CLEAN-Contact: Contrastive Learning-enabled Enzyme Functional Annotation Prediction with Structural Inference

Yuxin Yang[1,2,3], Abby Jerger[4], Song Feng[4], Zixu Wang[3], Margaret S. Cheung[4], Jeremy Zucker[5*], Qiang Guan[3*]

[1]Cleveland Clinic Genome Center, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH, USA, 44195.
[2]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH, USA, 44195.
[3*]Department of Computer Science, Kent State University, 800 E Summit St, Kent, OH, USA, 44242.
[4]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, 1100 Dexter Ave N, Seattle, WA, USA, 98109.
[5*]Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA, USA, 99354.

*Corresponding author(s). E-mail(s): jeremy.zucker@pnnl.gov; qguan@kent.edu;
Contributing authors: yangy9@ccf.org; abby.jerger@pnnl.gov; zwang55@kent.edu;

Abstract

Recent years have witnessed the remarkable progress of deep learning within the realm of scientific disciplines, yielding a wealth of promising outcomes. A prominent challenge within this domain has been the task of predicting enzyme function, a complex problem that has seen the development of numerous computational methods, particularly those rooted in deep learning techniques. However, the majority of these methods have primarily focused on either amino acid sequence data or protein structure data, neglecting the potential synergy of combining of both modalities. To address this gap, we propose a novel **C**ontrastive **L**earning framework for **E**nzyme functional **AN**notation prediction combined with protein amino acid sequences and **Contact** maps (CLEAN-Contact). We rigorously evaluated the performance of our CLEAN-Contact framework against

the state-of-the-art enzyme function prediction model using multiple benchmark datasets. Our findings convincingly demonstrate the substantial superiority of our CLEAN-Contact framework, marking a significant step forward in enzyme function prediction accuracy.

Keywords: deep learning, contrastive learning, enzyme function prediction

1 Introduction

The crucial role of enzyme function annotation in our understanding of the intricate mechanisms driving biological processes governed by enzymes is widely recognized. The Enzyme Commission (EC) number, a numerical classification system commonly used for enzyme function, is a widely recognized standard in these efforts. The depth of insights provided by the EC number ranges from broad categories of enzyme mechanisms to detailed biochemical reactions through its four hierarchical layers of digits. Traditionally, sequence similarity-based methods, such as the basic local alignment search tool for protein (BLASTP) [1] and HH-suite [2], were largely relied upon for annotating EC numbers. However, in recent times, the deep learning revolution has largely solved the protein structure prediction problem, and it is natural to ask how these twin scientific advances can aid in enzyme function prediction [3].

Predicting enzyme function is not merely an academic classification exercise; it holds immense practical value in systems biology and metabolic engineering, particularly in the construction of genome-scale metabolic models. Such predictive capabilities streamline the process of automating the curation of these models by improving the ability to predict which proteins are responsible for observed growth phenotypes under diverse nutrient conditions and distinct genetic backgrounds. Furthermore, precise knowledge of a genome's metabolic capabilities enables the design of microbial cell factories to fit for purpose to the metabolic engineering goal, be it medicine, biomanufacturing, or bioremediation.

Currently, the majority of deep learning-based models developed for predicting EC numbers focus on either amino acid sequence or structural data of the enzyme. Studies such as that of [4], for example, rely solely on amino acid sequence data for EC number prediction. Conversely, the work in [5] employs a single convolutional neural network to predict EC numbers, focusing more on the structural data. Additionally, a different approach integrating enzyme structure data into the training process is seen in [6], which provides a more comprehensive view of the enzyme's functionality. The recent addition of a novel method grounded on contrastive learning has also escalated the performance of EC number prediction [7].

Building upon this groundwork, we propose CLEAN-Contact, a contrastive learning framework that amalgamates both amino acid sequence data and protein structure data for superior enzyme function prediction. Our CLEAN-Contact framework has shown a notable improvement over the current state-of-the-art, CLEAN [7], under a variety of test conditions, further emphasizing the potential for combining protein sequence and structure-based deep learning in enzyme function predictive practices.

2 Results

2.1 Contrastive Learning Framework for Enzyme Functional Annotation Prediction Combined with Protein Contact Maps

We develop a deep learning framework aimed at predicting EC numbers. This framework integrates a protein language model (ESM-2) [8] and a computer vision model (ResNet50) [9] (See Section 4). ESM-2 operates as a feedforward neural network, extracting function-aware sequence representations from input protein amino acid sequences. Meanwhile, ResNet50 functions as a feedforward neural network, extracting structure representations from input 2D contact maps derived from protein structures. The CLEAN-Contact framework plays a key role by combining these sequence and structure representations and employing contrastive learning to learn the prediction of EC numbers. The schematic illustration of the CLEAN-Contact framework is shown in Fig. 1. The CLEAN-Contact framework consists of three key components: (1) The representations extraction segment (Fig. 1a), which is designed to extract structure representations from contact maps using ResNet50 [9] and sequence representations from amino acid sequences using ESM2 [8]. (2) The contrastive learning segment (Fig. 1b), which first employs a fully connected projector to combine structure and sequence representations. Subsequently, the combined representations are projected to a high-dimensional space to minimize the distances between enzymes sharing the same EC number while maximizing the distances between enzymes with different EC numbers. And (3) the EC number prediction segment (Fig. 1c), which is responsible for determining the EC number of a query enzyme based on the projector learned through contrastive learning.

2.2 Benchmark results

We conducted comprehensive evaluations of our proposed CLEAN-Contact framework, comparing it against two state-of-the-art EC number prediction models, namely CLEAN [7] and DeepECtransformer [10]. These models, along with the CLEAN-Contact, underwent testing on two independent test datasets. The first test dataset, New-392 [7] contains 392 enzyme sequences distributed over 177 different EC numbers. Predictive performance was assessed using metrics such as Precision, Recall, F1-score, and Area Under Receiver Operating Characteristic Curve (AUC). The CLEAN-Contact demonstrated superior performance, as illustrated in Fig. 2a. Specifically, CLEAN-Contact exhibited a 16.22% enhancement in Precision (0.652 vs. 0.561), a 9.04% improvement in Recall (0.555 vs. 0.509), a 12.30% increase in F1-score (0.566 vs. 0.504), and a 3.19% elevation in AUC (0.777 vs. 0.753) over CLEAN. Conversely, DeepECtransformer recorded the lowest performance metrics, achieving a Precision of 0.519, a Recall of 0.477, an F1-score of 0.465, and an AUC of 0.572.

The second test dataset, Price-149 [7] comprises 149 enzyme sequences distributed over 56 different EC numbers. Once more, as depicted in Fig. 2b, CLEAN-Contact exhibited superior performance. Specifically, CLEAN-Contact showcased enhancements across various metrics compared to CLEAN: a 16.95% improvement in Precision

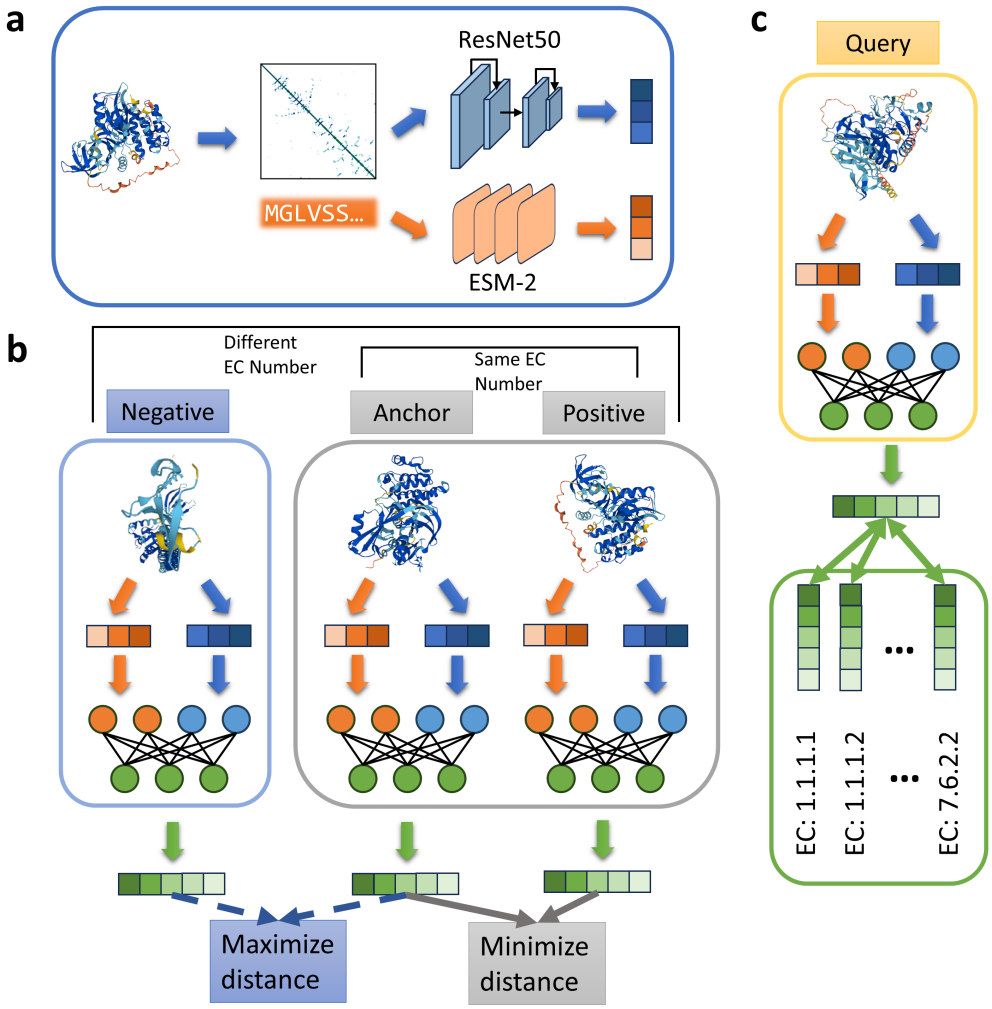

Fig. 1 Schematic illustration of the CLEAN-Contact framework. (a) Obtaining structure and sequence representations from contact map and amino acid sequence using ResNet50 and ESM2, respectively. (b) The contrastive learning segment. Sequence and structure representations are combined and projected into high-dimensional vectors using the projector. Positive samples are those with the same EC number as the anchor sample and negative samples are chosen from EC numbers with cluster centers close to the anchor. We perform contrastive learning to minimize distances between anchor and positive samples, and maximize distances between anchor and negative samples. (c) The EC number prediction segment. Cluster centers are computed for each EC number by averaging learned vectors within that EC number. Euclidean distances between the query enzyme's vector and the cluster centers are calculated to predict the EC number of a query enzyme. P-value EC number selection algorithm is used to determine predicted EC numbers.

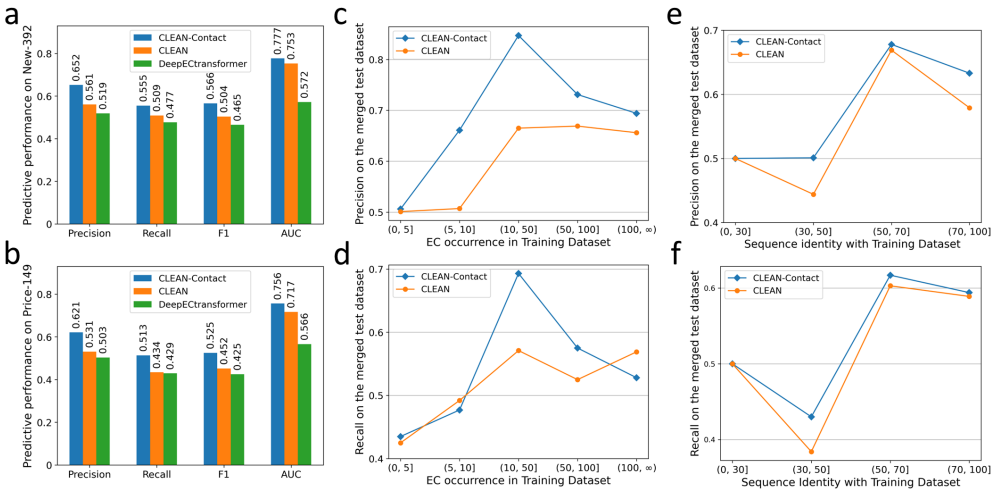

Fig. 2 Assessment of predictive performance between CLEAN-Contact and baseline models on two benchmark test datasets. (a) Evaluation of predictive performance, measured by Precision, Recall, F1 score, and AUC metrics, comparing CLEAN-Contact, CLEAN, and DeepECtransformer on the New-392 test dataset. (b) Evaluation of predictive performance, measured by Precision, Recall, F1 score, and AUC metrics, comparing CLEAN-Contact, CLEAN, and DeepECtransformer on the Price-149 test dataset. (c) Evaluation of predictive performance based on Precision across the merged test dataset, correlating with the frequency of occurrence of EC numbers in the training dataset. (d) Evaluation of predictive performance based on Recall across the merged test dataset, correlating with the frequency of occurrence of EC numbers in the training dataset. (e) Evaluation of predictive performance based on Precision across the merged test dataset, correlating with the maximum sequence identity in comparison to the training dataset. And (f) Evaluation of predictive performance based on Recall across the merged test dataset, correlating with the maximum sequence identity in comparison to the training dataset.

(0.621 vs. 0.531), an 18.20% increase in Recall (0.513 vs. 0.434), a 16.15% boost in F1-score (0.525 vs. 0.452), and a 5.44% elevation in AUC (0.756 vs. 0.717). Conversely, DeepECtransformer recorded the lowest performance metrics, achieving a Precision of 0.503, a Recall of 0.429, an F1-score of 0.425, and an AUC of 0.566.

To investigate the performance of CLEAN-Contact concerning understudied EC numbers, we merged the two test datasets and divided the merged test dataset based on the frequency of an EC number's occurrence in the training dataset. Subsequently, we measured the Precision and Recall of both CLEAN-Contact and CLEAN on these divided test datasets. As shown in Fig. 2c and Fig. 2d, CLEAN-Contact demonstrated a 30.4% improvement (0.661 vs. 0.507) in Precision while achieving comparable performance against CLEAN when the EC number was rare in the training dataset (occurring more than 5 times but less than 11 times) and a 27.4% improvement (0.847 vs. 0.665) in Precision and a 21.4% improvement (0.693 vs. 0.571) in Recall compared to CLEAN when moderately infrequent (occurring more than 10 times but less than 51 times). However, when the EC number was extremely rare in the training dataset (occurring less than 6 times) or very common (occurring more than 100 times), the improvement of CLEAN-Contact over CLEAN was less significant (0.506 vs. 0.501 in

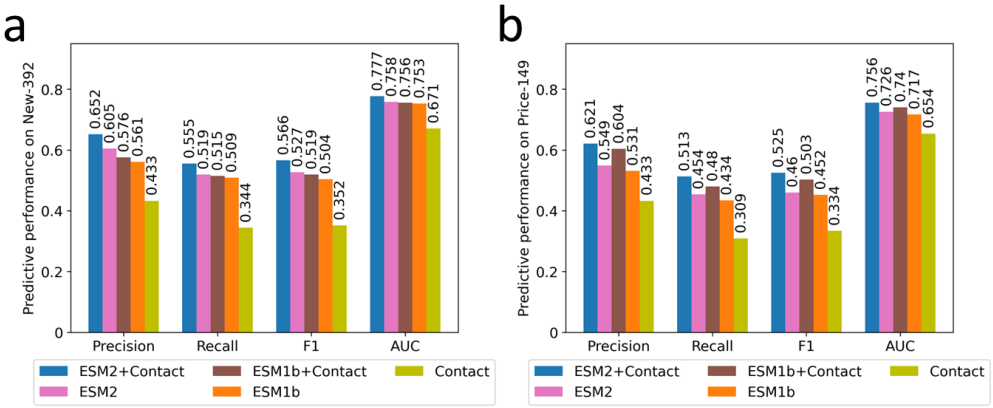

Fig. 3 Ablation results. We conducted a comparative analysis of several model variations, including our proposed CLEAN-Contact (ESM2+Contact), CLEAN-Contact without structure representation (ESM2), CLEAN-Contact with ESM-1b instead of ESM-2 (ESM1b+Contact), the original CLEAN (ESM1b), and a model with solely structure representation devoid of ESM-2 or ESM-1b generated sequence representation (Contact). (a) Evaluation of predictive performance metrics (Precision, Recall, F-1 score, and AUC) across different models on the New-392 test dataset. (b) Evaluation of predictive performance metrics (Precision, Recall, F-1 score, and AUC) across different models on the Price-149 test dataset.

Precision and 0.435 vs. 0.425 in Recall for EC numbers occurring less than 6 times, 0.731 vs. 0.669 in Precision and 0.575 vs. 0.525 in Recall for EC numbers occurring more than 50 times but less than 101 times, and 0.694 vs. 0.656 in Precision and 0.528 vs. 0.569 in Recall for EC numbers occurring more than 100 times). These findings underscore the significant enhancement in predictive performance achieved by integrating structural information into CLEAN-Contact.

Furthermore, we divided the merged test dataset based on the maximum sequence identity with the training dataset and evaluated the Precision and Recall of both CLEAN-Contact and CLEAN on these divided test datasets. As illustrated in Fig. 2e and Fig. 2f, when the sequence identity with the training dataset was very low (less than 30%), both models achieved comparable performance. However, as the maximum sequence identity ranged from 30% to 50%, CLEAN-Contact exhibited a 12.3% improvement in Precision (0.501 vs 0.446) and a 12.0% improvement in Recall (0.430 vs. 0.384) over CLEAN. As the maximum sequence identity increased to between 50% and 70%, CLEAN-Contact still maintained a 1.34% advantage in Precision (0.678 vs. 0.669) and a 2.32% advantage in Recall (0.617 vs. 0.603) over CLEAN. Remarkably, when the maximum sequence identity exceeded 70%, CLEAN-Contact showcased a 9.33% improvement over CLEAN (0.633 vs. 0.579) in Precision and a comparable performance in Recall (0.594 vs. 0.589).

2.3 Interpreting the neural network

We investigated the impact of various representation components, specifically comparing sequence representations derived from ESM-2 against those derived from ESM-1b,

along with including structure representations. As depicted in Fig. 3, our observations indicate that replacing ESM-1b with ESM-2, without incorporating structure representations, led to a marginal 1.39% average performance improvement across the two test datasets. However, integrating structure representations while retaining ESM-1b yielded a substantial 5.76% average performance increase across the two test datasets. Moreover, replacing ESM-1b with ESM-2 and including structure representations resulted in a 6.13% average performance improvement. Notably, utilizing solely structure representations as the model input yielded the poorest performance. We attributed this outcome to the fact that contact maps only offer information about residue contacts within the protein structure, lacking crucial details about the amino acids.

3 Discussion

In this work, we introduce a novel framework based on contrastive learning, which integrates both enzyme amino acid sequence and structural data to predict EC numbers. Our proposed CLEAN-Contact framework harnesses the power of ESM-2, a pretrained protein language model responsible for encoding amino acid sequences, and ResNet, a convolutional neural network utilized for encoding contact maps. Through comprehensive evaluations on diverse test datasets, we have meticulously assessed the CLEAN-Contact framework's performance. Our extensive comparisons and detailed analyses firmly established that the fusion of structural and sequence information substantially enhances the predictive performance of models used for enzyme functional annotation prediction.

However, our work does come with certain limitations. First, our utilization of structure information relies on contact maps a 2D matrix representation, rather than utilizing the full 3D protein structures of enzymes. One potential solution to overcome this limitation involves the incorporation of 3D interaction sequences (3Di) [11]. These sequences contain valuable information regarding geometric conformations between residues. Another possible way to enhance CLEAN-Contact involves considering EC numbers as hierarchical labels. This approach entails employing hierarchical losses, such as hierarchical contrastive loss [12] or exploring other hierarchical classification loss methodologies [13, 14].

Even though our focus in this work is solely on predicting enzyme functional annotations, we firmly believe that our proposed CLEAN-Contact framework has broader applications beyond this domain. A promising future direction could involve extending our model to predict general protein functional annotations, such as Gene Ontology (GO) numbers [15] and FunCat categories [16]. This expansion would significantly broaden the application and utility of our model.

4 Methods

4.1 Description of Dataset

The enzyme's amino acid sequences in the training dataset were retrieved from Swiss-Prot [17] released in April 2022. Sequences lacking structures available in

the AlphaFold Protein Structure Database https://alphafold.ebi.ac.uk/ were filtered out from the training dataset. The processed dataset comprises 224,742 amino acid sequences, covering 5,197 EC numbers. Test datasets, New-392 and Price-149, consist of 392 and 149 amino acid sequences, respectively, distributed across 177 and 56 EC numbers, as provided by [7] (See Table S5 and S6). For enzymes in the test datasets without available structures, We generated protein structures using AlphaFold2 [3].

4.2 Description of Framework

As shown in Fig. 1, our framework consists of two components: the contrastive learning segment and the EC number prediction segment.

We initiate our approach by obtaining contact maps $C \in \mathbb{R}^{n_r \times n_r}$ from protein structures, where n_r denotes the number of residues in a protein. To expand C into a three-channel matrix $C_3 \in \mathbb{R}^{n_r \times n_r \times 3}$, each channel holds identical values. Utilizing the ResNet50 [9] model pretrained on ImageNet [18], we extract high-dimensional structure representations from contact maps:

$$c = \mathcal{R}(C_3), \tag{1}$$

where $\mathcal{R}$ is the pretrained ResNet50 without its classification layer, and $c \in \mathbb{R}^{d_c}$ is the structure representation, with d_c as its dimension.

Recognizing that contact maps alone lack vital amino acid information, we leverage ESM-2 [8] to derive sequence representations from proteins' amino acid sequence:

$$a = \mathcal{E}(s_{aa}), \tag{2}$$

where $\mathcal{E}$ is the pretrained ESM-2 model, s_{aa} is the protein's amino acid sequence, and $a \in \mathbb{R}^{d_a}$ is the sequence representation, with d_a as its dimension.

To fuse the sequence and structure representations into a vector $z \in \mathbb{R}^{d_z}$, we employ a projector $\mathcal{F}_\Psi(\cdot)$. The projector contains three levels of linear layers, each followed by layer normalization [19]:

$$\begin{aligned} z &= \mathcal{F}_\Psi(a, c), \\ &= W_3 \left(\mathrm{LN}_2 \left(W_2 \left(\mathrm{LN}_1 (W_{1,a} a + W_{1,c} c) \right) \right) \right), \end{aligned} \tag{3}$$

where LN_i is the ith layer normalization, $W_{1,a} \in \mathbb{R}^{d_1 \times d_a}$, $W_{1,c} \in \mathbb{R}^{d_1 \times d_c}$, $W_2 \in \mathbb{R}^{d_2 \times d_1}$ and $W_3 \in \mathbb{R}^{d_z \times d_2}$ are the weights of the linear layers, while Ψ is the projector's trainable parameters.

Subsequently, we compute representations for EC numbers by concatenating the sequence and structure representations of proteins under a specific EC number and averaging these concatenated representations:

$$r_{\mathrm{EC},i} = \frac{1}{|\mathrm{EC}_i|} \sum_{j \in \mathrm{EC}_i} \left(\mathrm{Concat}(a_j, c_j) \right), \tag{4}$$

where $r_{\mathrm{EC},i}$ is the representation of the ith EC number, EC_i is the set of enzymes associated with the ith EC number, $|\mathrm{EC}_i|$ is the set's cardiality, and a_j and c_j are the sequence and structure representations, respectively, of the jth enzyme under the ith EC number.

Further, we compute the distance map between the representations of EC numbers utilizing Euclidean distance:

$$\begin{aligned} d(\mathrm{EC}_i, \mathrm{EC}_j) &= ||z_{\mathrm{EC},i} - z_{\mathrm{EC},j}||_2, \\ &= ||\mathcal{F}(a_{\mathrm{EC},i}, c_{\mathrm{EC},i}) - \mathcal{F}(a_{\mathrm{EC},j}, c_{\mathrm{EC},j})||_2, \end{aligned} \tag{5}$$

where $d(\mathrm{EC}_i, \mathrm{EC}_j)$ is the Euclidean distance between the projected vectors of the ith and j^{th} EC numbers.

During the training phase, we select anchor samples o for each EC number and positive samples p sharing the same EC number as o. For negative samples, we initially choose negative EC numbers with representations that are close to the anchor's EC number representation in Euclidean distance, then select negative samples from these negative EC numbers. We employ the triplet margin loss [20] for contrastive learning:

$$\mathcal{L} = \max(||z_o - z_p||_2 - ||z_o - z_n||_2 + \epsilon, 0), \tag{6}$$

where ϵ is the margin for triplet loss (sets as 1 for all experiments), z_o is the projected vector for the anchor sample, z_p is the projected vector for the positive sample, and z_n is the projected vector for the negative sample.

For the EC number prediction phase, we utilize the contrastively trained projector $\mathcal{F}$ to obtain vectors for enzymes from both the training dataset and test dataset. Subsequently. we compute cluster centers of EC numbers present in the training dataset by averaging the vectors of enzymes associated with each EC number. We then compute the Euclidean distance between the vectors of query enzymes in the test set and vectors of EC numbers. Finally, we predict potential EC numbers utilizing P-value algorithm [7] (See Section S1.2).

Supplementary information. Supplementary Information: Supplementary Figs. 1-3 and Tables 1-4 and Materials and Methods.

Supplementary Table 5: Details of enzymes within the New-392 dataset.

Supplementary Table 6: Details of enzymes within the Price-149 dataset.

Supplementary Table 7: Enzymes of New-392 test dataset lacking structures from AlphaFold Protein Structure Database.

Supplementary Table 8: Enzymes of Price-149 test dataset lacking structures from AlphaFold Protein Structure Database.

Supplementary Table 9: Details of EC number prediction outcomes for enzymes within the New-392 dataset using CLEAN-Contact. Predicted EC numbers were selected using the P-value algorithm.

Supplementary Table 10: Details of EC number prediction outcomes for enzymes within the New-392 dataset using CLEAN-Contact. Predicted EC numbers were selected using the Max-separation algorithm.

Supplementary Table 11: Details of EC number prediction outcomes for enzymes within the Price-149 dataset using CLEAN-Contact. Predicted EC numbers were selected using the P-value algorithm.

Supplementary Table 12: Details of EC number prediction outcomes for enzymes within the Price-149 dataset using CLEAN-Contact. Predicted EC numbers were selected using the Max-separation algorithm.

Acknowledgments. This work was performed on project award (10.46936/expl.proj.2022.60535/60008718) from the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program. This work was also supported by the Predictive Phenomics Initiative Laboratory Directed Research and Development Program. PNNL is a multi-program national laboratory operated by Battelle Memorial Institute for the DOE under Contract DEAC05-76RL01830. This work was also supported by the National Science Foundation (NSF) under Grant #2212465, #2217021, #2217104, #2230111, #2238734, and #2311950 to Qiang Guan.

Declarations

- Competing interests: The authors declare no competing interests.
- Ethics approval: Not applicable
- Code availability: All codes and datasets used in training and testing are available at https://github.com/PNNL-CompBio/CLEAN-Contact
- Authors' contributions: J.Z. and Q.G. conceived the study. Y.Y. developed data and codes, and performed all experiments. Y.Y., A.J., S.F., Z.W., J.Z., and Q.G. performed data analyses, and discussed and interpreted all results. Y.Y., M.C., J.Z., and Q.G. wrote the paper.

References

[1] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J.: Basic local alignment search tool. Journal of molecular biology **215**(3), 403–410 (1990)

[2] Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., Söding, J.: Hh-suite3 for fast remote homology detection and deep protein annotation. BMC bioinformatics **20**(1), 1–15 (2019)

[3] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al.*: Highly accurate protein structure prediction with alphafold. Nature **596**(7873), 583–589 (2021)

[4] Ryu, J.Y., Kim, H.U., Lee, S.Y.: Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. Proceedings of the National Academy of Sciences **116**(28), 13996–14001 (2019)

[5] Sanderson, T., Bileschi, M.L., Belanger, D., Colwell, L.J.: Proteinfer, deep neural networks for protein functional inference. Elife **12**, 80942 (2023)

[6] Gligorijević, V., Renfrew, P.D., Kosciolek, T., Leman, J.K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B.C., Fisk, I.M., Vlamakis, H., *et al.*: Structure-based protein function prediction using graph convolutional networks. Nature communications **12**(1), 3168 (2021)

[7] Yu, T., Cui, H., Li, J.C., Luo, Y., Jiang, G., Zhao, H.: Enzyme function prediction using contrastive learning. Science **379**(6639), 1358–1363 (2023)

[8] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., *et al.*: Evolutionary-scale prediction of atomic-level protein structure with a language model. Science **379**(6637), 1123–1130 (2023)

[9] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[10] Kim, G.B., Kim, J.Y., Lee, J.A., Norsigian, C.J., Palsson, B.O., Lee, S.Y.: Functional annotation of enzyme-encoding genes using deep learning with transformer layers. Nature Communications **14**(1), 7370 (2023)

[11] Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L., Söding, J., Steinegger, M.: Fast and accurate protein structure search with foldseek. Nature Biotechnology, 1–4 (2023)

[12] Zhang, S., Xu, R., Xiong, C., Ramaiah, C.: Use all the labels: A hierarchical multi-label contrastive learning framework. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16660–16669 (2022)

[13] Kosmopoulos, A., Partalas, I., Gaussier, E., Paliouras, G., Androutsopoulos, I.: Evaluation measures for hierarchical classification: a unified view and novel approaches. Data Mining and Knowledge Discovery **29**, 820–865 (2015)

[14] Binder, A., Kawanabe, M., Brefeld, U.: Efficient classification of images with taxonomies. In: Asian Conference on Computer Vision, pp. 351–362 (2009). Springer

[15] Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.*: Gene ontology: tool for the unification of biology. Nature genetics **25**(1), 25–29 (2000)

[16] Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Güldener, U., Mannhaupt, G., Münsterkötter, M., *et al.*: The funcat, a functional annotation scheme for systematic classification of proteins from whole genomes.

Nucleic acids research **32**(18), 5539–5545 (2004)

[17] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., *et al.*: The swiss-prot protein knowledgebase and its supplement trembl in 2003. Nucleic acids research **31**(1), 365–370 (2003)

[18] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). Ieee

[19] Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016)

[20] Balntas, V., Riba, E., Ponsa, D., Mikolajczyk, K.: Learning local feature descriptors with triplets and shallow convolutional neural networks. In: Bmvc, vol. 1, p. 3 (2016)