Acta Pharmaceutica Sinica B 2026;16(7):4051−4066

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

TOOLS

Accurate and task-agnostic modeling of enzymatic reactions through multimodal relational learning

Yuansheng Huang[a,†], Lanqing Li[b,c,†], Wenjia Qian[a], Jiahui Yu[d], Huifeng Zhao[a], Xiaorui Wang[a], Odin Zhang[e], Guangyong Chen[b], Shukai Gu[a], Pheng-Ann Heng[c,*], Tingjun Hou[a,*], Yu Kang[a,*]

[a]*College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China*
[b]*Research Center for Life Sciences Computing, Zhejiang Lab, Hangzhou 311121, China*
[c]*Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong 999077, China*
[d]*School of Computing, National University of Singapore, Singapore 117417, Singapore*
[e]*Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA 98195-2350, USA*

Received 8 May 2025; received in revised form 11 November 2025; accepted 30 March 2026

KEY WORDS

Molecular representation;
Enzymatic reaction;
Multimodal relational learning;
Protein language models;
Knowledge graph;
Substrate prediction;
Binding site prediction;
Enzymic reaction retrieval

Abstract Enzymatic reactions play an emerging role in a broad spectrum of scientific and industrial applications. The inherent complexity of enzymes, such as their substrate specificity, conformational flexibility, and the vast diversity of reactions involved, poses substantial challenges for the advanced computational prediction of enzymatic reactions with desirable accuracy. Moreover, existing approaches are mostly tailored for a specific sub-task, such as substrate prediction or binding site annotation, which limits their applicability. In this study, we introduce ERAM, a task-agnostic multimodal learning framework capable of addressing a broad range of downstream applications with both accuracy and efficiency. ERAM aligns pre-trained molecular representations from Protein Language Model with the knowledge of enzyme catalysis by modeling enzymatic reactions as multi-relational data. In enzyme retrieval tasks, ERAM achieves an improvement of 28.31% in mean average precision compared with the state-of-the-art (SOTA) method, CREEP. In substrate prediction tasks, ERAM outperforms the SOTA method ESP, achieving average improvements of 35.53% and 22.97% in Matthews correlation coefficient across

This article is part of special issue entitled Machine Learning in Drug Discovery.
*Corresponding authors.
E-mail addresses: yukang@zju.edu.cn (Yu Kang), tingjunhou@zju.edu.cn (Tingjun Hou), pheng@cse.cuhk.edu.hk (Pheng-Ann Heng).
[†]These authors made equal contributions to this work.
Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2026.03.052

two datasets. Additionally, ERAM exhibits commendable interpretability by assigning higher attention weights to binding sites, resulting in lower false-positive rates (42.36%) and higher overlap scores (70.59%) in the unsupervised binding site prediction task compared to RXNAA Mapper. By learning embeddings of substrates, enzymes, and products within a unified knowledge graph latent space, ERAM demonstrates its potential as a versatile and effective tool for enzyme catalysis research.

1. Introduction

Enzymes are a class of proteins that function as biological catalysts, accelerating chemical reactions in living organisms without being consumed. Recently, enzymes have emerged as crucial components in contemporary organic synthesis, making a significant impact in both academia and chemical/pharmaceutical industries, due to their remarkable efficiency, sustainability, and precise control of regio-, chemo-, and stereoselectivity *via* directing groups[1-4]. Therefore, understanding the functions of enzymes, such as Enzyme Commission (EC) numbers, substrate scope, turnover numbers, and catalytic residues, is central for innovations in enzyme rational design, metabolic engineering, and synthetic biology[5-7]. However, traditional approaches of purifying and characterizing one protein at a time, which is the foundation for enzyme databases such as BRENDA[8], are time-consuming, costly, and labor-intensive[9]. Developments of high-throughput technologies are underway but still in their infancy. As a result, the scarcity of high-quality enzyme annotations poses a major barrier to fully unlocking the potential of enzymatic reactions in pharmaceutical research and bioengineering. For example, the UniProt Knowledgebase[10] contains over 25 million enzyme sequences, yet only 0.91% of publicly accessible ones are manually annotated.

To address this challenge, machine learning (ML) and data-driven approaches have been introduced as viable alternatives for automating the functional annotation of uncharacterized enzymes, where a key step is to obtain an informative quantitative representation of the target molecules[11]. In early works, handcrafted features and experimental characterization were widely used to represent molecules. Mou et al.[12] developed models to predict substrates of bacterial nitrilases using input features derived from protein structures, active enzyme sites, and physicochemical property calculation with experimental activity assays. Yang et al.[13] trained a decision tree-based model based on systematically varied combinations of physicochemical properties and structural parameters to predict the substrate scope of plant glycosyltransferases. These models rely on highly specialized experimental training data and expert knowledge to achieve good performance in modeling specific enzymes, but their generalization and scalability are hindered by the narrow data distribution and limited expressive power of classical ML models.

With the surge of biological data and computational resources, deep learning (DL)-based methods have emerged as transformative tools for automatically extracting information-rich molecular representations, surpassing empirical or rule-based features[14-16]. Kim et al.[17] integrated a transformer[18] architecture with protein sequence alignment to effectively predict EC numbers. Teukam et al.[19] employed a BERT[20] model trained *via* multi-task learning on a combination of organic and enzymatic reaction SMILES to predict enzyme binding sites in an unsupervised manner. Li et al.[21] developed a DL-based approach that utilizes substrate graphs and protein sequences as inputs for high-throughput k_{cat} prediction in metabolic enzymes. However, the generalization and flexibility of these models are limited due to the data-intensive nature of DL, and moreover, the fact that enzyme-related data are highly imbalanced with insufficient coverage for the vast majority of enzyme families. To address this issue, recent studies have adopted large pre-trained models to represent enzymes and small molecules. For instance, Kroll et al.[22] developed a gradient boosting model, referred to as ESP, to predict small molecule substrates of enzymes. This model utilizes a fine-tuned ESM-1b[23] protein language model and a task-specific graph neural network (GNN) for generating informative molecular representations. Yu et al.[24] employed contrastive learning and ESM-1b outputs to develop rich enzyme representations, achieving superior performance and generalization in predicting EC numbers.

Despite the remarkable progress, pre-existing methods are mostly unimodal, by extracting protein representations exclusively from enzyme information[17,24-26]. Yet in nature, enzymes never act alone—their biochemical functions are realized through catalytic interactions with the substrates and products in enzyme-catalyzed reactions. Given the high specificity of enzymes for the type of substrates or reactions they catalyze, recent studies have shown that the integration of multimodal knowledge from substrate molecules and enzymatic reaction context can effectively enrich enzyme representations with improved performance for downstream tasks such as binding site prediction[19,27] and enzyme retrieval[28,29]. However, these methods are task-specific, resulting in enzyme representations highly attuned to a handful of problems or downstream tasks, and therefore limited applicability.

In addressing the aforementioned challenges, this study introduces a multimodal learning algorithm to obtain task-agnostic molecular representations for enhanced flexibility and accuracy across a broad spectrum of enzymatic-reaction-related tasks. This enzymatic-reaction-aware molecular representation learning (ERAM) framework is realized *via* the following innovations: (1) conceptualization of substrates, enzymes, and products involved in enzymatic reactions as triplets (head, relation, tail) within a knowledge graph, which enables the modeling of enzymes as translations connecting the substrate and product entities in a unified molecular embedding space, reminiscent of the classical TransE[30] method for modeling multi-relational data; (2) a cross-attention information fusion mechanism for generating substrate-aware enzyme representations, to capture reaction-specific effects such as enzyme promiscuity and induced fit[31,32]; (3) design of dual-grained contrastive learning scheme to perform two-levels of representational alignment based on the essential and mechanism of enzymatic reaction. Our model attains

remarkable performances in various downstream tasks. For the enzyme retrieval task, ERAM outperforms the state-of-the-art (SOTA) enzyme retrieval method CREEP[33], with a mean average precision (MAP) increase of 0.181 (28.31% relative improvement). Compared to the known SOTA substrate prediction model ESP, ERAM achieves an average improvement of 0.185 and 0.133 in Matthews correlation coefficient (MCC) for nitrilase and aminotransferase substrate prediction, respectively. Moreover, our model exhibits commendable interpretability, assigning high attention scores to the binding sites of enzymes and the reaction sites of substrates. In the unsupervised binding site prediction task, ERAM achieves a lower false positive rate (42.36%) and a higher overlap score (70.59%) than the attention mechanism-based self-supervised learning language models RXNAAMapper[19], among other baselines. Our results demonstrate ERAM's potential as a multipurpose tool in solving a variety of biocatalysis and enzyme engineering problems with enzymatic-reaction-aware molecular representations, thus facilitating advanced applications in fields including drug discovery, disease research, and synthetic biology.

2. Materials and methods

2.1. *Model overview*

The ERAM framework comprises two branches: a small molecule encoder and an enzyme encoder, as shown in Fig. 1A. Specifically, the small molecule involved in enzymatic reactions is first converted to the Simplified Molecular-Input Line-Entry System (SMILES) format and then processed using the pre-trained molecule representation model Uni-Mol. The atom-wise embeddings are further processed by a self-attention block, an MLP projector, and mean pooling to obtain final molecular representations. For the enzyme, the amino acid (AA) sequence is processed using ESM-2 to generate AA-wise embeddings, which are then processed through a transformer block with cross-attention, an MLP projector, and mean pooling to produce the representation. Unlike the small molecule branch, a cross-attention block is added to the enzyme encoder to integrate substrate knowledge into the enzyme representation. Therefore, enzymes that exhibit broad specificity can acquire distinct representations when they bind with different substrates, accounting for well-known problems such as enzyme promiscuity and induced fit.

2.2. *Dual-grained contrastive learning*

In biocatalysis, enzymes serve as a bridge, accelerating specific transformations of substrates to products, thereby ensuring the proper functioning of intracellular metabolic processes[34]. Therefore, we conceptualize substrates, enzymes, and products as triplets (head, relation, tail) within a multi-relational knowledge graph. By mapping them to a unified feature space, the enzymes (*i.e.*, relations) can be interpreted as translations operating on the substrate/product embeddings (*i.e.*, entities)[30]. Through minimizing the Euclidean distance between query embedding

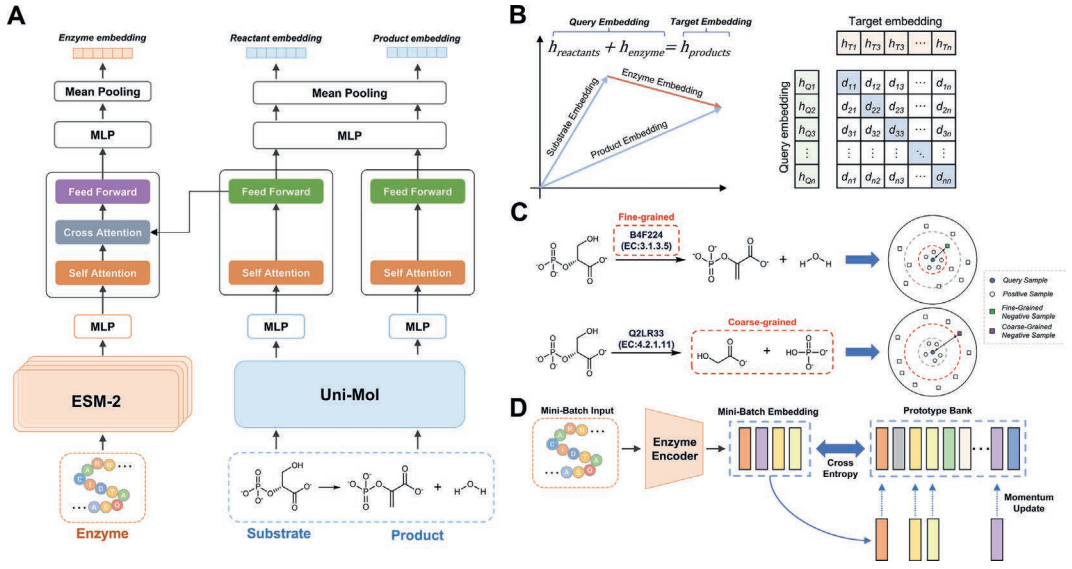

Figure 1 Framework and Methods of Model. (A) An overview of the model structure. The enzyme encoder comprises a frozen ESM-2 backbone, a self-attention block, a cross-attention block, MLP and mean pooling block. The small molecule encoder is composed of frozen Uni-Mol, followed by a self-attention block, an MLP, and mean pooling. Both substrates and products share the same encoder. (B) Relationships among reactants (substrates), enzyme, and products in a knowledge graph and the triplet loss for a minibatch. d_{ij} is Euclidean distance between query embedding (h_{Qi}) and target embedding (h_{Tj}). (C) Dual-grained contrastive learning method. Samples in which the product is permuted with an incorrect product are classified as coarse-grained negative samples, with a large margin as the lower bound distance. Conversely, samples in which the enzyme is removed or replaced by an incorrect enzyme are classified as fine-grained negative samples, with a small margin as the lower bound distance. (D) Enzyme Prototype Learning Procedure. During one iteration, the encoder is updated by computing the cross-entropy of cosine similarities between the enzyme embeddings and their corresponding prototypes within a mini-batch, which is then followed by updates to the prototypes using a momentum-based method (*e.g.*, exponential moving average).

($h_{\text{reactant}} + h_{\text{enzyme}}$) and target embedding ($h_{\text{product}}$) as illustrated in Fig. 1B, we ensure that the learned representations adhere to the translational principle where the head (substrate) plus relation (enzyme) closely approximates the tail (product). The proposed relational learning scheme effectively aligns multimodal molecular representations with enzymatic reaction knowledge by enforcing an equivalence of molecules with respect to reactions in the shared embedding space[35,36]. Since the alignment procedure is not tailored for any specific task, it produces molecular representations that are reaction-aware and task-agnostic, thus enabling a wide array of downstream applications. Moreover, endowed with task-agnostic geometric structure within the reaction triplets, the aligned representations naturally induce a reaction and enzyme retrieval model without additional training or fine-tuning.

To further enhance the learned representations, we observe that in chemical reactions, any changes in substrates or products can significantly alter the chemical equilibrium and even render the whole reaction invalid. In contrast, as with all catalysts, enzymes only accelerate the reaction without affecting the equilibrium or being consumed[34]. To address this functional disparity between the small molecules and enzymes, we propose a dual-grained contrastive learning approach as displayed in Fig. 1C. We generate coarse-grained negatives by replacing the product in a reaction with an incorrect one, which alters the chemical equilibrium. We also generate fine-grained negatives by removing the enzyme or replacing it with an incorrect one, which preserves the chemical equilibrium. To generate negative samples, reaction labels and enzyme labels are assigned. Let $i \in I \equiv \{1 \ldots N\}$ be the index of the sample within a mini-batch. Given the presence of labels, (N, N) masks are generated to distinguish between positive and negative samples. For coarse-grained negative sample generation, chemical reactions are treated as labels, with each unique chemical reaction constituting a separate label. Reaction labels within the mini-batch are utilized to create a coarse-grained (N, N) mask, wherein differences in reaction labels signify mismatched products, corresponding to coarse-grained negative samples. For fine-grained negative sample generation, the EC number serves as the enzyme label; if an enzyme is devoid of an EC number, it is substituted with the RHEA ID. Similar to the coarse-grained approach, enzyme labels in the mini-batch are employed to generate fine-grained (N, N) mask, wherein discrepancies in enzyme labels identify mismatched enzymes, representing fine-grained negative samples.

To distinguish the coarse-grained and fine-grained negative samples, we maximize the distance between the query embedding and the two types of negative embeddings using different margins. Specifically, we use a large margin as the lower bound distance for coarse-grained negative samples to ensure that these samples can be distinctly separated in the embedding space. For fine-grained negative samples, since the permutation of enzymes or removal of enzymes preserves the correct chemical reaction and chemical equilibrium, we employ a smaller margin to enhance the model's sensitivity to distinguish between the variations in catalytic activity without confusing them with the more significant errors of product mismatch.

For loss computation, Let $i \in I \equiv \{1, \ldots, N\}$ index samples in a mini-batch. Define the positive set $P(\mathrm{i}) \equiv \left\{ \mathrm{p} \in I : \tilde{y}_{\mathrm{p}} = \tilde{y}_{\mathrm{i}} \right\}$ with size $|P(\mathrm{i})|$, and the negative set $N(\mathrm{i}) \equiv \{n \in I : \tilde{y}_n \neq \tilde{y}_i\}$ with size $|N(\mathrm{i})|$. We denote h_{s}, h_{e} and h_{p} as the representations of the substrate, enzyme, and product, respectively, and treat them as the head, relation and tail in the triplet in the knowledge graph.

For positive samples, we minimize the Euclidean distance between the query embedding $(h_{\mathrm{s}} + h_{\mathrm{e}})$ and the target embedding h_{p}, as shown in Eqs. (1) and (2). This formulation enforces the translational principle expressed as $h_{\text{head}} + h_{\text{relation}} = h_{\text{tail}}$. For coarse-grained negative samples, which relate to product substitution, a large margin $\gamma_1 = 12$ is set as the lower boundary for distances between samples of different classes, as illustrated in Eq. (3). Conversely, for fine-grained negative samples, which involve the substitution or removal of enzymes, a smaller margin $\gamma_2 = 3$ defines the lower boundary for distances between different classes, as shown in Eqs. (4) and (5).

$$\mathscr{L}_{coarse_{pos}} = \frac{1}{N} \sum_{i \in I} \frac{\sum_{p \in P(i)} \left| h_{s_i} + h_{e_i} - h_{p_p} \right|}{|P(i)|} \tag{1}$$

$$\mathscr{L}_{fine_{pos}} = \frac{1}{N} \sum_{i \in I} \frac{\sum_{p \in P(i)} \left| h_{s_i} + h_{e_p} - h_{p_i} \right|}{|P(i)|} \tag{2}$$

$$\mathscr{L}_{coarse_{neg}} = \frac{1}{N} \sum_{i \in I} \frac{\sum_{n \in N(i)} \max\left(\gamma_1 - \left| h_{s_i} + h_{e_i} - h_{p_n} \right|, 0\right)}{|N(i)|} \tag{3}$$

$$\mathscr{L}_{fine_{neg1}} = \frac{1}{N} \sum_{i \in I} \frac{\sum_{n \in N(i)} \max\left(\gamma_2 - \left| h_{s_i} + h_{e_n} - h_{p_i} \right|, 0\right)}{|N(i)|} \tag{4}$$

$$\mathscr{L}_{fine_{neg2}} = \frac{\sum_{i \in I} \max\left(\gamma_2 - \left| h_{s_i} - h_{p_i} \right|, 0\right)}{N} \tag{5}$$

2.3. *Prototype learning*

In the context of contrastive learning, utilizing a small batch size can result in unstable and biased gradients and, consequently, suboptimal solutions[37-39]. However, computational constraints such as GPU memory often limit the feasible batch size. To address this issue, we propose a novel methodology that utilizes prototype learning to indirectly augment the number of positive samples and enhance the representational capacity of the model without requiring larger batch sizes and additional model parameters, as illustrated in Fig. 1D. We generate prototypes for enzyme classes and use the mean of the embeddings of all samples within the same class as the initial value for the prototype. The prototype embeddings are then updated using a momentum-based approach during training. For the computation of the prototype loss, we employ Cross Entropy as the loss function, shown as in Eq. (6). Here, $A(i) \equiv I$.

$$\mathscr{L}_{prototype} = -\frac{1}{N} \sum_{i \in I} \log \frac{\exp\left(z_i \cdot z_p / \tau\right)}{\sum_{a \in A(i) \exp} \left(z_i \cdot z_a / \tau\right)} \tag{6}$$

This process ensures that the embeddings of samples within the same class become more similar, effectively augmenting the number of positive samples by drawing embeddings closer to their

respective class prototypes. This enhances the model's capacity to learn more robust representations.

The final loss function is defined in Eq. (7).

$$\mathcal{L}_{total} = \mathcal{L}_{coarse\_neg} + \mathcal{L}_{coarse\_pos} + \mathcal{L}_{fine\_neg1} + \mathcal{L}_{fine\_neg2} + \mathcal{L}_{fine\_pos} + \mathcal{L}_{prototype} \tag{7}$$

2.4. Data processing

To prepare the data for training, validating, and testing ERAM, we compile a dataset containing chemical reaction SMILES, enzyme amino acid sequences, and enzyme types (EC numbers or RHEA IDs) based on UniProtKB/Swiss-Prot[10] and RHEA[40]. Because the input of ESM-2 (650 M) encoder is limited to 1024 residues, we exclude sequences longer than 1024 amino acids. Similarly, given the Uni-Mol model's limit of 256 atoms for small molecules, we filter out molecules exceeding this threshold using RDKit. In order to ensure the translational principle, we also exclude reactions where the substrate and product are identical. To ensure a sufficient number of positive samples for contrastive learning and the stability of training, we remove samples where the EC number appeared fewer than ten times. After processing, the final dataset contains 254,106 enzymatic reaction data samples, which have 197,352 unique enzyme sequences, 1718 different EC numbers, and 3048 chemical reactions. It is worth noting that the chemical reactions in the dataset are all balanced to preserve the atom conservation law, therefore preserving the equivalence principle in enzymatic reactions. The dataset distribution of EC numbers is displayed in Supporting Information Fig. S1. We randomly divide the dataset into training, validation, and test sets in an 8:1:1 ratio.

2.5. Experimental settings

To demonstrate that ERAM learns effective representations from enzymatic reactions and is capable of handling a wide array of downstream tasks, we evaluate it on three tasks: enzymatic-reaction retrieval, substrate prediction, and binding-site prediction.

2.5.1. Enzymatic reaction retrieval

We first conduct a series of retrieval tasks based on the enzymatic reactions to evaluate our model's ability to identify the correct reactions. The process primarily involves calculating the Euclidean distances of triples (substrate, enzyme, product) and selecting the best matches based on the ranking of these distances. We establish two specific tasks: product retrieval (finding correct products given substrates and enzyme) and enzyme retrieval (finding feasible enzymes given substrates and products). For the product retrieval task, we utilize Mean Reciprocal Rank (MRR) and Hit@1 as evaluation metrics, as shown in Eqs. (8) and (9).

$$\mathrm{MRR} = \frac{1}{|Q|}\sum_{i=1}^{Q}\frac{1}{\mathrm{rank}_i} \tag{8}$$

$$\mathrm{Hit@1} = \frac{\sum_{i=1}^{n} 1(r_i = 1)}{|Q|} \tag{9}$$

Here, Q represents queries, and rank_i denotes the ranking of the first relevant product retrieved by a query. The term $1(r_i = 1)$ is an indicator function that is set to one if the top recommended item for the i-th test instance is the correct item, otherwise it is 0.

Since multiple candidate enzymes may exist in the enzyme retrieval task, we use Mean Average Precision (MAP) as the evaluation metric. MAP is a common metric used in information retrieval and machine learning to evaluate the performance of ranking algorithms over query sets. This metric is particularly appropriate when multiple relevant items may exist per query, as formalized in Eqs. (10) and (11).

$$AP = \frac{\sum_{k=1}^{n}(P(k) \times \mathrm{rel}(k))}{N_{target}} \tag{10}$$

$$\mathrm{MAP} = \frac{1}{|Q|}\sum_{q=1}^{Q}\mathrm{AP}_q \tag{11}$$

The Average Precision (AP) for a query is calculated by summing the products of precision at each rank and the relevance indicator, divided by the total number of relevant enzymes for that query. Here, $P(k)$ is the precision at cutoff k, N_{target} represents the number of enzymes capable of catalyzing the queried chemical reaction, and $\mathrm{rel}(k)$ is an indicator function that equals one if the enzyme at rank k is relevant, and 0 otherwise. To compute MAP, average the AP scores for all individual queries. Q indicates the queries.

To ensure a comprehensive assessment, we employ MMseqs2[41] to cluster amino acid sequences, thereby evaluating our model's generalization capability by testing it on sequence sets that vary in identity to the training data. The test set is partitioned into stratified subsets based on sequence identity to the training set. The sequence identity ranges are categorized into five distinct groups: 0−40%, 40%−50%, 50%−60%, 60%−70%, and 70%−80%. The statistics of candidate samples in the test sets with different sequence identities are displayed in Supporting Information Table S1.

2.5.2. Baseline setting for enzymatic reaction retrieval

We compare the enzymatic reaction retrieval performance of ERAM against Reactzyme[29] and CREEP[33]. In all Reactzyme variants, we use pretrained embeddings from Uni-Mol (512-dimensional) for molecules and from ESM (1280-dimensional) for proteins as the official implementation. These embeddings are first projected into a shared latent space of 128 dimensions through multilayer perceptrons (MLPs). The projected molecular and protein representations are then integrated using cross-attention mechanisms and subsequently processed by one of the following three network architectures: MLPs, transformers, or recurrent neural networks (RNNs). The parameters of these network architectures are consistent with those in the original publication. All models are trained with the same set of hyperparameters: we use the AdamW optimizer with a learning rate of 1e-4 and a weight decay of 5e-10. The models are trained for up to 50 epochs, with early stopping triggered after 30 epochs without validation improvement. The batch size is set to 1024, and we use binary cross-entropy loss combined with a contrastive loss to encourage alignment between molecular and protein representations. All experiments are run on multiple GPUs using PyTorch DistributedDataParallel (DDP), and we ensure reproducibility by fixing all random seeds.

In the CREEP baseline configuration, we fine-tune the pretrained protein language model ProtT5 and the reaction model rxnfp to learn aligned representations of reactions and proteins using contrastive learning and generative tasks. The model is

trained using the Adam optimizer with a learning rate of 1e-5 for both protein and reaction encoders, with learning rate scaling factors set to 1. The weight decay is set to 0. We use a batch size of 16, with negative sampling for contrastive learning set to one negative sample per batch. Training lasts for 40 epochs, with early stopping based on the lowest loss observed. The contrastive loss is computed using the EBM-NCE loss function, and the generative loss is based on mean squared error (MSE) between the reaction-to-protein and protein-to-reaction facilitators. To ensure reproducibility, all random seeds are fixed across different stages of the training process.

2.5.3. *Substrate prediction datasets*

For the substrate prediction tasks, we choose the nitrilase datasets[12] for evaluation, as it was previously employed in the SOTA substrate prediction model ESP, which exhibits excellent performance on this dataset[22]. This dataset comprises 240 data points, covering all possible combinations of 12 enzymes and 20 substrates. The other dataset of aminotransferases[42,43] is also employed for external evaluation. This dataset includes 450 data points, with each of the 25 aminotransferases tested against 18 identical small molecules. Given that aminotransferases exhibit enzyme activity, we labeled samples with activity levels equal to or greater than 0.1 U/mg as 1, and the others as 0. Additionally, we excluded data that appeared during the training of our model, ultimately retaining 444 data entries. The glycosyltransferase substrate prediction dataset employed in ESP was excluded here because nucleophilic acceptors of sugars have multiple glycosylation sites, leading to unfixed enzymatic products. We divide these datasets into training, validation, and test sets in a 7:1:2 ratio. To ensure a comprehensive evaluation, we implement three different dataset partitioning strategies: random splitting, splitting by enzymes' amino acid sequences, and by substrates.

2.5.4. *Baseline setting for substrate prediction*

In substrate prediction, we first fine-tune our model on the training set by minimizing the Euclidean distance between the enzyme and the correct substrate, while maximizing the Euclidean distance between the enzyme and incorrect substrates. Upon completion of training, we establish a threshold for Euclidean distance based on the Matthews Correlation Coefficient (MCC) derived from the validation set. This threshold is subsequently applied to assign labels to samples in the test set.

For benchmarking purposes, we assess our model's performance against ESP. For ESP, we consider two training regimes. The first method, denoted as "ESP," performs hyperparameter optimization on the training and validation sets of the downstream task. Using Hyperopt[44], we perform a random grid search to optimize several hyperparameters, including learning rate, maximum tree depth, lambda and alpha coefficients for regularization, maximum delta step, minimum child weight, training epochs, and weight for negative data points. After optimization, the selected configuration is used to train gradient-boosting models. The training is conducted using the Python package xgboost[45]. The second method, "ESP + extra data", utilizes 5-fold cross-validation to determine the best hyperparameters for the gradient boosting model, applied on an experimentally validated dataset of 55,878 entries. This dataset is subsequently integrated into the downstream task's training set, and all data are used to train the gradient-boosting model. Evaluation is ultimately conducted on the held-out test set. Hyperparameters for this setup follow ESP's released source code.

2.5.5. *Binding site prediction*

Understanding the binding sites of enzymes is crucial for elucidating their catalytic mechanism and has profound implications for directed evolution and enzyme engineering[46,47]. To demonstrate our model's efficacy and potential in predicting binding sites, we directly utilize the attention scores from the enzyme encoder to predict the binding sites of enzymes. For this analysis, we employ a dataset previously studied by Teukam et al.[19], which originated from the Protein-Ligand Interaction Profiler (PLIP)[48]. This dataset comprises 777 co-crystallized ligand-protein pairs, offering a rich resource for evaluating our model against enzymatic interaction sites. We benchmark our model against the BERT-based RXNAAMapper[19] model developed by Teukam et al.[19], as well as the baseline models they identified. All models are trained without information on binding sites.

Following Teukam et al.[19], we utilize the overlap score and false positive rate as metrics.

$$\text{Overlap score} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{13}$$

where TP is the number of true positives, TN is the number of true negative, FP is the number of false negative and FN is the number of false negatives.

3. Results and discussion

3.1. *Embedding visualization*

In order to investigate whether ERAM has learned effective representations of enzymatic reactions, we use our model to embed molecules in the training set into a 512-dimensional latent space. To visualize the enzyme embeddings, we utilize T-distributed Stochastic Neighbor Embedding (T-SNE)[49] and randomly highlight 15 enzyme classes by distinct colors. For comparison, we also plot the 2D projections of pre-trained ESM-2 embeddings *via* T-SNE. As depicted in Fig. 2A and B, for enzymes belonging to the same class, representations learned by ERAM are effectively clustered but appear dispersed for ESM-2, illuminating that the model has learned more semantically compact and consistent representations during training.

Since ERAM simultaneously models enzymes and small molecules in the same representation space, we can jointly visualize them *via* T-SNE. As shown in Fig. 2C, enzymes form tight clusters, while the small molecule embeddings are more scattered, encapsulating the enzyme clusters. The distinct boundary between small molecules and enzymes indicates that the model, through our proposed dual-grained contrastive learning, has successfully captured the functional disparity between enzymes and small molecules in enzymatic reactions by mapping them into disjoint subspaces. Intuitively, switching product alone renders the whole reaction infeasible, effectively leaving a large gap (mismatch) in the knowledge graph triplet. In contrast, replacing the enzyme only affects the reaction rate without changing its feasibility or equilibrium, therefore creating a much smaller gap for the triplet in the embedding space. Based on such reasoning, perhaps counterintuitively, we expect the norm of small-molecule compounds to be much greater than that of enzymes in the shared feature space, which is confirmed by their norm distributions shown in Fig. 2D and E, respectively. Fig. 2C also provides a 2D

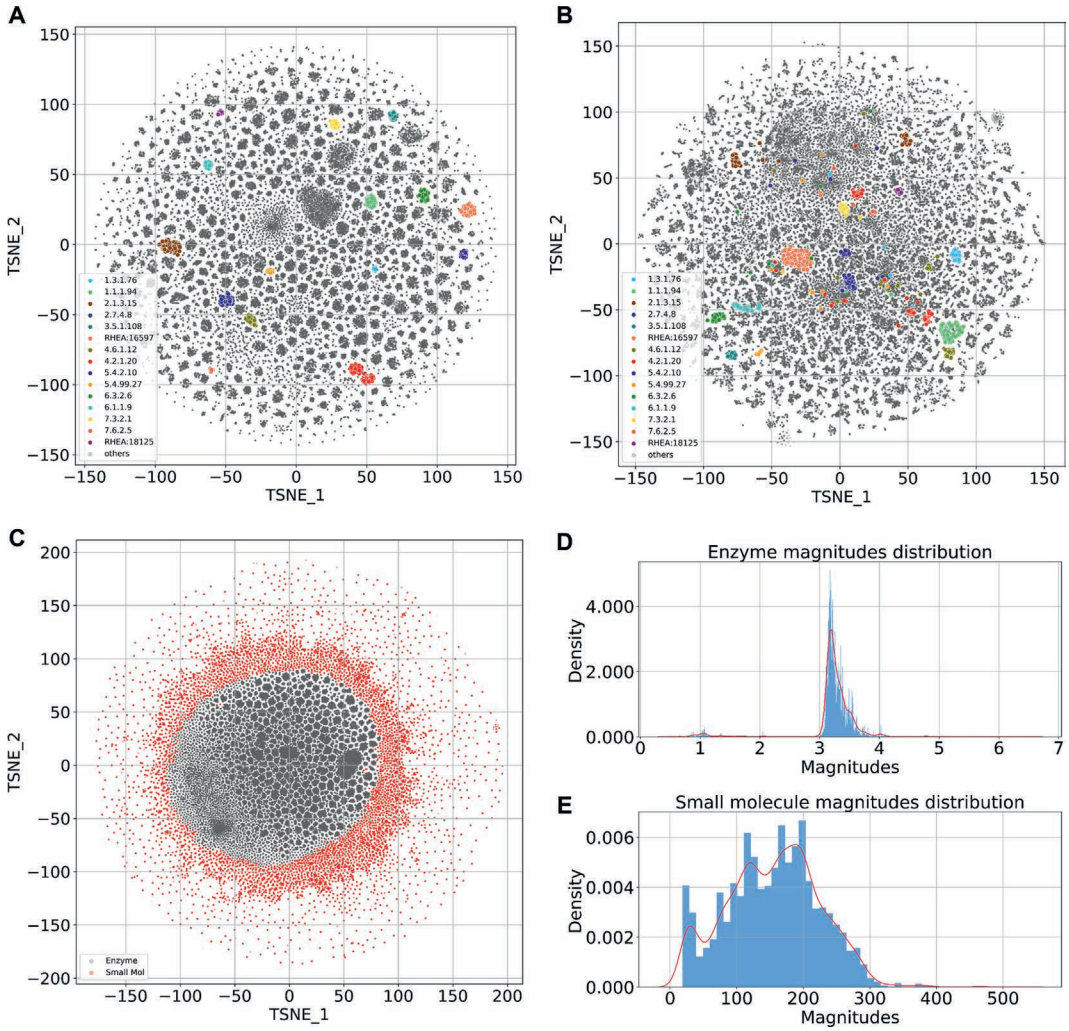

Figure 2 Embedding visualization. Two-dimensional T-SNE visualization of representations from (A) our model and those from (B) ESM-2. Each point represents the embedded representation of an enzyme. 15 enzyme categories are randomly selected for highlighting. The representations derive from our model cluster enzymes of the same category together, whereas the corresponding categories in the ESM-2 representations appear more dispersed. (C) Two-dimensional T-SNE visualization of enzyme (grey) and small molecule (red) representations learned by ERAM. Norm distribution of (D) enzyme and (E) small molecule representations are also given.

visualization of this phenomenon, where the bulk of enzyme embeddings is fully enclosed by the substrate and product molecules with a greater norm.

3.2. *Reaction retrieval evaluation*

3.2.1. *Outstanding performance in all test sets with varying sequence identity*

In the enzymatic reaction retrieval task, our model demonstrates exceptional performance across test sets with varying sequence identities, as shown in Table 1. Notably, in test sets characterized by low sequence identity (<40%), the model achieves higher MAP and MRR than those observed in the complete test set, indicating the model's robustness. We further divide the test set into subsets according to EC number and evaluate both product retrieval performance (using MRR and Hit@1) and enzyme retrieval performance (using MAP) for each subset. The distribution of EC numbers at the first level in the test set can be found in Supporting Information Fig. S2 and Table S2. As shown in Supporting Information Table S3, our model achieves sound performance across all EC subsets, including the one without EC annotations. Some EC-subset variability in retrieval performance can be observed, as expected, because of the sample size, enzyme diversity, and candidate set complexity. For example, subsets like EC1 and w/o EC with greater diversity among candidate products

Table 1 Enzymatic reaction retrieval performance on the test sets with different sequence identity.

Sequence identity	Product MRR ↑	Product Hit@1 ↑	Enzyme MAP ↑
0−100%	0.9836	0.9701	0.8202
70%−80%	0.9980	0.9961	0.9684
60%−70%	0.9988	0.9980	0.9733
50%−60%	0.9982	0.9968	0.9752
40%−50%	0.9949	0.9898	0.9723
0−40%	0.9952	0.9903	0.9770

tend to yield lower retrieval performance (MRR and Hit@1), suggesting that increased candidate variability may introduce additional challenges for the model.

To further investigate the performance of our model in reaction retrieval, we identify several scenarios where the model exhibits lower average precision (AP): First, enzymes that can catalyze multiple chemical reactions, including both forward and reverse reactions, present challenges; second, due to the hierarchical reaction classification in the RHEA Database, chemical reactions with hierarchical relationships often involve participants that are identical or related, either directly or indirectly, which confuses the model[40]. Additionally, the SMILES notation, which represents some variable groups as "R" groups, poses difficulties in distinguishing these reaction types during training. Lastly, enzymes represented by a smaller number of training samples also experience decreased performance. Representative cases are shown in Supporting Information Figs. S3−S5. Nevertheless, sequence identity exerts a minor influence on the model's effectiveness; even in test sets with low identity, the model continues to perform excellently in the reaction retrieval task.

3.2.2. ERAM effectively distinguishes between isomers in the product retrieval task

To further investigate ERAM's product retrieval capability, we analyze its performance in distinguishing between isomers. Differentiating between isomers is challenging because they share identical molecular formulas. We present product retrieval results for enzymatic reactions classified as EC:1−EC:6 involving isomers. EC:7 reactions, representing translocases that catalyze the transport of ions or molecules across membranes or their redistribution within membranes, are excluded since their substrates and products are identical. As illustrated in Fig. 3A−F, ERAM correctly ranks the appropriate isomer for various enzymatic reactions depending on the enzyme type. For instance, as depicted in Fig. 3A, the model accurately identifies chiral changes occurring during the oxidation of methionine catalyzed by Methionine-R-sulfoxide reductase (Q3MHL9). Moreover, ERAM identifies the chiral changes resulting from the type B cyclization of geranylgeranyl diphosphate (GGDP) catalyzed by Phyllocladan-16-alpha-ol synthase (B2DBF0), as shown in Fig. 3E. These results demonstrate ERAM's accurate understanding of enzymatic reactions.

3.2.3. ERAM outperforms reactzyme and CREEP in the enzyme retrieval task

To enable a systematic comparison of retrieval task performance, we train two enzyme-reaction retrieval models, Reactzyme[29] and CREEP[33], using our dataset. Since neither Reactzyme nor CREEP is designed for product retrieval, our evaluation is restricted to enzyme retrieval performance. Reactzyme utilizes ESM-2 for enzyme encoding and Uni-Mol for substrate and product encoding, similar to ERAM; however, it relies only on binary cross-entropy (BCE) or contrastive loss during training. In contrast, CREEP leverages fine-tuning of the pretrained language models rxnfp[50] and ProtT5[51] to learn aligned representations of reactions and proteins through contrastive learning, instead of using frozen pretrained models. For Reactzyme, we compare model performance using two types of loss functions and different encoder architectures—including multilayer perceptron (MLP), transformer, and recurrent neural network (RNN)—as described in the Reactzyme methodology. For negative sample generation, 1000 negative samples are generated for each reaction based on Levenshtein distance. For CREEP, we also follow the hyperparameter configuration described in the original publication, and negative samples are sampled within each batch. Subsequently, we compare the enzyme MAP across all test sets. As shown in Fig. 4, both Reactzyme and CREEP experience a dramatic decline in the 0−40% subset, although they can achieve sound performance for sequence identities ranging from 50% to 80%. This suggests that these two models may be difficult to retrieve correct enzymes effectively in low sequence identity test sets, even when the number of candidates is small. Notably, however, ERAM achieves superior performance across all test sets, especially in low sequence identity test sets, highlighting the robustness and effectiveness of our approach.

We further compare the performance of enzyme retrieval across all EC subsets. For Reactzyme, we adopt its best-performing variant, which combines binary cross-entropy (BCE) with a transformer encoder. As shown in Table 2, both baselines are sensitive to EC-specific diversity and experience performance drops in subsets such as EC1 and w/o EC. However, ERAM consistently achieves the highest MAP across all EC subsets. Notably, for the most challenging w/o EC subset, ERAM achieves MAP improvements of 93.01% over Reactzyme and 63.86% over CREEP, further underscoring its superior generalization and robustness in scenarios without EC annotations. These results collectively indicate that although variability in retrieval performance exists across EC subsets, driven by factors such as sample size, enzyme diversity, and candidate set complexity, ERAM demonstrates strong generalization across both common and rare enzyme families. Its limited sensitivity to data imbalance, combined with superior retrieval accuracy, highlights the effectiveness of the dual-grained contrastive learning strategy in capturing enzymatic specificity.

3.2.4. Ablation study proves the effectiveness of ERAM

To validate the efficacy of our model design, we conduct a series of ablation experiments on both the product retrieval and enzyme retrieval tasks within the test set. To demonstrate the effectiveness of dual-grained contrastive learning, we set identical margins in the contrastive loss and do not distinguish between coarse-grained and fine-grained negative samples. Here, we introduce the term Margin-Coarse, which assigns all margins as coarse-grained. Similarly, we define Margin-Fine, which assigns all margins as fine-grained. Additionally, we ablate the prototype learning to assess its impact. Moreover, to examine the effect of substrate-dependent enzyme representations enabled by the cross-attention module, we replace the cross-attention block in the enzyme block with a self-attention block. This is referred to as the Self-Attn control group.

Compared to ERAM, the results of Margin-Coarse and Margin Fine control groups in Table 3 are clear indications of the necessity of our proposed dual-grained learning strategy. The performance of ERAM is comparable to that of Margin-Fine. This can be attributed to the fact that, in contrast to models trained under contrastive learning frameworks with two distinct margins, the optimized training objective with a reduced margin is simpler and may yield similar results in retrieval tasks. ERAM outperforms the model trained with Margin-Coarse, as maintaining a substantial margin across all types of negative samples is challenging and may hinder the model from learning effective representations. Another pronounced shortcoming of uniform contrastive margin for enzymatic reaction is that the model can no longer accurately capture the functional disparity between substrates/products and enzymes by separating them into disjoint subspaces in their shared embedding space, which is evident by

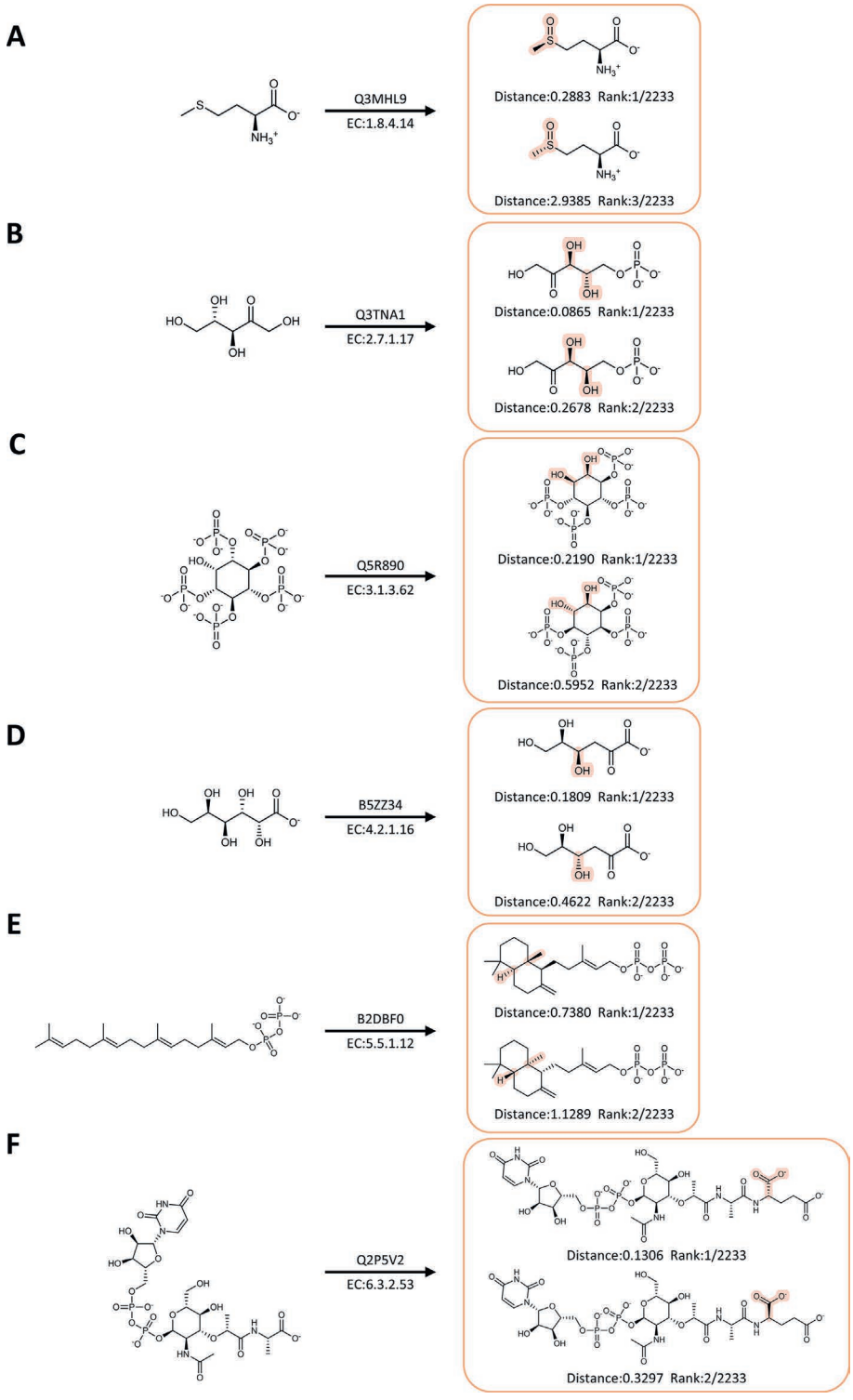

Figure 3 Product retrieval results for enzymatic reactions involving isomeric compounds. The results show the Euclidean distances calculated between the combined representation of reactant and enzyme, and candidate isomers, along with their corresponding rank positions.

comparing Supporting Information Fig. S6 with Fig. 2C. Moreover, ERAM's MRR and Hit@1 performance exhibit minimal differences compared to model lacking prototypes. However, ERAM demonstrates a significant advantage in the MAP for enzymes, suggesting that prototype learning effectively enhances the similarity of representations among enzymes within the same category, while implicitly increasing the number of positive samples in contrastive learning. This enhancement notably improves the model's performance on enzyme retrieval tasks. Furthermore, ERAM outperforms Self-Attn in both product

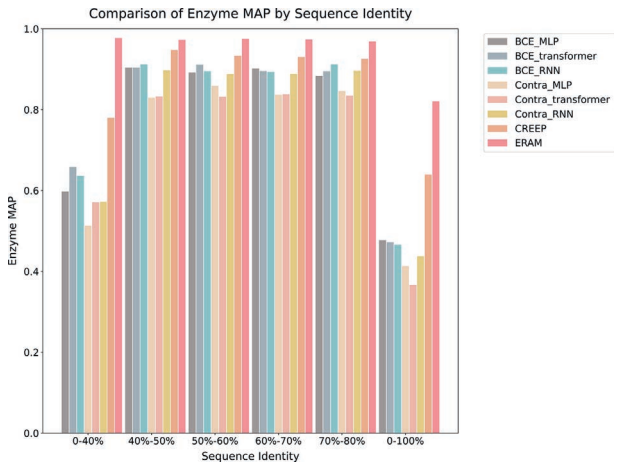

Figure 4 Performance comparison of Reactzyme model, CREEP and ERAM in the enzyme retrieval task. BCE denotes training with binary cross-entropy loss, whereas Contra denotes training with contrastive loss.

retrieval and enzyme retrieval tasks, which indicates that the cross-attention block successfully models the interactions between substrates and enzymes, thereby enhancing the performance of the reaction retrieval task.

3.3. *Enzyme substrate prediction*

3.3.1. *ERAM outperforms ESP in substrate prediction*

In the substrate prediction task, our model achieves SOTA performance across nearly all metrics and split settings in both nitrilase and aminotransferase substrate predictions, surpassing the performance of the ESP model, as shown in Fig. 5A−F. Although the "ESP + extra data", trained with an additional, experimentally validated dataset, outperforms the standard "ESP" model, ERAM achieves superior results without the need for additional data. Moreover, as illustrated in Fig. 5C and F, the performance of ESP notably declines under substrate-based splitting; in contrast, our model exhibits minimal decline, indicating superior robustness. Moreover, we evaluate two additional substrate prediction datasets: the OleA thiolase family and the DUF849 family. As shown in Supporting Information Fig. S7, while baseline models such as ESP and ESP + Extra data perform well in certain settings, ERAM achieves the best or comparable results across most metrics, particularly in scenarios that demand better generalization. Collectively, these results highlight the model's generalization capabilities and effectiveness, suggesting significant potential for the development of reliable enzymatic substrate prediction tools.

Table 2 Enzymatic reaction retrieval performance on the test sets across different EC numbers.

EC Number	Enzyme MAP ↑		
	Reactzyme	CREEP	ERAM
EC1	0.5688	0.7246	**0.7874**
EC2	0.7033	0.8089	**0.8913**
EC3	0.6747	0.7708	**0.9465**
EC4	0.7388	0.7858	**0.8102**
EC5	0.7801	0.8037	**0.8433**
EC6	0.8627	0.8075	**0.9513**
EC7	0.7794	0.6866	**0.9395**
W/o EC	0.4238	0.4992	**0.8180**

Table 3 Enzymatic reaction retrieval results on the ablation study.

Method	Product MRR ↑	Product Hit@1 ↑	Enzyme MAP ↑
Margin-fine	0.9773	0.9655	**0.8325**
Margin-coarse	0.9669	0.9502	0.7525
W/o prototype	0.9829	0.9696	0.8014
Self-Attn	0.9781	0.9593	0.6755
ERAM	**0.9836**	**0.9701**	0.8202

3.3.2. *Ablation study demonstrates the effectiveness of ERAM*

We conduct a series of ablation experiments on our model, as shown in Supporting Information Tables S4 and S5. In addition to the ablation settings in the reaction retrieval task, we replace balanced chemical reactions with unbalanced ones, which violate the laws of atom and mass conservation, and refer to this configuration as Unbalanced. The results of Margin-Fine and Margin-Coarse reveal that the performance in substrate prediction decreases when the model does not differentiate between the hard and the easy negative samples by using a uniform margin. This suggests that dual-grained contrastive learning enhances the model's ability to learn effective molecular representations, thereby improving performance and generalization in enzyme substrate prediction. Additionally, the omission of prototype learning from the model results in diminished substrate prediction performance, highlighting the role of prototype learning in augmenting the efficacy of contrastive learning by increasing the number of positive samples indirectly. Furthermore, training the model with unbalanced chemical reactions instead of balanced reactions leads to decrease performance in substrate prediction. This is expected since balanced chemical reactions adhere to the law of atom and mass conservation, while unbalanced reactions do not, leading to biased representations learned by the model.

3.4. *Enzymatic binding sites prediction*

3.4.1. *ERAM achieves superior results in the unsupervised binding site prediction task*

We evaluate performance using attention scores from different heads in the unsupervised enzymatic binding sites prediction. Results are shown in Supporting Information Table S6. Although head seven yields the highest overlap score (70.85%), it also comes with the highest FPR (45.28%), indicating a greater number of false positives. In contrast, head five offers the best trade-off, achieving a high overlap score (70.59%) while maintaining the lowest FPR (42.36%) among all heads. Accordingly, we select head five for downstream evaluation.

Using this setting, and as summarized in Table 4, ERAM achieves a 70.59% overlap score and a 42.36% false positive rate. Compared to the RXNAAMapper[19] and the Pfam-based method[52], our model exhibits a relatively low false positive rate while attaining a higher overlap score. This indicates that our model has a good understanding of the mechanism of enzymatic reaction and can accurately identify binding sites, thereby validating its effectiveness in recognizing critical enzymatic features.

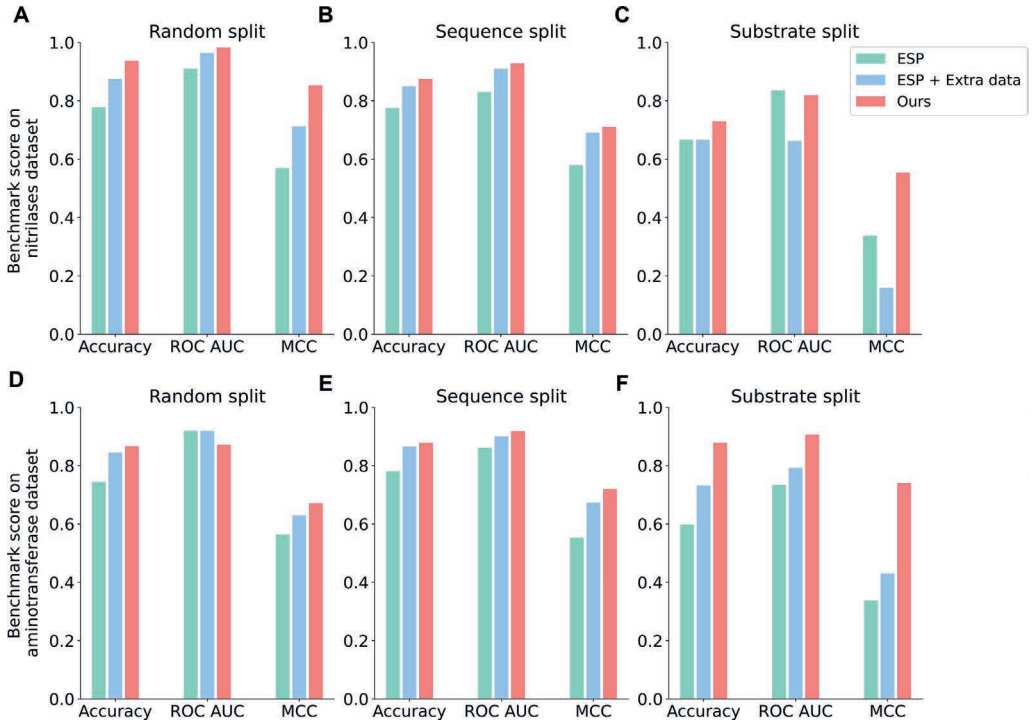

Figure 5 Performance of the model in substrate prediction tasks. (A−C) Displays the accuracy (ACC), area under the receiver operating characteristic curve (ROC-AUC), and the matthews correlation coefficient (MCC) for nitrilase substrate prediction under different data splitting methodologies: random split, sequence split, and substrate split. (D−F) ACC, ROC-AUC and MCC for aminotransferase substrate prediction under different data splitting methodologies.

Table 4 Binding site prediction results on the PLIP dataset.

Method	Overlap Score ↑	False Positive Rate ↓
Random model	14.07%	13.80%
BERT-base	15.27%	14.92%
BERT-large + BPE	45.91%	42.71%
ProtAlbert	18.27%	16.23%
ProBert	20.03%	16.42%
RXNAAMapper	52.13%	47.89%
Pfam-based	67.37%	61.68%
ERAM	70.59%	42.36%

3.4.2. Ablation study on the unsupervised binding site prediction task

Furthermore, we conduct ablation experiments to evaluate the influence of specific design choices of our methods. The ablation setting is identical to the substrate prediction task. The results are shown in Supporting Information Table S7. The superior performance of our model underscores the efficacy of the core design of ERAM, including dual-grained contrastive learning (ERAM *vs*. Margin-Fine, Margin-Coarse), cross-modal information fusion (ERAM *vs*. Self-Attn), and enzymatic-reaction-aware loss (ERAM *vs*. Unbalanced). All these techniques substantially enhance the ERAM's capacity in modeling the underlying mechanisms of enzymatic catalysis. The proposed synergistic integration of these design choices significantly enhances the model's capability to pinpoint critical enzyme structural regions essential for catalytic activity.

3.4.3. ERAM's attention scores accurately highlight enzyme binding sites

To further demonstrate the application of ERAM in binding sites prediction, we analyze the attention scores from our enzyme encoder to determine whether the model captures information pertinent to enzyme binding sites as labeled in UniprotKB[10]. As illustrated in Fig. 6, the enzyme encoder assigns high attention scores to the amino acid sequences from position 124 to 132 (DDVITVGTA) in phosphoribosyltransferase (EC:2.4.2.10) A1AXP4 and from position 88 to 96 (DDLVDTGGT) in phosphoribosyltransferase B5BDQ2. These regions correspond to the binding sites for 5-phospho-alpha-D-ribose 1-diphosphate (PRPP) in phosphoribosyltransferase. Similarly, the model assigns substantial attention to the ATP binding sites at positions 34 to 41 (GLSGSGKS) in Adenylyl-sulfate kinase (EC:2.7.1.25) A6KXG9. The enzyme encoder also assigns high attention scores to the amino acid sequences from position 204 to 209 (TAYAFS), which correspond to the NAD + binding sites in NAD kinase (EC:2.7.1.23) Q49897. Besides, we also visualize the attention scores of enzymes from other EC numbers (EC:1, EC:3, EC:4, EC:5, EC:6) in Supporting Information Figs. S8−S12. These results indicate that the model can identify key residues within enzyme amino acid sequences during training, even without explicitly being trained with supervised signals.

3.4.4. ERAM demonstrates significant potential for annotating the binding sites of understudied enzymes

However, many enzymes still lack high-quality binding site annotations, and it is both complex and costly to annotate these

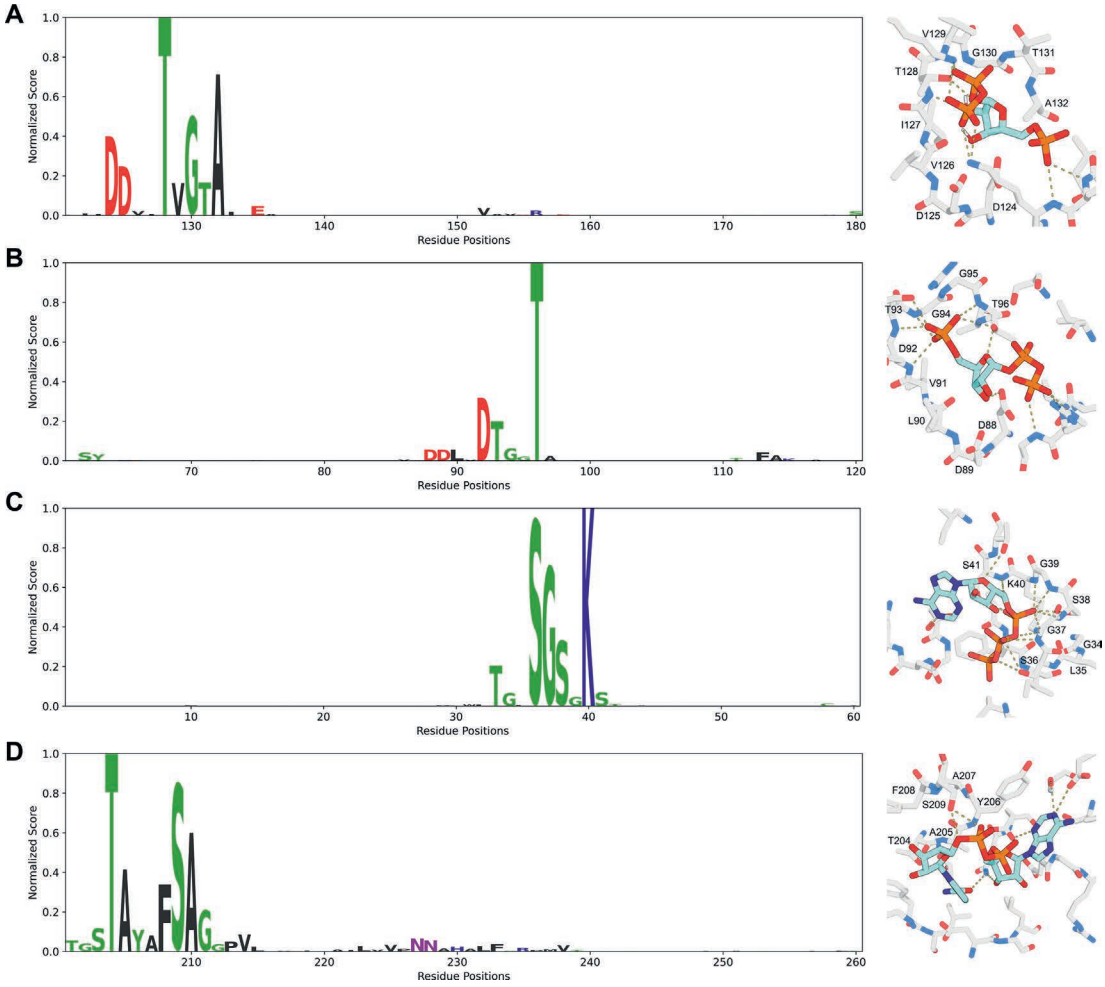

Figure 6 Highlighted amino acid residues by our model *vs*. annotated binding sites. On the left, the enzyme amino acid sequence logos provide a visualization of the attention scores, while the right side displays the binding sites of enzyme, as annotated in UniProtKB. (A, B) Phosphoribosyltransferase (Uniprot ID: A1AXP4, B5BDQ2, EC:2.4.2.10). (C) Adenylyl-sulfate kinase (Uniprot ID: A6KXG9, EC:2.7.1.25). (D) NAD kinase (Uniprot ID: Q49897, EC:2.7.1.23).

binding sites by experiment. Therefore, we utilize attention scores from the enzyme encoder to predict binding sites for proteins that are unannotated in UniProtKB. The protein under study, Squalene synthase (UniProt ID: A0A1D8PI71, EC:2.5.1.21), functions as a component of the third module in the ergosterol biosynthesis pathway, covering the pathway's later steps[53,54]. In the absence of a crystal structure, we employ AlphaFold2[55] to predict the protein's structure. Subsequently, we employ AutoDock Vina[56] to dock the protein with NADP and visualize the results using PyMOL. As depicted in Fig. 7, the amino acid residues within the binding pocket correspond to those with high attention scores in ERAM. To further validate the accuracy of our prediction, we employ the Basic Local Alignment Search Tool (BLAST)[57] to identify conserved domains within the protein. Consequently, the protein A0A1D8PI71 is identified as a member of the isoprenoid biosynthesis enzyme family. Residues Y178, A183, V186, G187, L190, G216, L219, and R226 are experimentally validated conserved sites and constitute the binding pocket[58-60], aligning with the residues which ERAM assign high attention scores. Furthermore, we provide additional examples of successful site annotation by ERAM attention scores with experimental validation in Supporting Information Figs. S13−S15. For instance, the amino acid residues of protein Q2FGD0, which have high attention scores, correspond to the conserved sites in nicotinate-nicotinamide nucleotide adenylyltransferase[61-63]. Similarly, the amino acid residues of protein Q2FSD3, scoring high in attention, correspond to the conserved sites in phosphoribosyl transferase[64,65]. These observations substantiate the model's comprehensive grasp of the enzymatic catalysis mechanism.

3.5. Small molecule encoder attention visualization

To further elucidate how ERAM comprehends enzymatic reactions, we visualize the attention scores generated by the small molecule encoder. The visualization results reveal that the small molecule encoder has learned to pay more attention to the reaction sites within chemical reactions. As illustrated in Fig. 8, the model assigns high attention scores to sites of bond cleavage, such as RHEA:17653, RHEA:23436, and RHEA:14369. Additionally, the model selectively emphasizes hydroxyl groups in elimination reactions at RHEA:20196, RHEA:11920, and RHEA:12460. Further examples in Fig. 8 include sites of stereoisomer transformations at RHEA:12813 and carbon-carbon double bonds involved in addition reactions at RHEA:12462, among others.

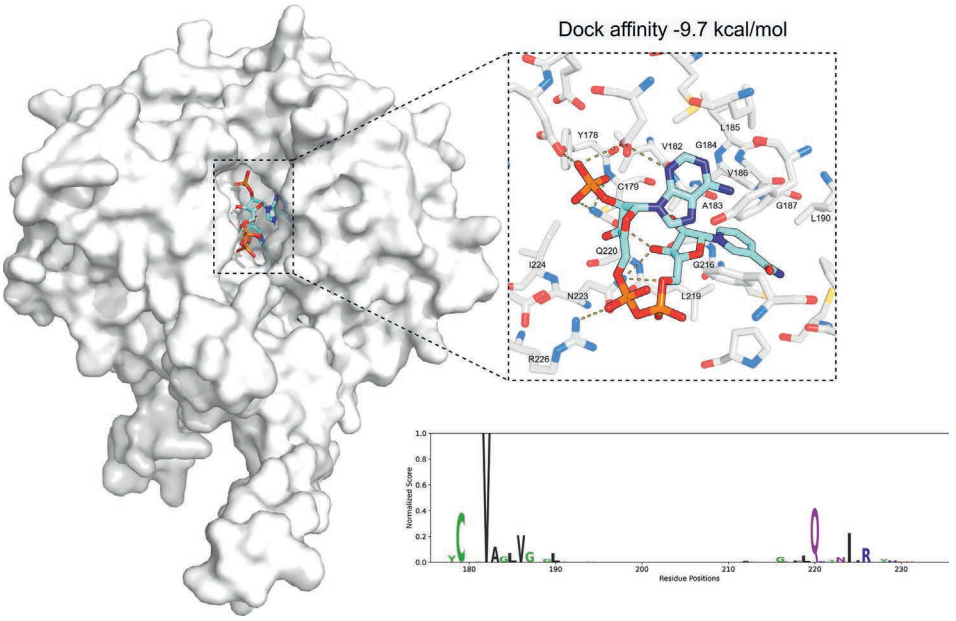

Figure 7 The docking result of A0A1D8PI71 with NADP and the amino acid residues with high attention scores. The amino acid residues exhibiting high attention scores are situated in the binding pocket. Furthermore, residues such as Y178, A183, V186, G187, L190, G216, L219, and R226, which exhibit high attention scores, are located within the conserved domain of the isoprenoid biosynthesis enzyme family.

This pattern indicates that the model possesses a comprehensive understanding of the underlying principles governing chemical reactions. By emphasizing key functional groups and reactive sites, the model effectively captures the dynamic nature of chemical transformations within molecules. This focused attention on critical reactive sites not only improves the model's predictive accuracy but also enhances its ability to generalize across various chemical contexts.

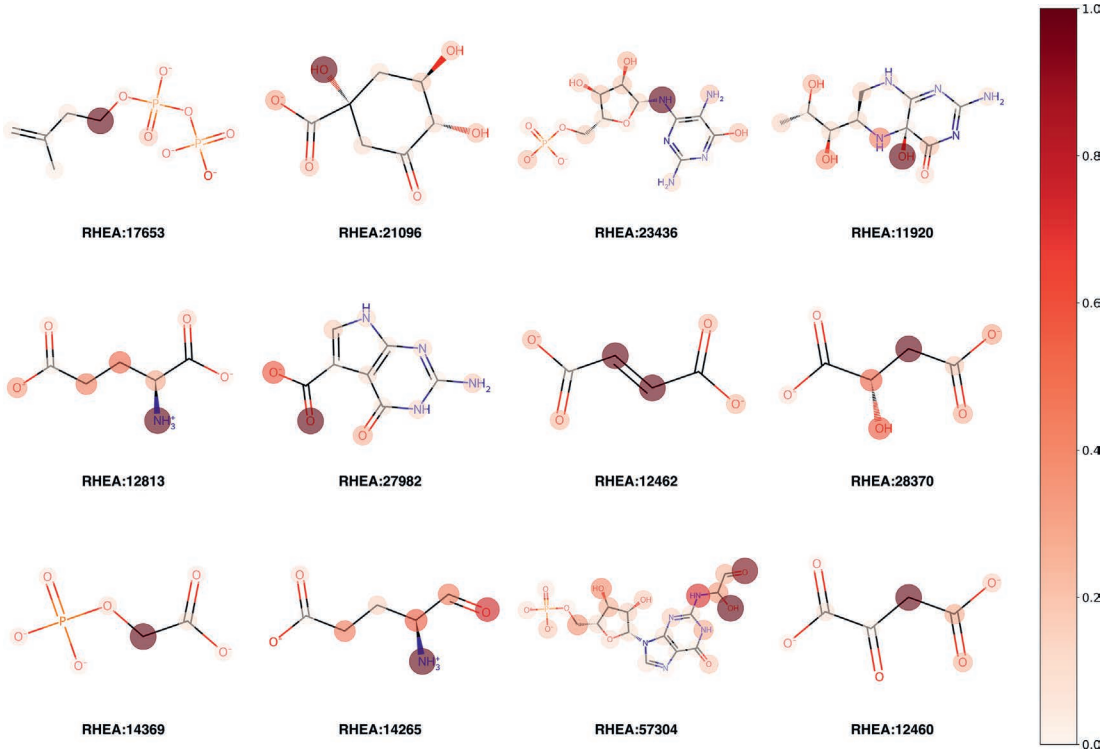

Figure 8 Visualization of small molecule encoder attention weight. The model allocates attention to the reactive sites where chemical reactions occur and to the important functional groups involved in the reactions.

4. Conclusions

This study presents the development of a task-agnostic Enzymatic-Reaction-Aware Molecular (ERAM) representation learning framework that aligns multimodal molecular representations with enzymatic reaction knowledge. The ERAM framework conceptualizes substrates, enzymes, and products involved in enzymatic reactions as triplets within a knowledge graph. This approach unifies the heterogeneous molecules involved in enzymatic reactions within a shared feature space. To further address the functional disparities between enzymes and chemical compounds, a dual-grained contrastive learning approach specifically tailored for enzymatic reactions is developed. The results of enzymatic reaction retrieval demonstrate that ERAM can accurately predict the feasibility and the ideal enzymes for a given reaction (and vice versa), based on the Euclidean distance in enzymatic reaction triplets. Furthermore, we have fine-tuned ERAM to predict the substrate scope of enzymes, outperforming the specific substrate predictor ESP. Moreover, the model exhibits commendable interpretability by assigning high attention scores to the binding sites of enzymes and the reaction sites of substrates, which results in a lower false positive rate and a higher overlap score compared to an unsupervised state-of-the-art binding site prediction model RXNAAMapper.

However, there is still room for improvement in ERAM. For example, incorporating the structural information of enzymes into the training process could help generate more representative embeddings, although this would require additional computational resources. Moreover, for negative sample construction, we employed dual-grained negative samples to model the enzymatic reaction. While it is feasible to generate more challenging and structured negative samples based on the hierarchical structure of EC numbers, this approach could help the model learn more nuanced embeddings.

In summary, our approach uniquely integrates knowledge from biochemistry with advanced machine learning techniques, thereby providing a deeper understanding of enzymatic functions. The model's success in these tasks underscores its significant potential in various biocatalysis-related applications. We have developed a web-based enzymatic reaction retrieval tool specifically designed to be accessible to researchers without coding experience, such as chemists and biologists. The platform is available at: http://cadd.zju.edu.cn/eram/. We believe that our model will serve as a valuable tool for advancing research in biocatalysis, offering insights and capabilities that can facilitate breakthroughs in enzyme engineering, drug discovery, and synthetic biology.

Acknowledgments

The work described in this paper was supported in part by the Young Scientists Fund of National Natural Science Foundation of China (62406295), by the National Natural Science Foundation of China (82373791, 22220102001), and by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project T45-401/22-N.

Author contributions

Lanqing Li and Yuansheng Huang designed the model architecture. Yuansheng Huang collected the data, developed the models, and analyzed the results. Yuansheng Huang and Lanqing Li wrote the manuscript. Wenjia Qian, Jiahui Yu, Huifeng Zhao, Xiaorui Wang, Odin Zhang, Guangyong Chen, and Shukai Gu contributed to result interpretation through constructive discussions. Pheng-Ann Heng, Tingjun Hou, and Yu Kang conceived and supervised the project, interpreted the results, and contributed to manuscript writing.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data availability

The data of this study is available at https://drive.google.com/file/d/1o-i4cl2u5j6cL5RDbutAeoQTuZxpD6ND/view?usp=sharing.

Code availability

The source code of this study is freely available at GitHub (https://github.com/YuanshengH/Dual-Enzy) to allow replication of the results.

Appendix A. Supporting information

Supporting Information to this article can be found online at https://doi.org/10.1016/j.apsb.2026.03.052.

References

1. Choi JM, Han SS, Kim HS. Industrial applications of enzyme biocatalysis: current status and future aspects. *Biotechnol Adv* 2015;**33**:1443−54.
2. Devine PN, Howard RM, Kumar R, Thompson MP, Truppo MD, Turner NJ. Extending the application of biocatalysis to meet the challenges of drug development. *Nat Rev Chem* 2018;**2**:409−21.
3. Basso A, Serban S. Industrial applications of immobilized enzymes—a review. *Mol Catal* 2019;**479**:110607.
4. Bell EL, Finnigan W, France SP, Green AP, Hayes MA, Hepworth LJ, et al. Biocatalysis. *Nat Rev Methods Primers* 2021;**1**:46.
5. Kuchner O, Arnold FH. Directed evolution of enzyme catalysts. *Trends Biotechnol* 1997;**15**:523−30.
6. Kirk O, Borchert TV, Fuglsang CC. Industrial enzyme applications. *Curr Opin Biotechnol* 2002;**13**:345−51.
7. Currin A, Swainston N, Day PJ, Kell DB. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem Soc Rev* 2015;**44**:1172−239.
8. Schomburg I, Chang A, Placzek S, Söhngen C, Rother M, Lang M, et al. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res* 2012;**41**:D764−72.
9. Nilsson A, Nielsen J, Palsson BO. Metabolic models of protein allocation call for the kinetome. *Cell Syst* 2017;**5**:538−41.
10. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res* 2023;**51**:D523−31.
11. Tkatchenko A. Machine learning for chemical discovery. *Nat Commun* 2020;**11**:4125.
12. Mou Z, Eakes J, Cooper CJ, Foster CM, Standaert RF, Podar M, et al. Machine learning-based prediction of enzyme substrate scope: application to bacterial nitrilases. *Proteins* 2021;**89**:336−47.
13. Yang M, Fehl C, Lees KV, Lim EK, Offen WA, Davies GJ, et al. Functional and informatics analysis enables glycosyltransferase activity prediction. *Nat Chem Biol* 2018;**14**:1109−17.
14. Van De Waterbeemd H, Gifford E. ADMET *in silico* modelling: towards prediction paradise?. *Nat Rev Drug Discov* 2003;**2**:192−204.

15. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature* 2018;**559**: 547−55.
16. Dong J, Wang NN, Yao ZJ, Zhang L, Cheng Y, Ouyang D, et al. ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J Cheminform* 2018;**10**:29.
17. Kim GB, Kim JY, Lee JA, Norsigian CJ, Palsson BO, Lee SY. Functional annotation of enzyme-encoding genes using deep learning with transformer layers. *Nat Commun* 2023;**14**:7370.
18. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*; 2017. Long Beach, CA.
19. Teukam YGN, Dassi LK, Manica M, Probst D, Schwaller P, Laino T. Language models can identify enzymatic binding sites in protein sequences. *Comput Struct Biotechnol J* 2024;**23**:1929−37.
20. Devlin J. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* 2018. Available from: https://doi.org/10.48550/arXiv.1810.04805.
21. Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MK, et al. Deep learning-based k cat prediction enables improved enzyme-constrained model reconstruction. *Nat Catal* 2022;**5**:662−72.
22. Kroll A, Ranjan S, Engqvist MK, Lercher MJ. A general model to predict small molecule substrates of enzymes based on machine and deep learning. *Nat Commun* 2023;**14**:2787.
23. Meier J, Rao R, Verkuil R, Liu J, Sercu T, Rives A. Language models enable zero-shot prediction of the effects of mutations on protein function. In: *Proceedings of the 35th international conference on neural information processing systems (NeurIPS 2021)*. Virtual Event; 2021.
24. Yu T, Cui H, Li JC, Luo Y, Jiang G, Zhao H. Enzyme function prediction using contrastive learning. *Science* 2023;**379**:1358−63.
25. Ryu JY, Kim HU, Lee SY. Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. *Proc natl acad Sci USA* 2019;**116**:13996−4001.
26. Shen X, Zhang S, Long J, Chen C, Wang M, Cui Z, et al. A highly sensitive model based on graph neural networks for enzyme key catalytic residue prediction. *J Chem Inform Model* 2023;**63**:4277−90.
27. Wang X, Yin X, Jiang D, Zhao H, Wu Z, Zhang O, et al. Multi-modal deep learning enables efficient and accurate annotation of enzymatic active sites. *Nat Commun* 2024;**15**:7348.
28. Liu Y, Hua C, Zeng T, Rao J, Zhang Z, Wu R, et al. EnzymeCAGE: a geometric foundation model for enzyme retrieval with evolutionary insights. *bioRxiv* 2024. Available from: https://doi.org/10.1101/2024.12.15.628585.
29. Hua C, Zhong B, Luan S, Hong L, Wolf G, Precup D, et al. Reactzyme: a benchmark for enzyme-reaction prediction. In: *Proceedings of the 38th international conference on neural information processing systems (NeurIPS 2024)*; 2024. Vancouver, Canada.
30. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS 2013)*. NV, USA: Lake Tahoe; 2013.
31. Krajewski WW, Collins R, Holmberg-Schiavone L, Jones TA, Karlberg T, Mowbray SL. Crystal structures of mammalian glutamine synthetases illustrate substrate-induced conformational changes and provide opportunities for drug and herbicide design. *J Mol Biol* 2008; **375**:217−28.
32. Koshland Jr DE. The key−lock theory and the induced fit theory. *Angew Chem Int Ed Engl* 1995;**33**:2375−8.
33. Yang J, Mora A, Liu S, Wittmann B, Anandkumar A, Arnold F, et al. Care: a benchmark suite for the classification and retrieval of enzymes. In: *Proceedings of the 38th international conference on neural information processing systems (NeurIPS 2024)*; 2024. Vancouver, Canada.
34. Berg JM, Tymoczko JL, Stryer L. *Biochemistry*. 6th ed. New York: W.H. Freeman and Company; 2007.
35. Wang H, Li W, Jin X, Cho K, Ji H, Han J, et al. Chemical-reaction-aware molecule representation learning. *arXiv preprint* 2021. Available from: https://doi.org/10.48550/arXiv.2109.09888.
36. Zeng L, Li L, Li J. Molkd: distilling cross-modal knowledge in chemical reactions for molecular property prediction. *arXiv* 2023. Available from: https://doi.org/10.48550/arXiv.2305.01912.
37. Grill J-B, Strub F, Altché F, Tallec C, Richemond P, Buchatskaya E, et al. Bootstrap your own latent—a new approach to self-supervised learning. In: *Proceedings of the 34th international conference on neural information processing systems (NeurIPS 2020)*. Virtual Event; 2020.
38. Gao Y, Zhuang J-X, Lin S, Cheng H, Sun X, Li K, et al. Disco: remedying self-supervised learning on lightweight models with distilled contrastive learning. In: *Proceedings of the 17th European conference on computer vision (ECCV 2022)*; 2022. Tel Aviv, Israel.
39. Vaessen N, van Leeuwen DA. The effect of batch size on contrastive self-supervised speech representation learning. *arXiv* 2024. Available from: https://doi.org/10.48550/arXiv.2402.13723.
40. Bansal P, Morgat A, Axelsen KB, Muthukrishnan V, Coudert E, Aimo L, et al. Rhea, the reaction knowledgebase in 2022. *Nucleic Acids Res* 2022;**50**:D693−700.
41. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017; **35**:1026−8.
42. Li T, Cui X, Cui Y, Sun J, Chen Y, Zhu T, et al. Exploration of transaminase diversity for the oxidative conversion of natural amino acids into 2-ketoacids and high-value chemicals. *ACS Catal* 2020;**10**: 7950−7.
43. Goldman S, Das R, Yang KK, Coley CW. Machine learning modeling of family wide enzyme-substrate specificity screens. *PLoS Comput Biol* 2022;**18**:e1009853.
44. Bergstra J, Yamins D, Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: *Proceedings of the 30th international conference on machine learning (ICML 2013)*; 2013. Atlanta, GA, USA.
45. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*; 2016. San Francisco, CA, USA.
46. Kaur J, Sharma R. Directed evolution: an approach to engineer enzymes. *Crit Rev Biotechnol* 2006;**26**:165−99.
47. Powell KA, Ramer SW, Del Cardayré SB, Stemmer WP, Tobin MB, Longchamp PF, et al. Directed evolution and biocatalysis. *Angew Chem Int Ed* 2001;**40**:3948−59.
48. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein−ligand interaction profiler. *Nucleic Acids Res* 2015;**43**:W443−7.
49. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**:2579−605.
50. Schwaller P, Probst D, Vaucher AC, Nair VH, Kreutter D, Laino T, et al. Mapping the space of chemical reactions using attention-based neural networks. *Nat Mach Intell* 2021;**3**:144−52.
51. Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, et al. ProtTrans: towards cracking the language of life's code through self-supervised learning. *IEEE Trans Pattern Anal Mach Intell* 2021; **44**:7112−27.
52. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer EL, et al. Pfam: the protein families database in 2021. *Nucleic Acids Res* 2021;**49**:D412−9.
53. Fegueur M, Richard L, Charles A, Karst F. Isolation and primary structure of the ERG9 gene of Saccharomyces cerevisiae encoding squalene synthetase. *Curr Genet* 1991;**20**:365−72.
54. Jennings SM, Tsay YH, Fisch TM, Robinson GW. Molecular cloning and characterization of the yeast gene for squalene synthetase. *Proc Natl Acad Sci U S A* 1991;**88**:6038−42.

55. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583−9.
56. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2. 0: new docking methods, expanded force field, and python bindings. *J Chem Inf Model* 2021;**61**:3891−8.
57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;**215**:403−10.
58. Pandit J, Danley DE, Schulte GK, Mazzalupo S, Pauly TA, Hayward CM, et al. Crystal structure of human squalene synthase: a key enzyme in cholesterol biosynthesis. *J Biol Chem* 2000;**275**:30610−7.
59. Gu P, Ishii Y, Spencer TA, Shechter I. Function-structure studies and identification of three enzyme domains involved in the catalytic activity in rat hepatic squalene synthase. *J Biol Chem* 1998;**273**: 12515−25.
60. Tansey TR, Shechter I. Structure and regulation of mammalian squalene synthase. *Biochim Biophys Acta* 2000;**1529**:49−62.
61. Zhou T, Kurnasov O, Tomchick DR, Binns DD, Grishin NV, Marquez VE, et al. Structure of Human Nicotinamide/Nicotinic Acid Mononucleotide Adenylyltransferase: basis for the dual substrate specificity and activation of the oncolytic agent tiazofurin. *J Biol Chem* 2002;**277**:13148−54.
62. Magni G, Amici A, Emanuelli M, Orsomando G, Raffaelli N, Ruggieri S. Structure and function of nicotinamide mononucleotide adenylyltransferase. *Curr Med Chem* 2004;**11**:873−85.
63. Zhang X, Kurnasov OV, Karthikeyan S, Grishin NV, Osterman AL, Zhang H. Structural characterization of a human cytosolic NMN/NaMN adenylyltransferase and implication in human NAD biosynthesis. *J Biol Chem* 2003;**278**:13503−11.
64. Sinha SC, Smith JL. The PRT protein family. *Curr Opin Struct Biol* 2001;**11**:733−9.
65. Musick WDL, Nyhan WL. Structural features of the phosphoribosyltransferases and their relationship to the human deficiency disorders of purine and pyrimidine metabolism. *Crit Rev Biochem* 1981;**11**:1−34.