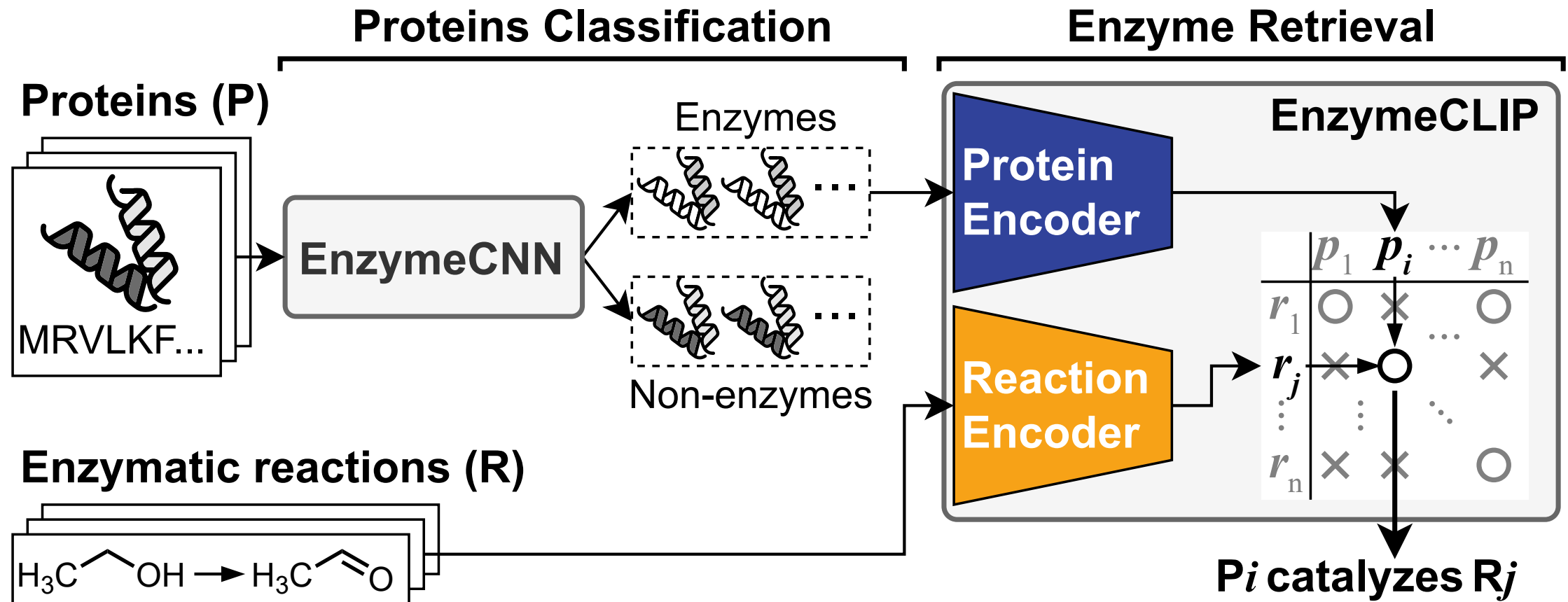
1. Introduction

2. Development of DeepRES

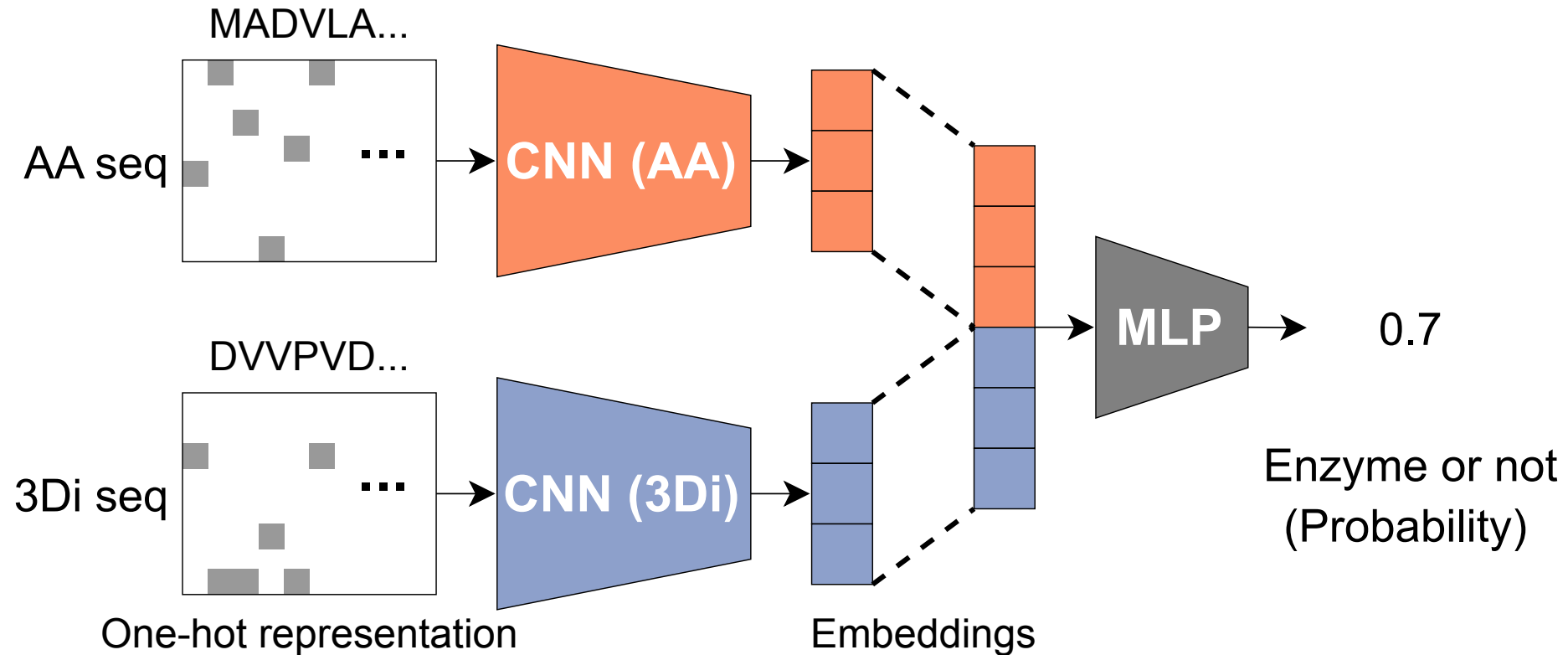
3. Exploration of orphan enzyme candidates

4. Conclusion

DeepRES overview



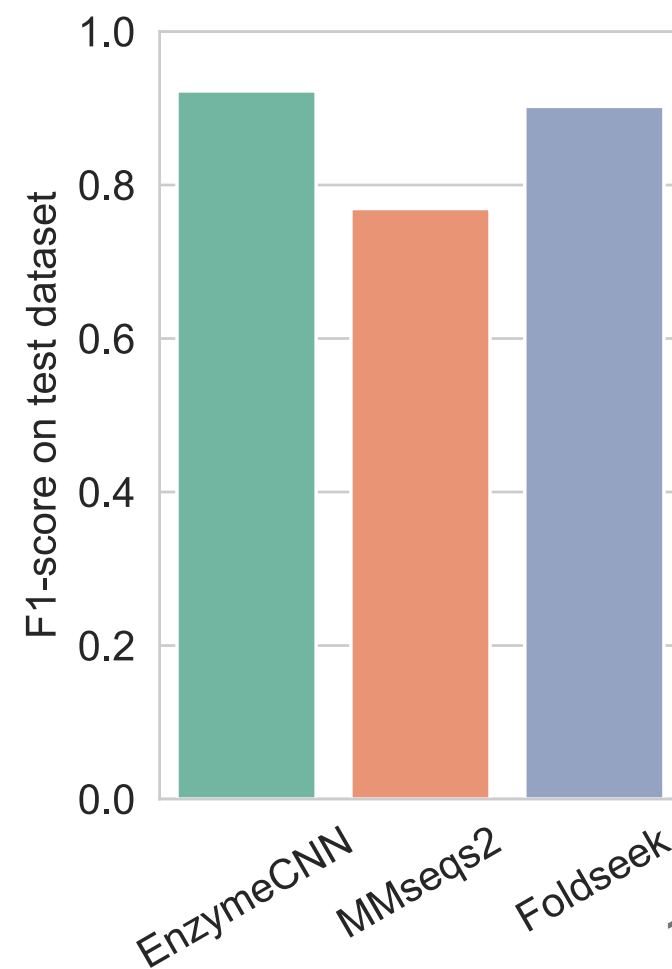
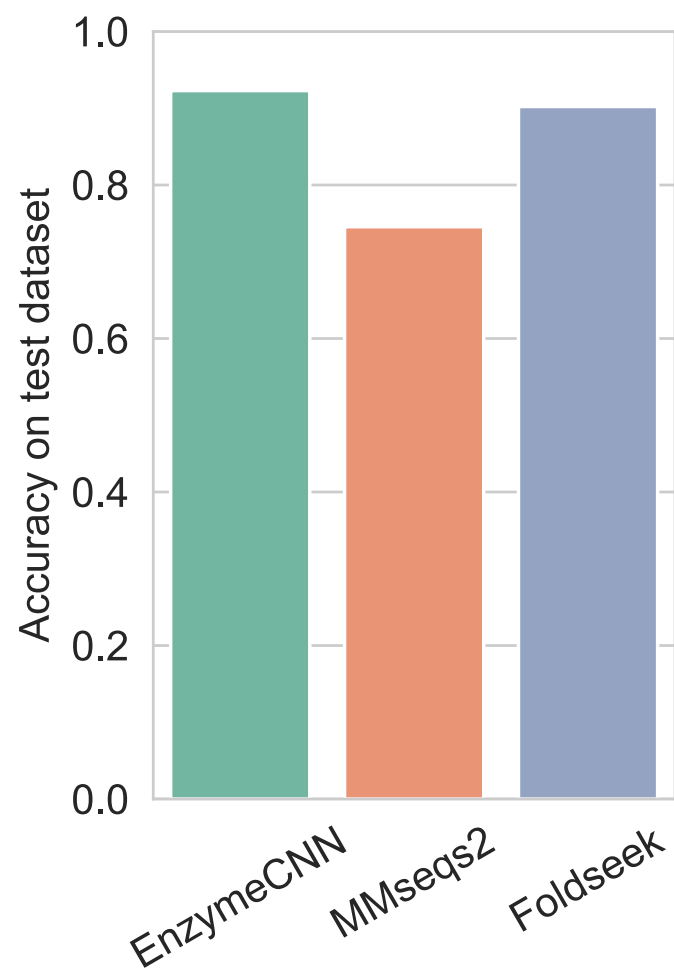
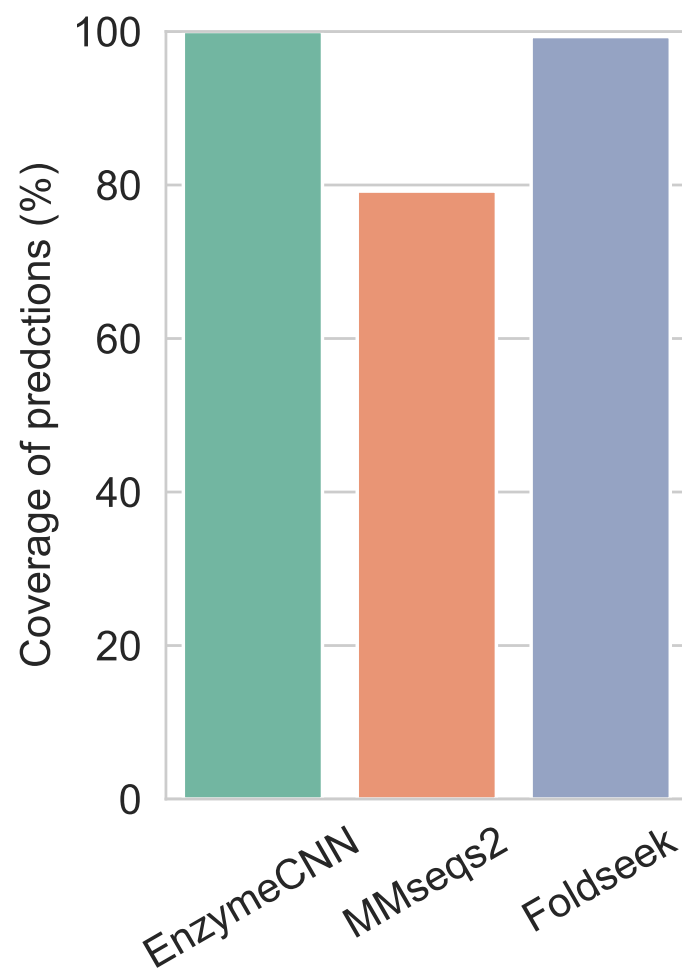
Protein classification via EnzymeCNN



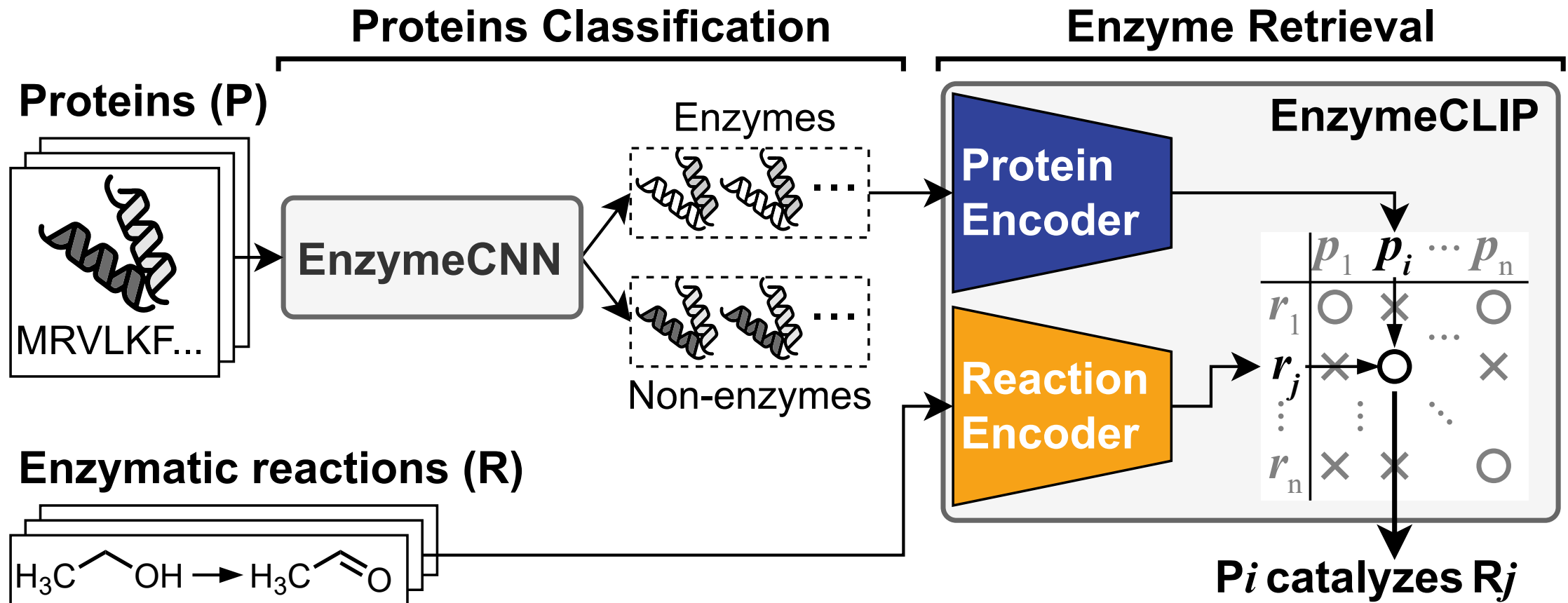
EnzymeCNN outperformed existing tools

Task: Classification of enzymes and non-enzymes

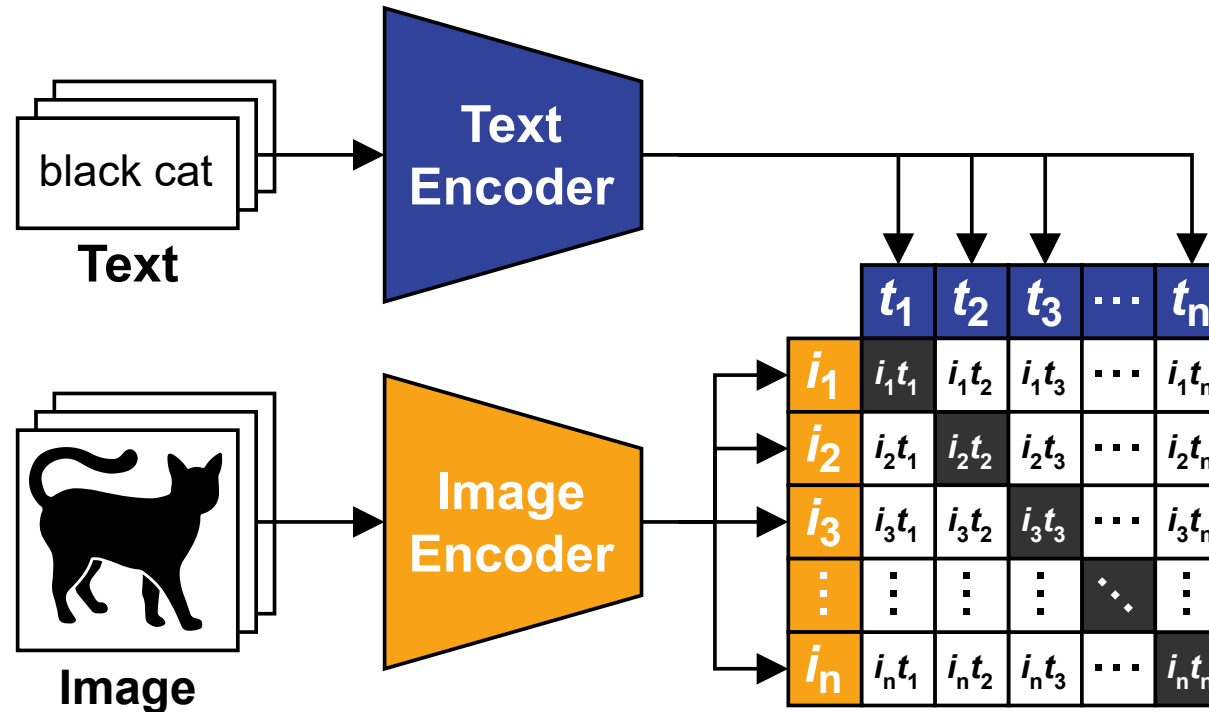
Dataset: Swiss-Prot



DeepRES overview

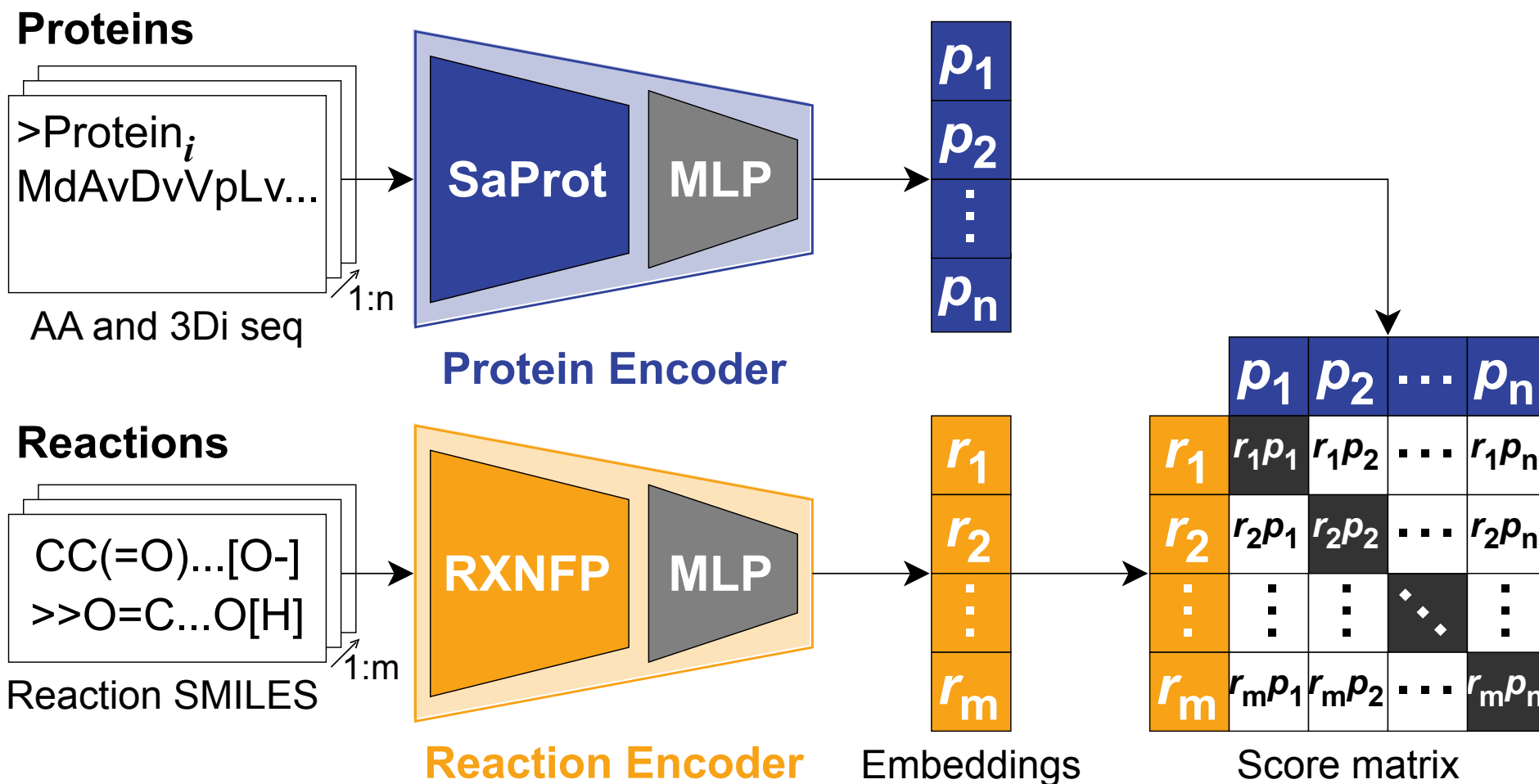


Contrastive Language-Image Pre-training (CLIP) [1]



- CLIP strategy is **applicable to any multimodal data**
- CLIP strategy enables **zero-shot prediction** by encoding label information

Enzyme-Reaction retrieval via EnzymeCLIP



SaProt ^[1]: structure-aware PLM

RXNFP ^[2]: transformer-based Reaction SMILES encoder

Structural information improve EnzymeCLIP performance

Task: Enzyme screening

Dataset: Swiss-Prot + Rhea

Enrichment Factor

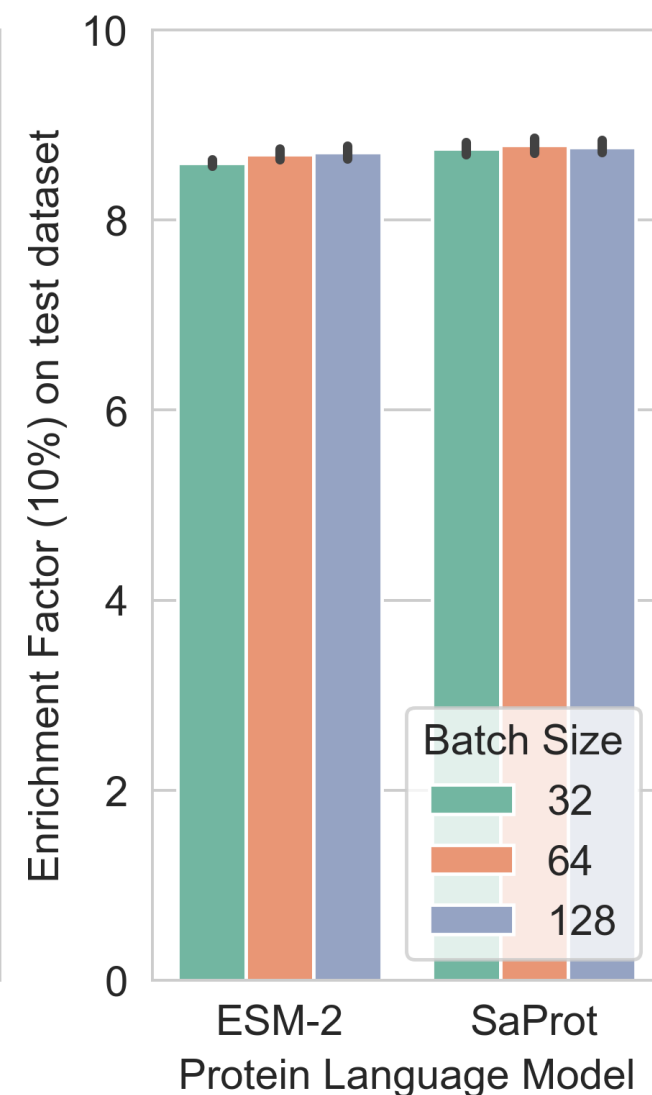
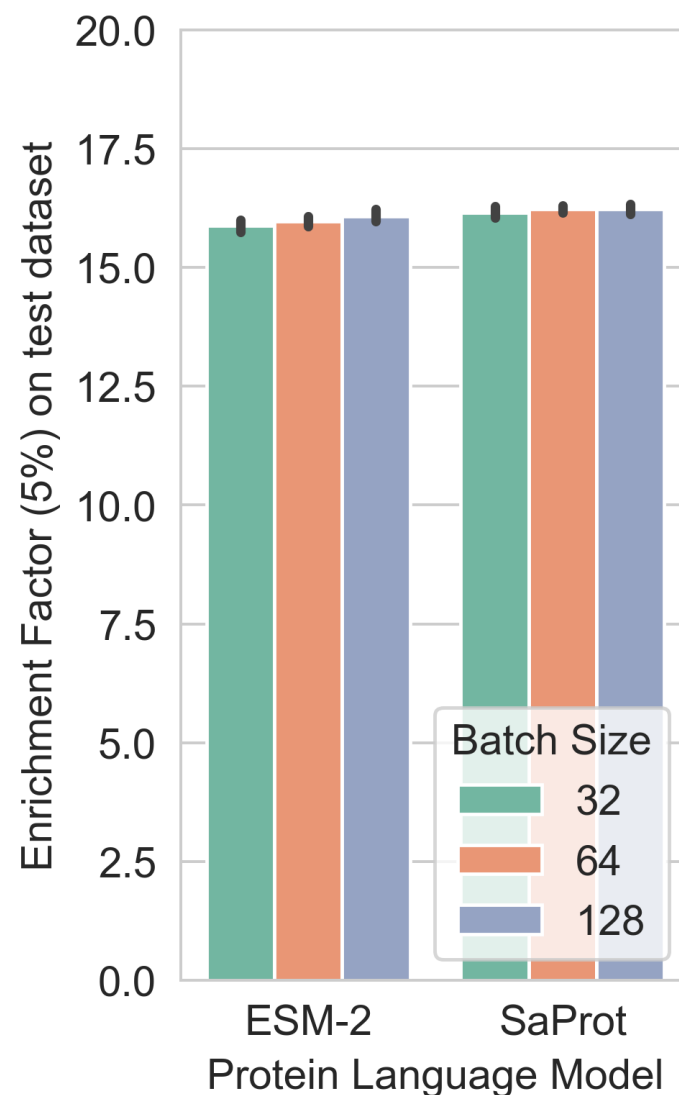
$$EF_{\chi} = \frac{n_{\text{pos}}/(\chi N)}{N_{\text{pos}}/N}$$

$$1 \leq EF_{\chi} \leq \frac{1}{\chi}$$

Bad  Good

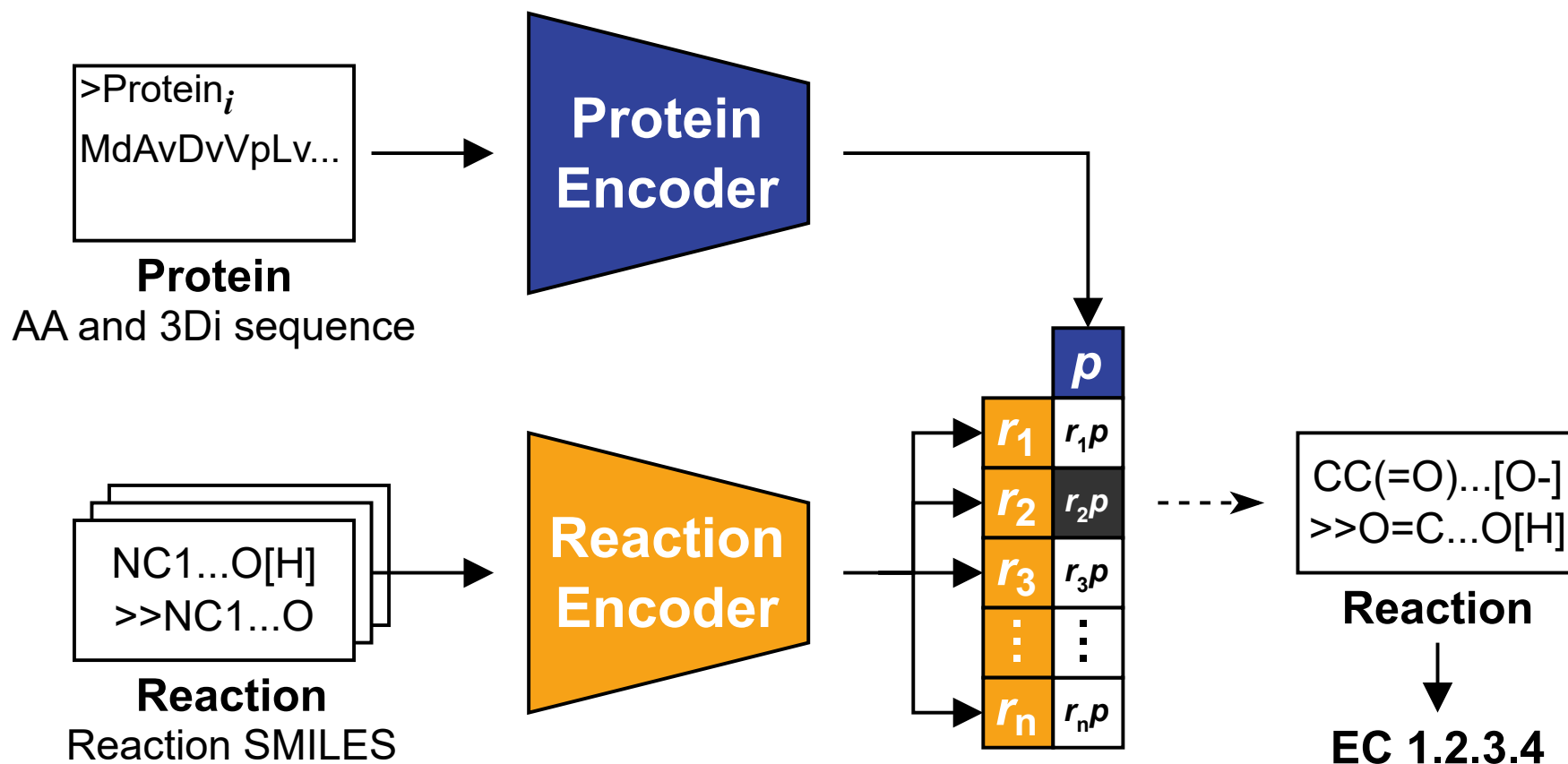
SaProt ^[1]: structure-aware PLM

ESM-2 ^[2]: transformer-based PLM



Performance comparison with CLEAN [1]

EnzymeCLIP was compared with CLEAN, the state-of-the-art EC number prediction method, as the EC number prediction task.



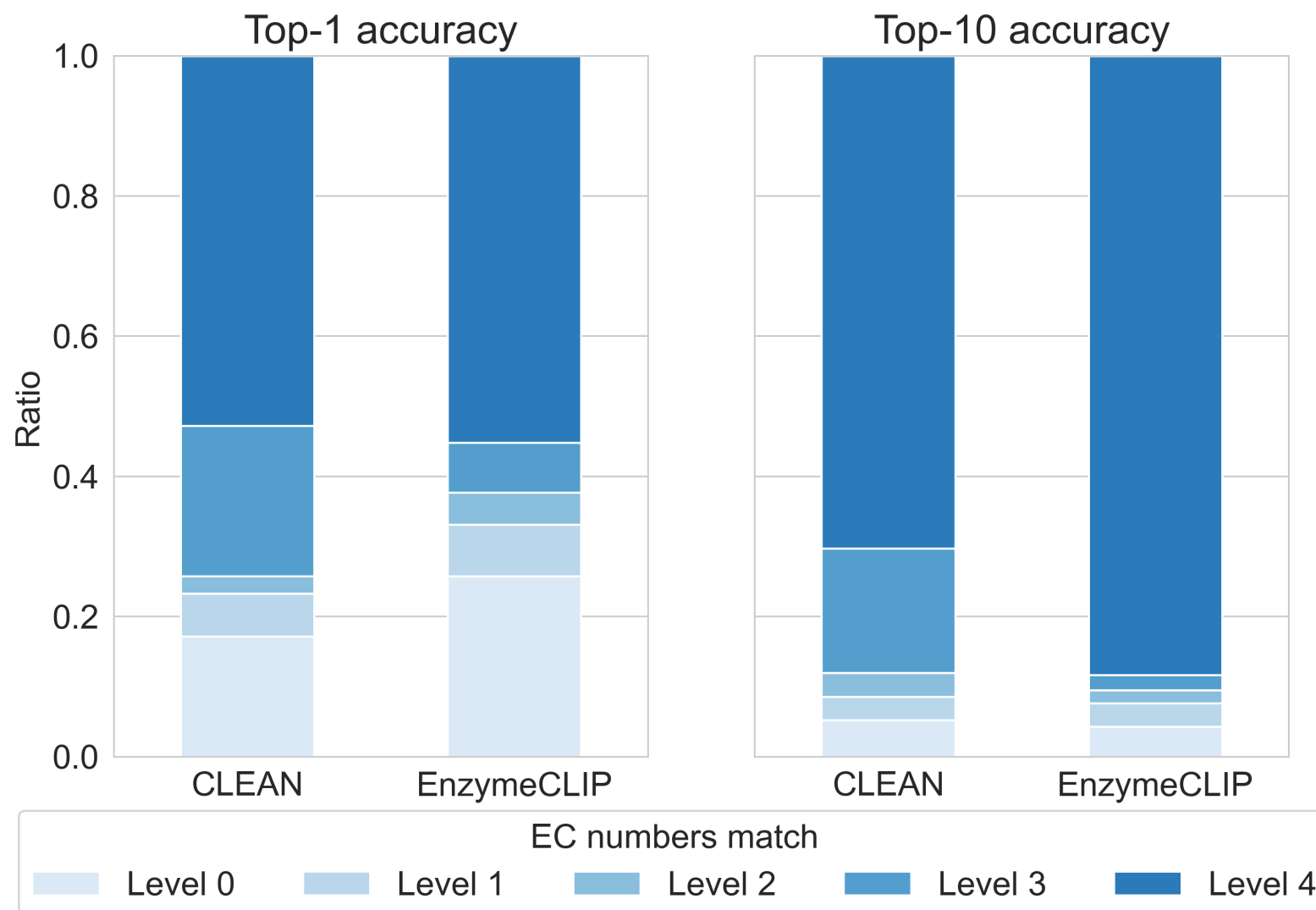
EnzymeCLIP outperformed CLEAN [1]

Label: EC 1.2.3.4

Pred: EC 1.2.1.1



Level 2 match !



Development of DeepRES

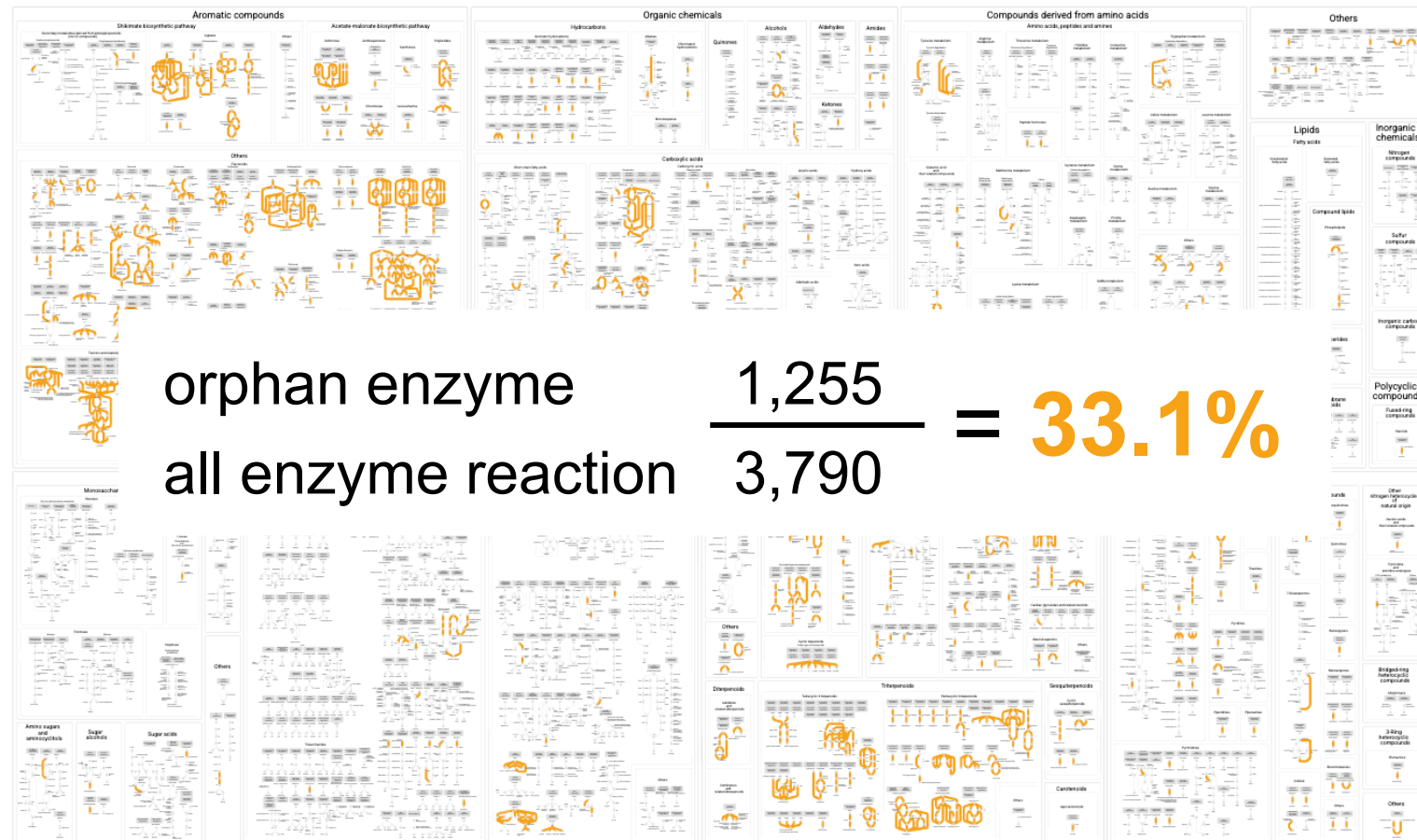
- DeepRES comprises two deep learning models; EnzymeCNN and EnzymeCLIP.
- EnzymeCNN outperformed MMseqs2 and Foldseek.
- EnzymeCLIP performed well and outperformed CLEAN, the state-of-the-art EC number prediction tool.

Table of contents

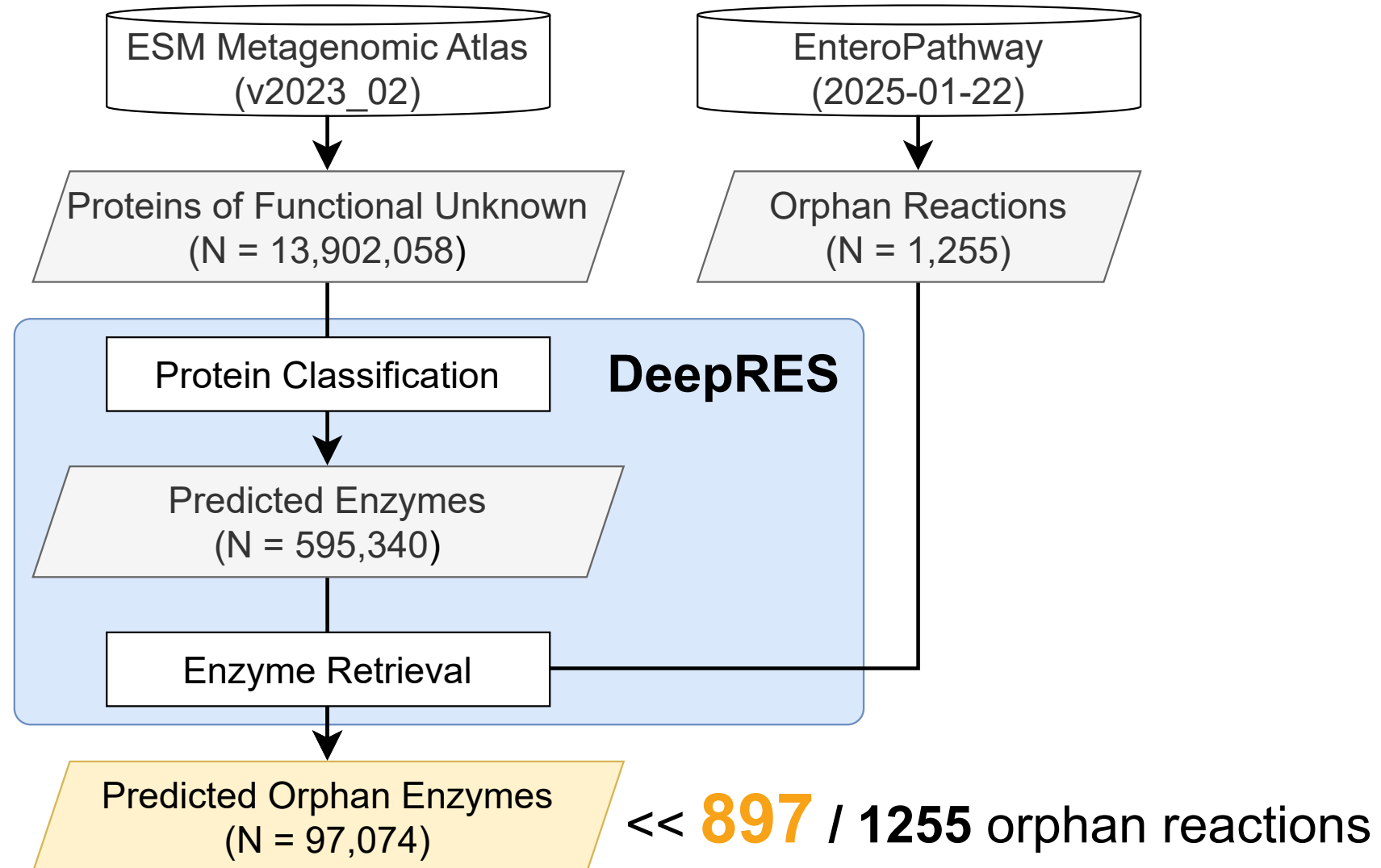
1. Introduction
2. Development of DeepRES
- 3. Exploration of orphan enzyme candidates**
4. Conclusion

EnteroPathway [1]

- Metabolic pathway database for the human gut microbiota

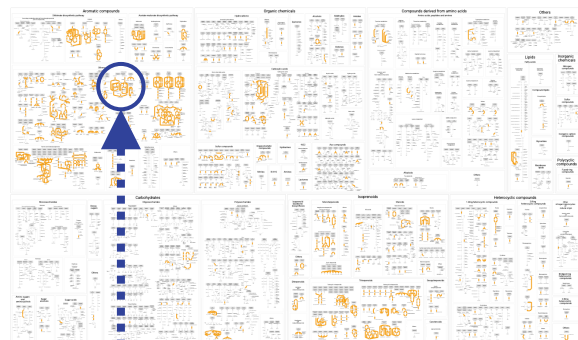


Exploration of orphan enzyme candidates



Orphan enzymes identified by DeepRES

EnteroPathway



Anthocyanin degradation

- EPM0520

- EPM0522

These affect host various diseases ^[1,2].

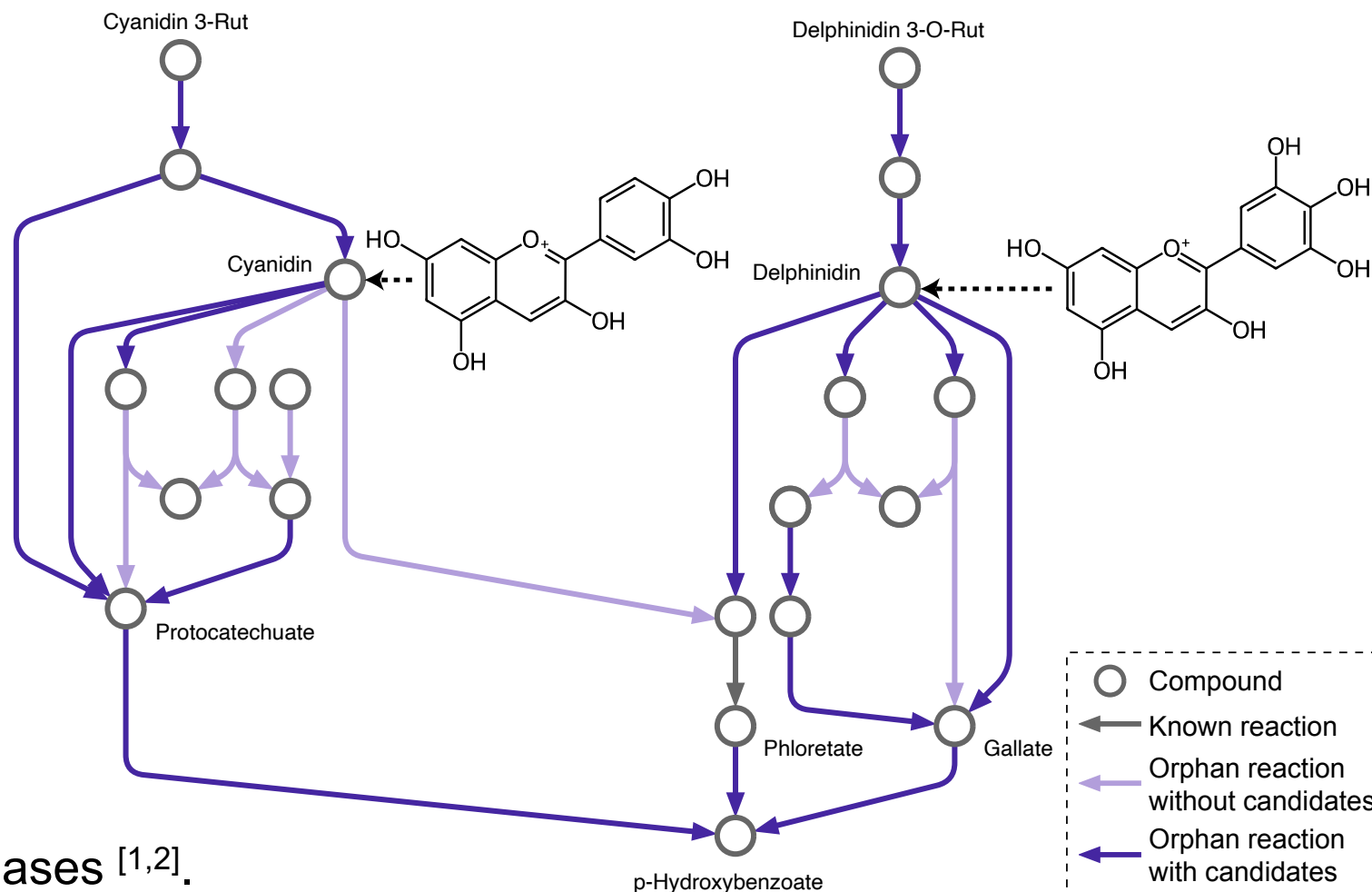


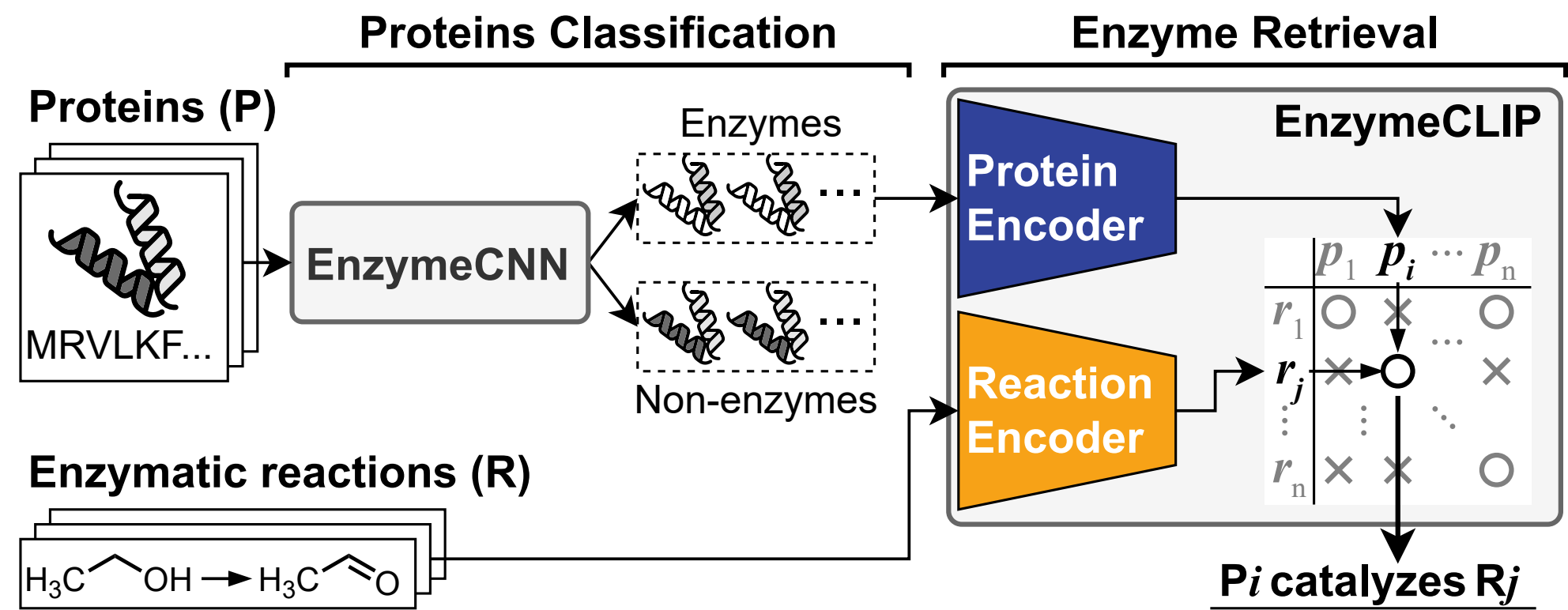
Table of contents

1. Introduction
2. Development of DeepRES
3. Exploration of orphan enzyme candidates
- 4. Conclusion**

DeepRES

Objectives >> link proteins of unknown function with orphan enzymes

Deep learning enables Reaction-based Enzyme Screening



Summary

- This study developed DeepRES, the AI-based framework to predict catalytic availability for any protein-reaction pair.
 - DeepRES consists of EnzymeCNN and EnzymeCLIP.
- EnzymeCNN and EnzymeCLIP outperformed existing tools.
- DeepRES annotated 97,074 proteins of unknown function with 897 orphan reactions involved in human gut bacteria metabolism.

DOI: 10.1101/2025.07.28.667344

GitHub: <https://github.com/yamada-lab/DeepRES>

Acknowledgement

- This study was carried out using the TSUBAME4.0 supercomputer at Institute of Science Tokyo

Funding

- JSPS KAKENHI Grant Number 16H06279 (PAGS)
- JST SPRING (JPMJSP2180)

