nature portfolio

Peer Review File

Accurately predicting enzyme functions through geometric graph learning on ESMFold-predicted structures

Reviewer #1 (Remarks to the Author):

This paper presents framworkk for predicting enzyme functions through geometric graph learning on ESMFold-predicted structures. The proposed method called GraphEC predicts the enzyme active sites, optimum pH and EC numbers. The presented results showed that GraphEC outperforms the state-of-the-art methods in the predictions of active sites, EC number, and optimum pH.

The paper is well presented and well structured. The idea to use both structure prediction tools, protein language models is really interesting. Additionnaly, label diffusion algorithm and ProtTrans embeddings were used to improve the predictions by incorporating homology information.

I have two main concerns about the paper:
1 - The way that the authors choose state-of-the-art methods is not justified especially for the EC number prediction part.

2 - Some graph-based methods for EC number prediction where not mentionned and not included in the experimental study:

Bishnu Sarker, David Ritchie, Sabeur Aridhi. GrAPFI: predicting enzymatic function of proteins from domain similarity graphs. BMC Bioinformatics, 2020, 10.1186/s12859-020-3460-7.

So-Ra Han, Mingyu Park, Sai Kosaraju, JeungMin Lee, Hyun Lee, Jun Hyuck Lee, Tae-Jin Oh, Mingon Kang, Evidential deep learning for trustworthy prediction of enzyme commission number, Briefings in Bioinformatics, Volume 25, Issue 1, January 2024, bbad401, https://doi.org/10.1093/bib/bbad401

Reviewer #1 (Remarks on code availability):

The provided git repository is well organized. However, the code is not well commented for future use.
Instructions for installing and running the application are provided.

Reviewer #2 (Remarks to the Author):

The authors have developed GraphEC, an enzyme function prediction tool utilizing geometric graph learning. GraphEC predicts active sites by employing a protein language model and structure prediction model. By combining homologous search, GraphEC performed EC number prediction, outperforming the state-of-the-art EC number prediction tools. Overall, the manuscript introduced a high-performing AI tool that leverages both sequence and structure features for the analysis of enzyme functionality. However, I suggest a few comments to improve the study.

1. The manuscript should provide more detailed methods. For example, the authors mentioned

using a label diffusion algorithm to improve the prediction performance, yet they do not describe this method. Also, it is unclear how the performance of the active site prediction has been evaluated. Was the prediction regarded as correct only if all of the active sites for an enzyme were correctly predicted? If the performance was evaluated per residue, the disproportionate number of active site residues to non-active site residues may hinder using AUC as a fair evaluation metric. For EC number prediction, does GraphEC employ homologous search? How is this integrated into the tool? Likewise, a more detailed explanation of all the methods would improve the manuscript.

2. Since GraphEC employs ESMFold to extract structural features, its inference is likely longer than the inference time of the other state-of-the-art EC number prediction tools. The authors should compare the computational efficiency of these tools and assess whether GraphEC is suitable for high-throughput enzyme function analysis.

3. Please clearly explain the coverage of EC number prediction. Additionally, please compare how many EC numbers can be predicted by GraphEC compared with state-of-the-art EC number prediction tools.

4. GraphEC primarily employs active site features. Could mutations in these sites lead to altered predictions? Is GraphEC capable of identifying mutated enzymes as non-enzymes?

5. If GraphEC can predict the optimum pH for enzymes, the authors should demonstrate its ability to discern differences in optimum pHs among homologous enzymes from acidophiles and others.

6. The manuscript mentions that geometric graph learning employs protein structure predicted by ESMFold, as described in Fig. 1. If experimentally (or even predicted by the state-of-the-art 3D structure prediction tools, such as AF2, AF3, or RoseTTAFold) known protein structures are available, could they replace the ESMFold-based predicted structures? If so, is there potential to improve the prediction performance?

7. The text in the figures is too small and should be enlarged for clarity.

Reviewer #2 (Remarks on code availability):

The code is being updated during the review process. Even though the update was only minor and might not affect the total results, authors should keep the finalized version for the review process.

Reviewer #3 (Remarks to the Author):

In this manuscript, the authors develop the GraphEC model for the prediction of enzyme properties, including enzyme active sites, enzyme commission (EC) number, and optimum pH. The key part of GraphEC is to combine a geometric graph neural network model with a pre-trained protein language model. Further, a homology-based label diffusion algorithm is incorporated into the model to improve the prediction. Based on extensive benchmark

datasets, it has been found that the developed model can achieve superior performance than other state-of-the-art methods.
The paper is well-written, and the results are very promising. I would recommend the publication of the paper if the following questions were well addressed.
Major concerns,
My major concern is the notations in the paper.
In section “Featurizer layer”, line 314, the authors use “C” to denote an atom in the atom pair “A and C. But “C” has been used as carbon atom before.
Further, the author use “N_i” to denote nitrogen atom, but later they also use it to denote atom neighbours.
Moreover, in Equation (3), “exp w_{ji}^l” and “exp w_{ki}^l” are very bad notations.

Minor concerns,
1) In the model, a radius with cutoff distance 10 \AA is used. How importance is the cut-off distance? Will a larger cut-off distance be better?
2) The node and edge features are very important for GNN models. It seems in this model, most of the features are from structural properties, i.e., distances, angles, directions, etc. How about physical properties, such as solvent surface areas, hydrophilic and hydrophobic information, etc.

Reviewer #3 (Remarks on code availability):

The code looks good!

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This paper presents framworkk for predicting enzyme functions through geometric graph learning on ESMFold-predicted structures. The proposed method called GraphEC predicts the enzyme active sites, optimum pH and EC numbers. The presented results showed that GraphEC outperforms the state-of-the-art methods in the predictions of active sites, EC number, and optimum pH.

The paper is well presented and well structured. The idea to use both structure prediction tools, protein language models is really interesting. Additionnaly, label diffusion algorithm and ProtTrans embeddings were used to improve the predictions by incorporating homology information.

I have two main concerns about the paper:
1 - The way that the authors choose state-of-the-art methods is not justified especially for the EC number prediction part.

2 - Some graph-based methods for EC number prediction where not mentionned and not included in the experimental study:

Bishnu Sarker, David Ritchie, Sabeur Aridhi. GrAPFI: predicting enzymatic function of proteins from domain similarity graphs. BMC Bioinformatics, 2020, 10.1186/s12859-020-3460-7.

So-Ra Han, Mingyu Park, Sai Kosaraju, JeungMin Lee, Hyun Lee, Jun Hyuck Lee, Tae-Jin Oh, Mingon Kang, Evidential deep learning for trustworthy prediction of enzyme commission number, Briefings in Bioinformatics, Volume 25, Issue 1, January 2024, bbad401, https://doi.org/10.1093/bib/bbad401

Reviewer #1 (Remarks on code availability):

The provided git repository is well organized. However, the code is not well commented for future use. Instructions for installing and running the application are provided.

Reviewer #2 (Remarks to the Author):

The authors have developed GraphEC, an enzyme function prediction tool utilizing geometric graph learning. GraphEC predicts active sites by employing a protein language model and structure prediction model. By combining homologous search, GraphEC performed EC number prediction, outperforming the state-of-the-art EC number prediction tools. Overall, the manuscript introduced a high-performing AI tool that leverages both sequence and structure features for the analysis of enzyme functionality. However, I suggest a few comments to improve the study.

1. The manuscript should provide more detailed methods. For example, the authors mentioned using a label diffusion algorithm to improve the prediction performance, yet they do not describe this method. Also, it is unclear how the performance of the active site prediction has been evaluated. Was the prediction regarded as correct only if all of the active sites for an enzyme were correctly predicted? If the performance was evaluated per residue, the disproportionate number of active site residues to non-active site residues may hinder using AUC as a fair evaluation metric. For EC number prediction, does GraphEC employ homologous search? How is this integrated into the tool? Likewise, a more detailed explanation of all the methods would improve the manuscript.

2. Since GraphEC employs ESMFold to extract structural features, its inference is likely longer than the inference time of the other state-of-the-art EC number prediction tools. The authors should compare the computational efficiency of these tools and assess whether GraphEC is suitable for high-throughput enzyme function analysis.

3. Please clearly explain the coverage of EC number prediction. Additionally, please compare how many EC numbers can be predicted by GraphEC compared with state-of-the-art EC number prediction tools.

4. GraphEC primarily employs active site features. Could mutations in these sites lead to altered predictions? Is GraphEC capable of identifying mutated enzymes as non-enzymes?

5. If GraphEC can predict the optimum pH for enzymes, the authors should demonstrate its ability to discern differences in optimum pHs among homologous enzymes from acidophiles and others.

6. The manuscript mentions that geometric graph learning employs protein structure predicted by ESMFold, as described in Fig. 1. If experimentally (or even predicted by the state-of-the-art 3D structure prediction tools, such as AF2, AF3, or RoseTTAFold) known protein structures are available, could they replace the ESMFold-based predicted structures? If so, is there potential to improve the prediction performance?

7. The text in the figures is too small and should be enlarged for clarity.

Reviewer #2 (Remarks on code availability):

The code is being updated during the review process. Even though the update was only minor and might not affect the total results, authors should keep the finalized version for the review process.

Reviewer #3 (Remarks to the Author):

In this manuscript, the authors develop the GraphEC model for the prediction of enzyme properties, including enzyme active sites, enzyme commission (EC) number, and optimum pH. The key part of GraphEC is to combine a geometric graph neural network model with a pre-trained protein language model. Further, a homology-based label diffusion algorithm is incorporated into the model to improve the prediction. Based on extensive benchmark datasets, it has been found that the developed model can achieve superior performance than other state-of-the-art methods.

The paper is well-written, and the results are very promising. I would recommend the publication of the paper if the following questions were well addressed.
Major concerns,
My major concern is the notations in the paper.
In section “Featurizer layer”, line 314, the authors use “C” to denote an atom in the atom pair “A and C. But “C” has been used as carbon atom before.
Further, the author use “N_i” to denote nitrogen atom, but later they also use it to denote atom neighbours.
Moreover, in Equation (3), “exp w_{ji}^l” and “exp w_[1]^l” are very bad notations.

Minor concerns,
1) In the model, a radius with cutoff distance 10 \AA is used. How importance is the cut-off distance? Will a larger cut-off distance be better?
2) The node and edge features are very important for GNN models. It seems in this model, most of the features are from structural properties, i.e., distances, angles, directions, etc. How about physical properties, such as solvent surface areas, hydrophilic and hydrophobic information, etc.

Reviewer #3 (Remarks on code availability):

The code looks good!

Response to the reviewers

Our responses are in blue.
We thank the reviewers for acknowledging the novelty of our work and noting the superior performance compared to other state-of-the-art methods. For the constructive and insightful questions raised by the reviewers, we provide the responses point-by-point as follows.

Reviewer 1 (Remarks on code availability):

This paper presents framework for predicting enzyme functions through geometric graph learning on ESMFold-predicted structures. The proposed method called GraphEC predicts the enzyme active sites, optimum pH and EC numbers. The presented results showed that GraphEC outperforms the state-of-the-art methods in the predictions of active sites, EC number, and optimum pH.

The paper is well presented and well structured. The idea to use both structure prediction tools, protein language models is really interesting. Additionally, label diffusion algorithm and ProtTrans embeddings were used to improve the predictions by incorporating homology information.

Response: Thanks for your positive reviews on our paper, as well as your commendation of our innovative approach that integrates protein language models and structure prediction.

I have two main concerns about the paper:

1. The way that the authors choose state-of-the-art methods is not justified especially for the EC number prediction part.

Response: Thanks for your valuable comments. The state-of-the-art methods for EC number prediction were chosen followed by a recent method (CLEAN [2]) published in *Science*, and other methods are selected from recently published articles. Now, we have added two additional methods as required for a more comprehensive comparison. The corresponding descriptions are added in the introduction section.

"*Utilizing the InterPro signatures as domain information, GrAPFI [22] performed label propagation on a weighted undirected graph. For ECPICK [23], the protein sequence was encoded using one-hot embedding, which was subsequently employed to compute the posterior probabilities of around 5000 EC numbers through convolutional and hierarchical layers. CLEAN [11], another deep learning method that learned abundant embeddings through contrastive learning [24], achieved better accuracy and EC coverage for EC number identification.*"

2. Some graph-based methods for EC number prediction were not mentioned and not included in the experimental study:

Bishnu Sarker, David Ritchie, Sabeur Aridhi. GrAPFI: predicting enzymatic function of proteins from domain similarity graphs. BMC Bioinformatics, 2020, 10.1186/s12859-020-3460-7.
So-Ra Han, Mingyu Park, Sai Kosaraju, JeungMin Lee, Hyun Lee, Jun Hyuck Lee, Tae-Jin Oh, Mingon Kang, Evidential deep learning for trustworthy prediction of enzyme commission number, Briefings in Bioinformatics, Volume 25, Issue 1, January 2024, bbad401, https://doi.org/10.1093/bib/bbad401

Response: Thanks for your suggestions. We have added these two methods, GrAPFI [3] and ECPICK [4], to facilitate a more comprehensive comparison. GrAPFI achieved acceptable performance utilizing a protein domain similarity graph, with AUC values of 0.5095 and 0.5407 on Price-149 and NEW-392 (Table S2), respectively. ECPICK attained the third-best performance through the application of a convolutional neural network and hierarchical module, achieving AUC values of 0.5888 and 0.6502 on Price-149 and NEW-392, respectively. GraphEC outperforms these methods with the AUC values of 0.8404 and 0.8910 on Price-149 and NEW-392. The corresponding descriptions are added in section "Enzyme EC number identification (GraphEC)".

"*Relying on the label propagation on a protein domain similarity graph, GrAPFI [22] achieved acceptable performance, with AUC values of 0.5095 and 0.5407 on Price-149 and NEW-392 (Table S2). ECPICK [23] attained the third-best performance through the implementation of a convolutional neural network and hierarchical module, achieving the AUC values of 0.5888 and 0.6502 on Price-149 and NEW-392, respectively.*"

Reviewer 1 (Remarks on code availability):

The provided git repository is well organized. However, the code is not well commented for future use. Instructions for installing and running the application are provided.

Response: Thanks for your suggestions. We have added more detailed comments to the code for future use. The source code of GraphEC is available at https://github.com/biomed-AI/GraphEC.

Reviewer 2 (Remarks to the Author):

The authors have developed GraphEC, an enzyme function prediction tool utilizing geometric graph learning. GraphEC predicts active sites by employing a protein language model and structure prediction model. By combining homologous search, GraphEC performed EC number prediction, outperforming the state-of-the-art EC number prediction tools. Overall, the manuscript introduced a high-performing AI tool that leverages both sequence and structure features for the analysis of enzyme functionality. However, I suggest a few comments to improve the study.

1.1 The manuscript should provide more detailed methods. For example, the authors mentioned using a label diffusion algorithm to improve the prediction performance, yet they do not describe this method.

Response: Thanks for your valuable comments. We have added a more detailed explanation of the methods in the paper for clearer understanding. To provide a more comprehensive explanation of the label diffusion algorithm, we added a detailed algorithm description in section "Label diffusion algorithm".

"*To enhance the initial predictions of EC numbers, a label diffusion algorithm [44, 66] was applied during the testing phase. First, the sequences in the training set similar to the test sequences were found using DIAMOND [15]. Second, based on the sequence identity of protein pairs, a homology network* $M \in R^{T\times T}$ *was constructed (*T *represents the sum of the number of proteins in the test set and the number of hits in the training set). Then, to measure the degree to which a pair of proteins belongs to the same community within the homology network, a Jaccard similarity matrix was defined as follows:*

$$J_{ij} = \frac{\sum_z M_{iz} M_{jz}}{\sum_z M_{iz} + \sum_z M_{jz} - \sum_z M_{iz} M_{jz}} \tag{6}$$

For a target EC number x*, the* x^{th} *column of the final annotation matrix* S *(*S_x*) was learned by minimizing the cost function* $P(S_x)$*:*

$$P(S_x) = \sum_{i=1}^{T} (S_{ix} - Y_{ix})^2 + \frac{\varepsilon}{2} \sum_{i=1}^{T} \frac{1}{d_i} \sum_{j=1}^{T} J_{ij} M_{ij} \left(S_{ix} - S_{jx}\right)^2 \tag{7}$$

Where ε *represents the regularization parameter. The first term serves to preserve the initial labels (*Y_{ix}*), and the consistency of the labels of adjacent nodes is accounted for through the second term. And* $\frac{1}{d_i}$ *is defined as:*

$$\frac{1}{d_i} = \frac{1}{\sum_j J_{ij} M_{ij}} \tag{8}$$

Furthermore, we define M^1 *as:*

$$M^1{}_{ij} = \frac{1}{2}\left(\frac{1}{d_i} + \frac{1}{d_j}\right) J_{ij} M_{ij} \quad (9)$$

its Laplacian matrix L is:

$$L = DM - M^1 \quad (10)$$

where DM is the diagonal degree matrix of M^1. The closed-form solution that minimizes $P(S_x)$ can be converted to:

$$S = (I + \varepsilon L)^{-1} Y \quad (11)$$

where S is the updated annotation matrix, $I \in R^{T \times T}$ indicates an identity matrix, and Y represents the combination of the training set labels along with the initial predictions for the test set."

1.2 Also, it is unclear how the performance of the active site prediction has been evaluated. Was the prediction regarded as correct only if all of the active sites for an enzyme were correctly predicted? If the performance was evaluated per residue, the disproportionate number of active site residues to non-active site residues may hinder using AUC as a fair evaluation metric.

Response: Thanks for your comments. For the active site prediction, the performance was evaluated based on residue. Although AUC can evaluate imbalanced data to some extent, we also employed MCC, precision, and recall for a comprehensive evaluation. The MCC, precision, and recall of GraphEC-AS are 69.2%, 133.2%, and 17.0% higher than the second-best method (PREvaIL_RF), respectively, demonstrating its superior performance (Table S1). Additionally, we have added F1 for further evaluation. The F1 of GraphEC-AS on TS124 is 0.4698 (Table S1), surpassing the second-best method (PREvaIL_RF, 0.240) by 95.8%. These results indicate the outstanding performance of GraphEC-AS. The corresponding descriptions are added in section "Enzyme active site prediction (GraphEC-AS)".

"*We first evaluated GraphEC-AS for enzyme active site prediction based on residue using the independent test TS124 (details shown in Methods).*"
"*In terms of MCC (Matthews correlation coefficient), recall, and precision (Fig. 2B), our method consistently performed the best. The second-best method (PREvaIL_RF) achieved 0.2939, 0.6223, and 0.1487, lower than GraphEC-AS by 40.9%, 14.5%, and 57.1%, respectively. Additionally, the F1 score for GraphEC-AS on TS124 is 0.4698 (Table S1), while the second-best method, PREvaIL_RF, achieves a score of 0.240, reflecting a decrease of 48.9% relative to GraphEC-AS. The PREvaIL needs the calculation of time-consuming evolutionary profiles using PSI-BLAST [48], whereas GraphEC-AS can identify the enzyme active sites rapidly and accurately.*"

Table S1. The Recall, Precision, MCC, and F1 of GraphEC-AS and comparison methods on TS124.

Methods	Recall	Precision	MCC	F1
HA (combination filter)	0.2401	0.1413	0.1757	0.1779
HA (residue identity filter)	0.2770	0.1606	0.2026	0.2033
Crpred (all residues)	0.5013	0.1438	0.2586	0.2235
CRpred (residues with coordinates)	0.5013	0.1469	0.2613	0.2272

PREvaIL_LR	0.4973	0.1406	0.2539	0.2192
PREvaIL_RF	0.6223	0.1487	0.2939	0.2400
GraphEC-AS	**0.7282**	**0.3467**	**0.4973**	**0.4698**

1.3 For EC number prediction, does GraphEC employ homologous search? How is this integrated into the tool? Likewise, a more detailed explanation of all the methods would improve the manuscript.

Response: Thanks for your worthwhile suggestions. For EC number prediction, we have employed homologous search through a label diffusion algorithm using DIAMOND[5]. The importance of homologous search was evaluated, with results indicating that the incorporation of this module led to slight AUC improvements of 0.4% on NEW-392 and 0.2% on Price-149 (Fig. 3D). Specifically, we utilized DIAMOND[5] to conduct a search of the entire test set against the training set in order to identify training sequences that are similar to the test sequences. Then, a homology network was constructed, and subsequently, an annotation matrix was computed to update the initial predictions (more details can be seen in section "Label diffusion algorithm"). The corresponding descriptions are added in sections "The identification of EC numbers (GraphEC)" and "The ablation studies of GraphEC".

"*After pooling, the initial prediction was obtained, and a label diffusion algorithm was employed to enhance the prediction using DIAMOND. The label diffusion algorithm was used to extract homologous information, as referenced by S2F [44].*"
"*When removing label diffusion, the AUC values slightly decreased (Fig. 3D) likely because of the ability of GraphEC to learn homology information.*"

Additionally, we added a more detailed explanation of the protein language model (ProtTrans) and the structure prediction method (ESMFold). The corresponding descriptions are added in sections "The protein language model (ProtTrans)" and "Protein structure prediction using a language model (ESMFold)".

"The informative sequence embeddings were generated through a pre-trained language model ProtT5-XL-U50 (ProtTrans [39]). ProtTrans is a transformer-based autoencoder known as T5 [66], which has been pre-trained on UniRef50 [67] to facilitate the prediction of masked amino acids. The features derived from the final layer of the ProtTrans encoder were employed to enhance the node representations."
"ESMFold [30] is a large language model with up to 15B parameters, developed on the premise that language models can capture evolutionary patterns across millions of sequences. Achieving accurate and fast structure prediction, ESMFold reduces inference time by as much as 60 times in comparison to the state-of-the-art method. Benefiting from its high efficiency, the first evolutionary scale structural characterization of a metagenomic resource has been presented. In this study, we employed ESMFold to predict the protein structures, which were then applied in subsequent geometric graph learning."

2. Since GraphEC employs ESMFold to extract structural features, its inference is likely longer than the inference time of the other state-of-the-art EC number prediction tools. The authors should compare the computational efficiency of these tools and assess whether GraphEC is suitable for high-throughput enzyme function analysis.

Response: Thanks for your valuable suggestions. We have compared the computational time of different methods on Price-149 (Fig. S8). The average inference time for each enzyme of GraphEC is 0.26 seconds (s), and that of CLEAN, ProteInfer, and DeepEC are 1.28s, 0.21s, and 0.14s, respectively (Fig. S8). Due to the extensive time required to calculate the pairwise distances between the query sequence and each EC number cluster center in CLEAN, the inference speed of GraphEC is 392.3% faster than that of CLEAN. The average time for ESMFold to calculate protein structures is 11.44s. With the inference time of GraphEC (0.26s), the total process needs 11.7s, so the calculation of 1,000 enzymes only takes 3.25 hours. In summary, GraphEC is suitable for high-throughput analysis and is expected to be faster since many databases already have pre-predicted ESMFold structures. The corresponding descriptions are added in section "Enzyme EC number identification (GraphEC)".

"*Additionally, we have compared the average computational time per protein of different methods on Price-149. The average inference time for GraphEC is 0.26 seconds (s), while CLEAN, ProteInfer, and DeepEC have inference times of 1.28s, 0.21s, and 0.14s, respectively (Fig. S8). Due to the considerable time needed to compute the pairwise distances between the query sequence and each EC number cluster center in CLEAN, GraphEC's inference speed is 392.3% faster than that of CLEAN. By combining the time required for ESMFold to compute protein structures (11.44s) with the inference time of GraphEC (0.26s), a total of 11.7s is necessary for each enzyme. In this case, computing the functions of 1,000 enzymes requires just 3.25 hours, thereby meeting the needs for high-throughput analysis.*"

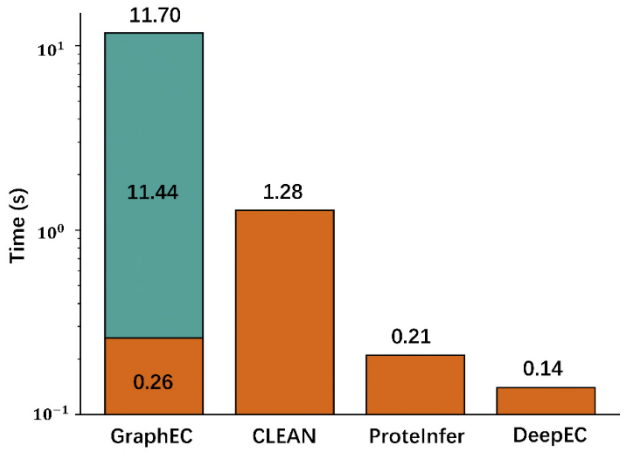

Fig. S8. Comparison of computational time of different methods. Green denotes the time taken by ESMFold to compute protein structures, while orange represents the inference time of models.

3. Please clearly explain the coverage of EC number prediction. Additionally, please compare how many EC numbers can be predicted by GraphEC compared with state-of-the-art EC number prediction tools.

Response: Thanks for your valuable comments. GraphEC is able to predict 5106 EC numbers that are included in the training set obtained from SwissProt. The number of predictable EC numbers of different methods was compared in Table S3. GraphEC and CLEAN can predict the same number of EC numbers (5106) because of their utilization of the same training set, which is 9.4% higher than that of DeepEC (4669). These results indicate that GraphEC can achieve high EC number coverage while maintaining high performance. And the EC number coverage is expected to increase further with the expansion of enzyme data in the future. The corresponding descriptions are added in sections "Enzyme EC number

identification (GraphEC)" and "Hierarchy of catalytic functions".

"*As shown in Table S3, GraphEC is able to achieve high EC number coverage (5106 EC numbers) while maintaining high performance.*"
"*In this study, we collected 5,106 EC numbers from the training set and defined a label of length 5,106, where each position corresponds to a specific EC number.*"

Table S3. The number of predictable EC numbers for different methods.

Methods	Number of predictable EC numbers	Release time
GraphEC, CLEAN	5106	2023
DeepEC	4669	2019
ECPred	858	2018

4. GraphEC primarily employs active site features. Could mutations in these sites lead to altered predictions? Is GraphEC capable of identifying mutated enzymes as non-enzymes?

Response: Thanks for your suggestions. We have evaluated the impact of mutations in the active sites and find that the predicted scores for true EC numbers have decreased a lot after mutation (Fig. S6). After mutating the active sites to Alanine (A), the average predicted score for true EC numbers decreased to 0.2389 (Fig. S6), representing a 66.5% reduction compared to the original score (0.7123). This indicates the impact of mutations in the active sites on EC number prediction. Among the mutated enzymes, 59.1% can be identified as non-enzymes, including L-2-hydroxyglutarate dehydrogenase (Uniprot ID: A0A011QK89) and Farnesyl pyrophosphate synthase (Uniprot ID: B4YA15) (additional cases are presented in Table S4). Furthermore, we analyzed the predicted scores for active sites before and after the mutation, and found that the predicted scores for active sites decreased after mutation (Fig. S7), suggesting a reduced focus of the model on the mutated active sites. The corresponding descriptions are added in section "Enzyme EC number identification (GraphEC)".

"*Considering the utilization of active sites in EC number prediction, we have evaluated the impact of mutations in the active sites. We first identified the active sites of enzymes on NEW-392 and Price-149 based on the predicted results (score > 0.5). Subsequently, these active sites were mutated to Alanine (A), and the predicted scores for true EC numbers were compared before and after the mutation. After mutation, the predicted scores for true EC numbers have decreased (Fig. S6), demonstrating the influence of mutations in the active sites on the prediction of EC numbers. Among the mutated enzymes, 59.1% can be identified as non-enzymes, such as L-2-hydroxyglutarate dehydrogenase (Uniprot ID: A0A011QK89) and Farnesyl pyrophosphate synthase (Uniprot ID: B4YA15) (more cases can be seen in Table S4). Furthermore, the predicted scores for active sites before and after the mutation were compared, discovering a reduction in predicted scores for active sites after mutation (Fig. S7). This indicates a reduced focus of the model on the mutated active sites.*

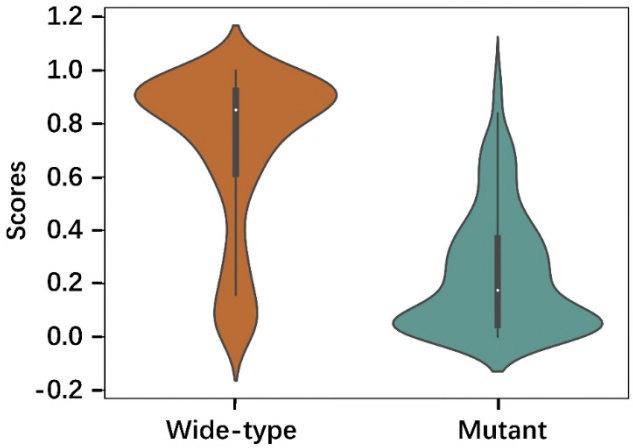

Fig. S6. The scores predicted by our model for the true EC numbers before (left) and after (right) the mutation.

Table S4. The top 10 enzymes selected based on the degree of score decrease after mutation.

Uniprot ID	Predicted EC number (wide-type)	Predicted EC number (Mutant)	True EC number
Q59VG6	1.14.19.20	non-enzymes	1.14.19.20
A0A011QK89	1.1.99.2	non-enzymes	1.1.99.2
B4YA15	2.5.1.10, 2.5.1.1, 2.5.1.29	non-enzymes	2.5.1.10, 2.5.1.1
B1VB71	2.3.1.222	non-enzymes	2.3.1.222
Q0E2F9	3.4.19.12	non-enzymes	3.4.19.12
Q53FA7	1.6.5.5	non-enzymes	1.6.5.5
B1XJV9	2.5.1.29, 2.5.1.1, 2.5.1.10, 2.5.1.81	non-enzymes	2.5.1.29, 2.5.1.1, 2.5.1.10
WP_011717048	5.1.3.3	non-enzymes	5.1.3.3
WP_014879336	4.1.2.29, 4.1.2.13, 4.1.2.40	non-enzymes	4.1.2.29
WP_014235627	5.1.99.1, 4.4.1.5	non-enzymes	5.1.99.1

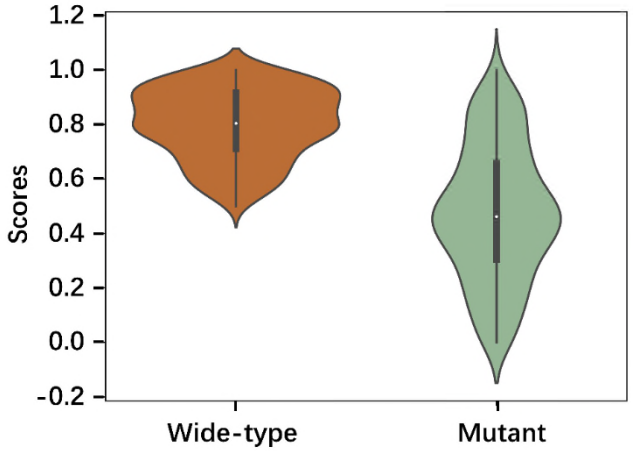

Fig. S7. The scores predicted by the model at active sites before (left) and after (right) mutation.

5. If GraphEC can predict the optimum pH for enzymes, the authors should demonstrate its ability to discern differences in optimum pHs among homologous enzymes from acidophiles and others.

Response: Thanks for your valuable comments. To evaluate the model's ability to discern differences

among homologous enzymes, we employed DIAMOND [5] to search for homologous enzymes in Brenda-test and found 289 pairs of sequences. Among these pairs, more than 87.9% (254 pairs) have the same type of optimum pH (i.e., “acidic” - “acidic” and “non-acidic” - “non-acidic”). GraphEC-pH can correctly identify 95.7% of pairs (243 pairs) with the same type of optimum pH, exceeding the comparison method EpHod by 85.5% (131 pairs). Only 35 pairs of enzymes have different types of optimum pH (i.e., “acidic” - “non-acidic”), and GraphEC-pH can correctly distinguish 14 pairs of them (Table S8), 75% more than EpHod (8 pairs). These results demonstrate that GraphEC-pH can discern the differences among homologous enzymes to a certain degree. The corresponding descriptions are added in section “The prediction of enzyme optimum pH”.

“*We then evaluate the model’s ability to discern differences among 289 homologous enzyme pairs searched by DIAMOND in Brenda-test. More than 87.9% (254 pairs) of the homologous enzyme pairs have the same type of optimum pH (i.e., “acidic” - “acidic” and “non-acidic” - “non-acidic”), and GraphEC-pH can correctly identify 95.7% of them (243 pairs). Only 35 pairs of enzymes exhibit different optimal pH types (i.e., “acidic” - “non-acidic”), with GraphEC-pH correctly distinguishing 14 pairs (Table S8), which is 75% more than EpHod (8 pairs). These results indicate that GraphEC-pH can discern the differences among homologous enzymes to some extent.*”

Table S8. The homologous enzymes (acidophilic and non-acidophilic enzymes) correctly identified by GraphEC-pH. The E-value is calculated through DIAMOND.

Acidophilic enzymes	Non-acidophilic enzymes	E-value
A0A286JZ72	A5A699	5.89e-48
A0A286JZ72	K7ZLW6	1.43e-39
A0A286JZ72	Q7PC53	2.51e-39
A0A286JZ72	E5FGC5	2.97e-28
I0B0T0	A9ZND1	5.01e-07
D0QF43	A0A0U5B5Z4	1.77e-19
D0QF43	A0A075EEX4	3.15e-18
D0QF43	A1XM14	7.56e-08
D0QF43	A4XIF7	9.64e-33
A0A0C5H8Q7	C7NVN0	1.43e-08
I0B0T0	A3DD67	5.87e-72
D0QF43	A0A0F6MV31	1.5e-45
D0QF43	B9VSZ3	7.95e-41
E2GDF4	Q1KLC8	4.21e-17

6. The manuscript mentions that geometric graph learning employs protein structure predicted by ESMFold, as described in Fig. 1. If experimentally (or even predicted by the state-of-the-art 3D structure prediction tools, such as AF2, AF3, or RoseTTAFold) known protein structures are available, could they replace the ESMFold-based predicted structures? If so, is there potential to improve the prediction performance?

Response: Thanks for your valuable suggestion. We replaced the ESMFold-predicted structures with AlphaFold2-predicted structures, which resulted in a slight improvement. Since only 6.7% and 12.1% of proteins in the training set and test set have experimental structures, we hope to test the experimental structures when more data are available in the future. When using the AlphaFold2-predicted structures, the model achieves an AUC, recall, precision, and F1 of 0.9004, 0.8267, 0.5745, and 0.6044 on NEW-392 (Table S6), slightly higher than those of using ESMFold-predicted structures (the AUC, recall, precision, and F1 are 0.8910, 0.7988, 0.5710, and 0.5910, respectively). On Price-149, comparable performance was achieved when utilizing AlphaFold2-predicted and ESMFold-predicted structures. Although high-quality structures can be predicted by AlphaFold2, the high computational cost limits its high-throughput application. And ESMFold can generate structures with comparable accuracy in much less time [1] than AlphaFold2. The corresponding descriptions are added in section "The ablation studies of GraphEC".

"*To evaluate the importance of predicted structures, we replace the ESMFold-predicted structures with those predicted by AlphaFold2. Utilizing the AlphaFold2-predicted structures, the AUC, recall, precision, and F1 on NEW-392 are 0.9004, 0.8267, 0.5745, and 0.6044, respectively (Table S6), slightly higher than those of using ESMFold-predicted structures. On Price-149, comparable performance was obtained when utilizing AlphaFold2-predicted and ESMFold-predicted structures. These results indicate that ESMFold can generate structures with comparable accuracy in much less time than AlphaFold2.*"

Table S6. Comparison of model performance when using AlphaFold2-predicted and ESMFold-predicted structures.

Datasets	Methods	Recall	Precision	F1	AUC
Price-149	GraphEC (AlphaFold2)	**0.6908**	0.6129	0.6127	**0.8406**
	GraphEC (ESMFold)	0.6908	**0.6132**	**0.6131**	0.8404
NEW-392	GraphEC (AlphaFold2)	**0.8267**	**0.5745**	**0.6044**	**0.9004**
	GraphEC (ESMFold)	0.7988	0.5710	0.5910	0.8910

7. The text in the figures is too small and should be enlarged for clarity.

Response: Thanks for your valuable suggestion. We have enlarged the text in the figures so that they can be clearly seen.

Reviewer 2 (Remarks on code availability):

The code is being updated during the review process. Even though the update was only minor and might not affect the total results, authors should keep the finalized version for the review process.

Response: Thanks for your suggestion. We made minor adjustments in response to the user feedback.

And we will avoid doing this in the future.

Reviewer 3 (Remarks to the Author):

In this manuscript, the authors develop the GraphEC model for the prediction of enzyme properties, including enzyme active sites, enzyme commission (EC) number, and optimum pH. The key part of GraphEC is to combine a geometric graph neural network model with a pre-trained protein language model. Further, a homology-based label diffusion algorithm is incorporated into the model to improve the prediction. Based on extensive benchmark datasets, it has been found that the developed model can achieve superior performance than other state-of-the-art methods.

The paper is well-written, and the results are very promising. I would recommend the publication of the paper if the following questions were well addressed.

Response: Thanks for your positive reviews on our paper, and we appreciate your recognition of the promising results of our model.

Major concerns,
My major concern is the notations in the paper.
In section "Featurizer layer", line 314, the authors use "C" to denote an atom in the atom pair "A and C. But "C" has been used as carbon atom before. Further, the author use "N_i" to denote nitrogen atom, but later they also use it to denote atom neighbors. Moreover, in Equation (3), "$\exp w\_\{ji\}\^{}l$" and "$\exp w\_\{ki\}\^{}l$" are very bad notations.

Response: Thanks for your valuable suggestion. We have modified $C \in \{C_j,\ C_{\alpha_j},\ N_j,\ O_j,\ R_j\}$ to $D \in \{C_j,\ C_{\alpha_j},\ N_j,\ O_j,\ R_j\}$, while modifying N_i (neighbors of node i) to NB_i. In Equation (3), $\exp w_{ji}^l$ and $\exp w_{ki}^l$ are revised into $e^{w_{ji}^l}$ and $e^{w_{ki}^l}$. The corresponding descriptions are added in sections "Featurizer layer" and "Geometric graph learning".

"*For atom pairs $A \in \{C_i,\ C_{\alpha_i},\ N_i,\ O_i,\ R_i\}$ and $D \in \{C_j,\ C_{\alpha_j},\ N_j,\ O_j,\ R_j\}$ representing residues i and j respectively, the edge features are defined similarly, including distance, direction, and orientation features. The distance features between residues i and j are $RBF(\| A - D \|)$, indicating the distance characteristics of given residue pairs.*"
"*NB_i represents the neighbors of node i. The queries, keys, and values are translated multiple times, with parallel attention functions being performed before concatenating them together.*"

Minor concerns,
1. In the model, a radius with cutoff distance 10 Å is used. How importance is the cut-off distance? Will a larger cut-off distance be better?

Response: Thanks for your valuable comments. We have tested the impact of different cut-off distances on model performance on NEW-392 and Price-149. When the distance is 12 Å and 14 Å, the AUC of the model are 0.8876 and 0.8753 (Table S7) on NEW-392, which is 0.4% and 1.8% lower than when the distance is 10 Å. This might be due to the fact that a larger distance enables each node to have more edges, resulting in excessive aggregation of information from neighbor nodes during the iterative process, which ultimately reduces the specificity of the nodes. At the same time, a larger distance will result in increased computation and memory overhead. Additionally, when the distance is 8 Å, the AUC of the model is 0.8761 on NEW-392, representing a 1.7% decrease compared to when the distance is 10 Å. This is likely due to the reduction in distance, which leads to a decrease in the number of neighbor nodes per node, resulting in the loss of some information. Similar results on Price-149 can be seen in Table S7. The corresponding descriptions are added in section "The ablation studies of GraphEC".

"*In addition, we also evaluated the impact of various cut-off distances (8 Å, 12 Å, and 14 Å) relative to 10 Å on model performance. When the distance is 8 Å, the AUC, recall, precision, and F1 of the model are 0.8761, 0.7729, 0.5577, and 0.5459 on NEW-392 (Table S7), lower by 1.7%, 3.2%, 2.3%, and 7.6% when the distance is 10 Å. This may be due to the decreased distance, which reduces the number of neighbor nodes associated with each node, ultimately causing some information loss. When the distance is 12 Å and 14 Å, the AUC of the model are 0.8876 and 0.8753 on NEW-392, respectively, 0.4% and 1.8% lower than when the distance is 10 Å (0.8910). This might be because a larger distance allows each node to have more edges, resulting in excessive aggregation of information from neighbor nodes during the iterative process, which eventually reduces the node specificity. Similar results on Price-149 are presented in Table S7.*"

Table S7. Comparison of model performance using graphs with different radii.

	Methods	Recall	Precision	F1	AUC
Price-149	GraphEC (radius = 8 Å)	0.6382	0.6123	0.5808	0.8150
	GraphEC (radius = 10 Å)	**0.6908**	**0.6132**	**0.6131**	**0.8404**
	GraphEC (radius = 12 Å)	0.6776	0.6119	0.6076	0.8344
	GraphEC (radius = 14 Å)	0.6776	0.6012	0.5962	0.8346
NEW-392	GraphEC (radius = 8 Å)	0.7729	0.5577	0.5459	0.8761
	GraphEC (radius = 10 Å)	**0.7988**	**0.5710**	**0.5910**	**0.8910**
	GraphEC (radius = 12 Å)	0.7928	0.5604	0.5481	0.8876
	GraphEC (radius = 14 Å)	0.7689	0.5637	0.5529	0.8753

2. The node and edge features are very important for GNN models. It seems in this model, most of the features are from structural properties, i.e., distances, angles, directions, etc. How about physical properties, such as solvent surface areas, hydrophilic and hydrophobic information, etc.

Response: Thanks for your valuable suggestion. The solvent surface areas, calculated using DSSP [6], have already been incorporated into our model. We also tried to test additional 7 physicochemical properties referring to previous works [7, 8] (including hydrophobicity, steric parameters, volume, polarizability, isoelectric point, helix probability, and sheet probability). Adding these physicochemical

properties failed to further improve the performance of GraphEC, with AUC values of 0.8343 and 0.8899 on Price-149 and NEW-392 (Table S5), suggesting that the geometric features and language model embeddings employed in this study may have already captured the physicochemical properties inherently. The corresponding descriptions are added in section "The ablation studies of GraphEC".

"*Additionally, we have evaluated the effects of physicochemical properties in reference to previous studies [52, 53]. The incorporation of these physicochemical properties failed to further improve the performance of GraphEC (Table S5), suggesting that the geometric features and language model embeddings used in this study may have already inherently captured the physicochemical properties.*"

Table S5. The analysis of model performance in relation to the inclusion of physical properties.

	Methods	Recall	Precision	F1	AUC
Price-149	GraphEC	**0.6908**	0.6132	**0.6131**	**0.8404**
	GraphEC (add physical properties)	0.6776	**0.6228**	0.6084	0.8343
NEW-392	GraphEC	0.7988	**0.5710**	0.5910	**0.8910**
	GraphEC (add physical properties)	**0.8008**	0.5678	**0.6038**	0.8899

Reviewer 3 (Remarks on code availability):

The code looks good!

Response: Thanks for your positive comments on our code.

Reference

1. Lin, Z., et al., *Evolutionary-scale prediction of atomic-level protein structure with a language model.* Science, 2023. **379**(6637): p. 1123-1130.
2. Yu, T., et al., *Enzyme function prediction using contrastive learning.* Science, 2023. **379**(6639): p. 1358-1363.
3. Sarker, B., D.W. Ritchie, and S. Aridhi, *GrAPFI: predicting enzymatic function of proteins from domain similarity graphs.* BMC bioinformatics, 2020. **21**: p. 1-15.
4. Han, S.-R., et al., *Evidential deep learning for trustworthy prediction of enzyme commission number.* Briefings in Bioinformatics, 2024. **25**(1): p. bbad401.
5. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND.* Nature methods, 2015. **12**(1): p. 59-60.
6. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers: Original Research on Biomolecules, 1983. **22**(12): p. 2577-2637.
7. Chen, J., et al., *Structure-aware protein solubility prediction from sequence through graph convolutional network and predicted contact map.* Journal of cheminformatics, 2021. **13**: p. 1-10.
8. Meiler, J., et al., *Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks.* Molecular modeling annual, 2001. **7**(9): p. 360-369.

Reviewer #1 (Remarks to the Author):

All the raised comments have been adressed in teh revised verison of the paper.

Reviewer #1 (Remarks on code availability):

The code of GraphEC is available, well structured and well commented.

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed my concerns.

Reviewer #2 (Remarks on code availability):

The repository is clearly prepared.

Reviewer #3 (Remarks to the Author):

All my comments have been well addressed and the paper quality has been improved. I have no further comments.