

ENZYME-UNIFIED: LEARNING HOLISTIC REPRESENTATIONS OF ENZYME FUNCTION WITH A HYBRID INTERACTION MODEL

Anonymous authors

Paper under double-blind review

ABSTRACT

Predicting diverse functional properties of enzymes is a crucial challenge in biotechnology. Current machine learning approaches often fall short due to two key limitations: they predict properties in isolation using simplistic feature concatenation, thereby missing crucial inter-property relationships, and their performance is frequently overestimated on biased, homology-unaware datasets. To overcome these challenges, we introduce ENZYME-UNIFIED, a unified framework built upon a single, powerful architecture. We train distinct instances of our model to predict a comprehensive suite of five key properties: turnover number, Michaelis constant, catalytic efficiency, optimal temperature, and optimal pH. At its core is our novel Hybrid Interaction Model, which dynamically integrates fine-grained local interactions via cross-attention with global feature representations through a trainable gate, enabling a more holistic representation of enzyme function. For robust evaluation, we developed three new large-scale, sequence-dissimilar datasets. Our experiments show that ENZYME-UNIFIED achieves state-of-the-art results and significantly outperforms previous models on our rigorously curated datasets, demonstrating the powerful synergy of its hybrid architecture. The code and dataset will be available upon acceptance.

1 INTRODUCTION

Enzymes are nature’s catalysts, driving biochemical reactions that are fundamental to life sciences and a cornerstone of modern industry, from pharmaceuticals to biofuel production. (Wu et al., 2021; Zhang et al., 2015; Guldhe et al., 2015) Harnessing their full potential, however, requires tailoring these biocatalysts to the specific demands of industrial applications, a task for which natural enzymes are rarely optimized. A central goal of enzyme engineering is to design or discover novel catalysts with desired functional properties. Key metrics such as catalytic rate (k_{cat}), the Michaelis constant (K_m), catalytic efficiency (k_{cat}/K_m), optimal temperature (T_{opt}), and optimal pH dictate an enzyme’s utility. However, the experimental characterization of these properties is often a laborious, costly, and low-throughput process. In recent years, deep learning has emerged as a powerful paradigm to predict these properties from sequence and substrate information, promising to accelerate computational enzyme design. This progress is largely driven by advances in protein language models, which learn deep, context-aware representations of protein sequences from vast unlabeled databases (Lin et al., 2023; Elnaggar et al., 2022). For instance, models like UniKP have shown success in predicting key kinetic parameters from sequence and substrate information (Zheng et al., 2023), while others, such as PRIME, have focused on environmental properties like thermostability (Linder et al., 2024), promising to replace slow experimental cycles with rapid in silico screening.

Despite promising progress, current approaches are plagued by two fundamental limitations. First, they are fragmented into a single-task research paradigm, where models are designed to predict a single property in isolation. This disregards the intricate biophysical interplay between an enzyme’s catalytic activity and the environmental conditions in which it functions. Moreover, this fragmentation often leads to models being evaluated on improperly split datasets that fail to account for sequence homology. Consequently, reported performance can be artificially inflated, masking poor generalization to novel enzymes, which is a critical failure for real-world discovery applications.

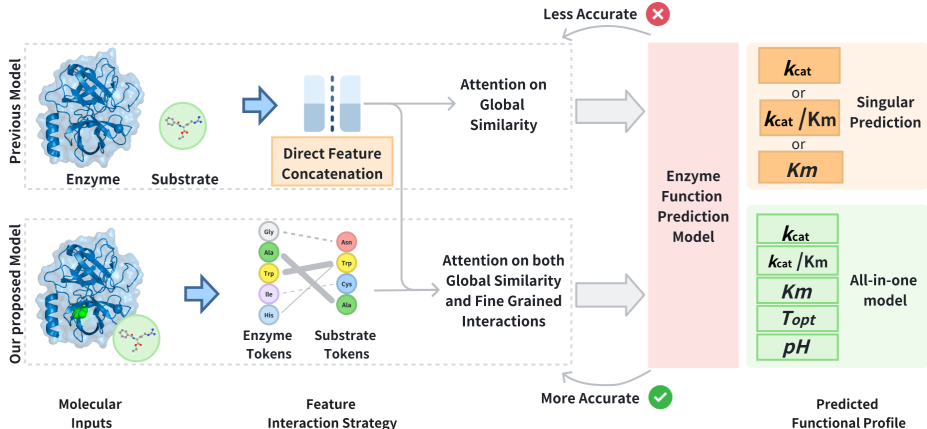
054
055
056
057
058
059
060
061
062
063
064
065
066
067
068069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091

Figure 1: Conceptual comparison of our framework with previous models. **(Top)** Prior methods, based on coarse global features, yield fragmented singular predictions. **(Bottom)** By capturing fine-grained interactions, our model provides a holistic, *all-in-one* prediction of the enzyme’s entire functional profile, including kinetics and environmental optima.

This fragmentation significantly curtails the practical utility of predictive models. A highly active enzyme (high k_{cat}/K_m) is of little industrial value if its optimal conditions (T_{opt} , pH) are impractically narrow or extreme (Bornscheuer et al., 2012). To guide meaningful experiments, scientists require a holistic view of an enzyme’s performance profile (Choi et al., 2019). Therefore, a unified framework that can co-predict multiple, interdependent properties, a challenge well-suited for a multi-task learning paradigm, which has proven effective for improving generalization and data efficiency in modern deep learning (Zhang & Yang, 2021), is not merely an incremental improvement, but a necessary step towards building truly applicable computational tools.

To bridge this gap, we introduce ENZYME-UNIFIED, a multi-task learning framework designed to provide a comprehensive and robust profile of enzyme function by simultaneously predicting catalytic rate (k_{cat}), the Michaelis constant (K_m), catalytic efficiency (k_{cat}/K_m), optimal temperature (T_{opt}), and optimal pH. At the architectural level, we address another key weakness of prior work—the reliance on simple feature concatenation. We propose a novel Hybrid Interaction Model that dynamically integrates two distinct paradigms: a global feature concatenation network and a token-level cross-attention network designed to capture fine-grained molecular interactions. This latter approach has shown significant promise for explicitly modeling interactions between distinct molecular entities, such as proteins and small molecules (Huang et al., 2021). Crucially, our model adaptively balances these two pathways through a single, end-to-end trainable gating parameter, allowing it to learn the optimal feature abstraction strategy for a given enzyme-substrate pair.

092
093
094
095
096
097
098
099
100
101
102

Our primary contributions are summarized as follows:

- We propose ENZYME-UNIFIED, a unified framework for holistic enzyme property prediction, powered by a novel HYBRID INTERACTION MODEL that adaptively fuses global and local interaction features for more powerful and flexible representations.
- We construct three new large-scale, rigorously partitioned datasets for multi-property prediction, which are carefully split by sequence similarity to ensure unbiased evaluation and promote research into model generalization.
- Our framework achieves state-of-the-art performance on the public, unbiased CataPro benchmark and significantly outperforms specialized single-task baselines across all properties on our new datasets.

103
104
105
106
107

2 RELATED WORK

From Single-Task to Holistic Prediction. Deep learning can handle multiple tasks simultaneously (Zhang & Yang, 2021; Zeng et al., 2025). Early machine learning models laid a foundation by predicting isolated enzyme properties, such as catalytic rate (k_{cat}) or substrate affinity (K_m) (Li

et al., 2022; Kroll et al., 2021), while a parallel line of research targeted thermostability (T_m) (Linder et al., 2024). These single-task approaches, however, fail to capture intricate biophysical trade-offs. Recognizing this, recent models such as UniKP advanced towards jointly predicting kinetic parameters (k_{cat} , K_m) (Zheng et al., 2023). Still, a critical gap remains: these models treat environmental factors like pH and temperature as conditional inputs rather than properties to be predicted. This prevents the modeling of a holistic functional profile, including an enzyme’s apparent optimal conditions (T_{opt} and optimal pH) ¹. In contrast, our work, ENZYME-UNIFIED, is to our knowledge the first framework to treat an enzyme’s kinetics and its optimal conditions as coequal targets, thus providing a truly holistic functional profile.

The Imperative for Unbiased Evaluation. The field has increasingly recognized that model performance on randomly split benchmarks is often an illusion, inflated by “data leakage” from sequence homology (Rao et al., 2019; Jumper et al., 2021). Recent works like CataPro (Yu et al., 2025) and CatPred (Gao et al., 2025) have established a new, more rigorous standard by introducing benchmarks with strict, non-homologous data splits and a focus on out-of-distribution generalization. Our work fully adheres to this philosophy and the demand for rigorous, unbiased evaluation. We validate our model’s performance on the public CataPro benchmark and further contribute to this effort by developing and releasing three new, large-scale datasets partitioned by a strict sequence similarity threshold to facilitate robust, multi-property model evaluation.

Architectures for Enzyme-Substrate Interaction. At the architectural level, a common strategy involves encoding enzyme and substrate using pre-trained models like ProtT5 (Elnaggar et al., 2022) and SMILES transformers (Kreutter et al., 2021), respectively. However, the fusion of these representations is often accomplished via simple concatenation (Zheng et al., 2023; Yu et al., 2025; Gao et al., 2025), which may not adequately capture fine-grained, token-level interactions at the active site. We directly address the limitation for modeling enzyme-substrate interactions with a novel Hybrid Interaction Model, which moves beyond static concatenation by adaptively fusing global features with detailed, token-level interactions learned via a trainable cross-attention mechanism. Beyond these directions, several recent works have explored enzyme–substrate modeling for different problem settings. For example, Yu et al. (2023) leverages contrastive learning and multi-level feature aggregation to tackle enzyme function annotation. EnzymeCAGE Liu et al. (2024) focuses on hierarchical multi-label prediction to improve multi-label enzyme function prediction. EZSpecificity Cui et al. (2025) treats reactivity as a ranking problem to select likely substrates. Inspired by these works, we develop ENZYME-UNIFIED.

3 METHOD

Our methodology is structured into three key components. We begin in Section 3.1 by detailing the creation and rigorous, non-homologous partitioning of three new, large-scale datasets, which form the foundation for both training and unbiased evaluation. Next, in Section 3.2, we introduce the ENZYME-UNIFIED framework, a novel hybrid architecture designed to overcome the limitations of previous models by holistically predicting multiple enzyme properties. We describe its multi-modal input representations and its core hybrid interaction model. Finally, Section 3.3 outlines the objective function and training protocol used to optimize the framework for each predictive task.

3.1 DATASET CONSTRUCTION AND UNBIASED PARTITIONING FOR MULTI-TASK LEARNING

To rigorously train and evaluate our multi-task framework, we first established a suite of three new, large-scale datasets for predicting catalytic efficiency (k_{cat}/K_m), optimal temperature (T_{opt}), and optimal pH. Recognizing that simple random splits are insufficient for robust model evaluation due to sequence homology, which can lead to overly optimistic performance estimates (Rao et al., 2019), we implemented a meticulous, cluster-based partitioning strategy from the outset.

Our standardized protocol ensures a reproducible and unbiased partitioning for all datasets. The protocol employs MMseqs2 (Steinegger & Söding, 2017) to cluster all unique protein sequences based on a strict 40% sequence identity threshold, which guarantees that highly similar proteins are

¹Importantly, in our setting, the labels for these optimal conditions are drawn from heterogeneous assay measurements. Where multiple optima are reported for the same enzyme–substrate pair under varying experimental setups, we resolve them by taking the median within that pair (but never across different substrates).

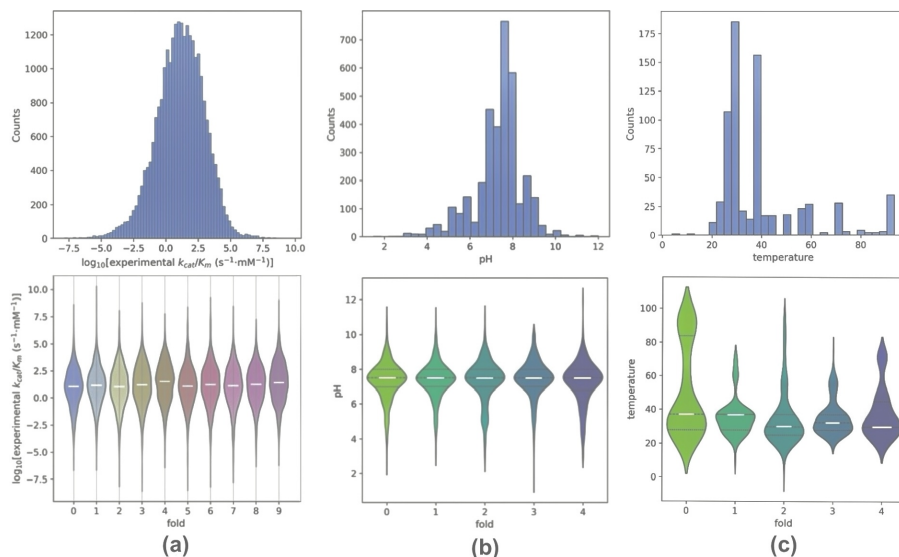


Figure 2: Statistical distribution of the curated datasets. The figure displays the overall distribution (top) and the per-fold distribution (bottom) for (a) catalytic efficiency (k_{cat}/K_m), (b) optimal pH, and (c) optimal temperature (T_{opt}). The consistency across folds demonstrates our balanced, cluster-based data partitioning.

grouped into the same cluster (Rost, 1999). Subsequently, we distribute these entire clusters, rather than individual data points, into distinct and non-overlapping folds. This assignment is performed greedily: clusters are sorted by size and then iteratively allocated to the fold with the fewest currently assigned samples to maintain balance. For more details, please refer to the Appendix E Applying this rigorous protocol, we construct and partition the following three datasets.

The Catalytic Efficiency (k_{cat}/K_m) Dataset. The construction of this dataset involved a comprehensive, multi-stage pipeline. We began by aggregating all entries containing k_{cat} or K_m values from the BRENDA (Chang et al., 2021) and SABIO-RK (Wittig et al., 2018) databases. Of 30,995 unique substrate names, 17,610 could not be reliably mapped to a SMILES string and were excluded. The remaining data were cleaned, and for entries with multiple annotations, we retained the maximum k_{cat} and minimum K_m to capture peak performance. After integrating all sources and performing a final deduplication, our dataset consists of 26,361 high-confidence samples. The \log_{10} -transformed values are approximately normally distributed, as is shown in Figure 2(a). Following our partitioning protocol, the sequences were grouped into 3,329 clusters and distributed into 10 balanced folds for cross-validation, with consistent value distributions across all folds.

The Optimal pH Dataset. Our curation for this dataset began with an initial collection of 6,521 raw entries. The data was subjected to a multi-stage filtering pipeline, removing entries without a valid substrate SMILES string or enzyme sequence. We retained the median value for enzymes with multiple reported optimal pH values. This pipeline yielded our final dataset of 3,435 high-quality samples exhibiting bimodal distribution, as shown in Figure 2(b). The sequences were grouped into 2,238 clusters and partitioned into 5 balanced folds, maintaining a consistent pH distribution across them.

The Apparent Temperature Optimum (T_{opt} , °C) Dataset. This dataset was constructed from an initial pool of 2,233 entries. We applied a stringent cleaning protocol, discarding records lacking essential information and taking the median value for enzymes with multiple reported optima. The resulting 706 unique entries represent the apparent temperature optimum, which reflects the balance between catalytic activity and thermal stability under the specific assay timescale and experimental matrix, rather than a purely intrinsic property of the enzyme. As shown in Figure 2(c), the values are concentrated around 30–40°C. The sequences were clustered into 251 groups and partitioned into 5 folds, with consistent value distributions across folds to ensure a fair evaluation.

This unified methodology of curation and cluster-based partitioning ensures that any given protein sequence and its close homologs are strictly confined to a single fold. Consequently, during cross-

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269

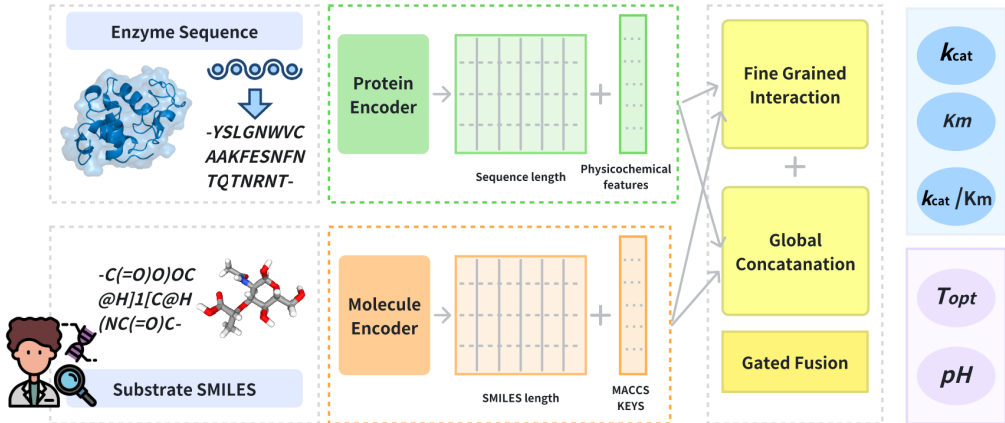


Figure 3: Architecture of the Enzyme-Unified hybrid framework. The framework processes encoded enzyme and substrate inputs through parallel pathways: a fine-grained interaction path to capture localized interactions, a global concatenation path for pooled features, and a gated fusion mechanism to predict the complete functional profile of kinetics and environmental optima dynamically.

validation, the model is always evaluated on enzymes that are significantly different from those seen during training. This forces the model to learn generalizable sequence-function relationships (Park et al., 2021) and provides a robust foundation for all experiments reported in this work.

3.2 THE ENZYME-UNIFIED FRAMEWORK

Current deep learning approaches for enzyme function are limited by a single-task paradigm that ignores the **intrinsic biophysical coupling between kinetics and environmental optima**, and by an oversimplified concatenation of global features that misses fine-grained interactions. To overcome these challenges, we propose ENZYME-UNIFIED, a framework built on a novel hybrid architecture. It moves beyond simple concatenation by explicitly modeling and dynamically balancing both global and fine-grained, token-level interactions to learn a more comprehensive representation. This allows our framework to provide a truly holistic functional profile by predicting five distinct properties together: k_{cat} , K_m , k_{cat}/K_m , T_{opt} , and pH.

3.2.1 MULTI-MODAL INPUT REPRESENTATION

The framework takes a multi-modal input, generating representations for both the enzyme and substrate from pre-trained language models and chemical fingerprints.

Enzyme Representation: We use the ProtT5-XL-U50 (Elnaggar et al., 2022) encoder to generate token-level embeddings from the amino acid sequence, denoted as $\mathbf{E} \in \mathbb{R}^{L \times 1024}$. We also use a pre-computed, mean-pooled global embedding, $\mathbf{h}_E \in \mathbb{R}^{1024}$.

Substrate Representation: The substrate’s SMILES string is encoded using the MolT5 (Edwards et al., 2022) encoder, yielding token-level embeddings $\mathbf{S} \in \mathbb{R}^{M \times 768}$. A global representation is formed by concatenating the pre-computed, mean-pooled MolT5 embedding ($\mathbf{h}_S \in \mathbb{R}^{768}$) with a 167-dimensional MACCS chemical fingerprint ($\mathbf{h}_{maccs} \in \mathbb{R}^{167}$) (Durant et al., 2002).

Auxiliary Physicochemical Features: As an optional input, we incorporate a 22-dimensional vector of pre-computed physicochemical properties of the enzyme, denoted as $\mathbf{h}_{phys} \in \mathbb{R}^{22}$. These descriptors, including amino acid composition, molecular weight, and isoelectric point, were calculated from the primary sequence using the ProtParam module of the Biopython library (Cock et al., 2009).

3.2.2 THE HYBRID INTERACTION ARCHITECTURE

Our key architectural innovation is a hybrid model comprising two distinct sub-models, or pathways, that capture different interaction philosophies. The final prediction is a learned interpolation of the outputs from these two pathways.

The Fine-Grained Interaction Pathway: This pathway is designed to model localized, molecular-level interactions through a stack of N bidirectional cross-attention layers. Let $\mathbf{E}^{(l)}$ and $\mathbf{S}^{(l)}$ be the token-level representations of the enzyme and substrate at layer l , respectively. With initial inputs $\mathbf{E}^{(0)}$ and $\mathbf{S}^{(0)}$ from projected embeddings of T5 encoders, output of each layer is defined as:

$$[\mathbf{E}^{(l)}, \mathbf{S}^{(l)}] = \text{BiCrossAttn}^{(l)}(\mathbf{E}^{(l-1)}, \mathbf{S}^{(l-1)}), \quad (1)$$

where the `BiCrossAttn` block internally computes parallel multi-head attention updates (enzyme-attending-to-substrate and vice-versa), followed by feed-forward networks. After the final layer, the contextualized representations $\mathbf{E}^{(N)}$ and $\mathbf{S}^{(N)}$ are pooled and combined with the projected MACCS fingerprint embedding \mathbf{h}'_{maccs} to form a final feature vector, \mathbf{z}_{attn} . A dedicated MLP then generates the prediction:

$$\mathbf{z}_{attn} = [\text{MeanPool}(\mathbf{E}^{(N)}); \text{MeanPool}(\mathbf{S}^{(N)}); \mathbf{h}'_{maccs}], \quad \hat{y}_{attn} = \text{MLP}_{attn}(\mathbf{z}_{attn}). \quad (2)$$

This hierarchical process allows the model to build progressively more sophisticated representations of the enzyme-substrate complex.

The Global Concatenation Pathway: This pathway reflects the traditional modeling approach by operating exclusively on global feature vectors. It takes the mean-pooled enzyme embedding (\mathbf{h}_E), the mean-pooled substrate embedding (\mathbf{h}_S), and the substrate’s MACCS fingerprint (\mathbf{h}_{maccs}). The language model embeddings are first independently normalized using `BatchNorm1d` (Ioffe & Szegedy, 2015) to stabilize distributions. These normalized features are then concatenated to form a single feature vector, \mathbf{z}_{cat} , which is processed by a dedicated MLP decoder to yield a prediction:

$$\mathbf{z}_{cat} = [\text{BatchNorm}(\mathbf{h}_E); \text{BatchNorm}(\mathbf{h}_S); \mathbf{h}_{maccs}], \quad \hat{y}_{cat} = \text{MLP}_{cat}(\mathbf{z}_{cat}). \quad (3)$$

This pathway serves as a robust baseline, capturing the overall context of the enzyme-substrate pair.

Gated Prediction Fusion: The final prediction of the hybrid model, \hat{y} , is a learned interpolation of the outputs from the two pathways. The balance is controlled by a single, globally trainable scalar parameter, α , allowing the optimization process to find an optimal, task-specific balance between the two modeling strategies. The final prediction is computed as:

$$\hat{y} = \sigma(\alpha) \cdot \hat{y}_{attn} + (1 - \sigma(\alpha)) \cdot \hat{y}_{concat}, \quad (4)$$

Where σ is the sigmoid function, this formulation allows the optimization process to find a single, optimal balance between the two modeling strategies for each predictive task.

3.3 OBJECTIVE FUNCTION AND TRAINING PROTOCOL

Our unified architectural framework is designed to support both specialized training (one model per task) and true multi-task training (shared encoder with multiple prediction heads). For the benchmark results reported in this paper, we adopt the specialized training protocol: a separate instance of the unified architecture is trained for each of the five target properties. This avoids the complexities of multi-task loss balancing and allows each model to fully optimize its regression objective.

Formally, each task is trained with a standard Mean Squared Error (MSE) loss. Because kinetic parameters span several orders of magnitude, we apply a task-specific transformation $T(y)$ before regression:

$$T(y) = \begin{cases} \log_{10}(y), & \text{if the task involves kinetic parameters } (k_{cat}, K_m, k_{cat}/K_m), \\ y, & \text{otherwise (pH, } T_{opt}). \end{cases} \quad (5)$$

Given predictions $\hat{y}_i = \mathcal{F}_\theta(x_i)$, the single-task objective is:

$$\mathcal{L}_{\text{single}}(\theta) = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - T(y_i))^2. \quad (6)$$

The models are trained using AdamW (Loshchilov & Hutter, 2019), with early stopping based on validation RMSE (evaluated on the original scale) to prevent overfitting.

Although our main results use specialized training, our architecture naturally supports a true multi-task regime. In the multi-task setting, the five task-specific losses are optimized jointly, with gradients flowing through a shared encoder and task-specific heads. The multi-task objective is:

$$\mathcal{L}_t = \frac{1}{N_t} \sum_{i=1}^N \mathbf{1}_{\{y_i^{(t)} \text{ available}\}} (\hat{y}_i^{(t)} - T_t(y_i^{(t)}))^2, \quad t \in \{k_{\text{cat}}, K_m, k_{\text{cat}}/K_m, \text{pH}, T_{\text{opt}}\}, \quad (7)$$

This formulation enables the encoder to capture shared biophysical representations across tasks, while allowing each head to model property-specific variations. Our expanded experiments demonstrate that the same unified architecture remains competitive or superior under both specialized and multi-task training protocols.

4 EXPERIMENTS

4.1 DATASETS AND PREPROCESSING

A robust and diverse collection of data is essential for training and validating our multi-task framework. Our strategy involved two key components: leveraging an established, high-quality benchmark for foundational kinetic tasks (k_{cat} , K_m), and constructing novel, rigorously partitioned datasets for the new tasks of catalytic efficiency and optimal environmental condition prediction (k_{cat}/K_m , pH, T_{opt}).

4.2 EXPERIMENTAL SETUP

All models were trained on 8 NVIDIA A100 GPU using the AdamW optimizer (Loshchilov & Hutter, 2019). We employed a consistent training protocol with a maximum of 100 epochs, using an early stopping mechanism that monitored validation RMSE with a patience of 25 epochs to prevent overfitting. Recognizing the distinct statistical characteristics of each dataset, we optimized key hyperparameters such as learning rate and batch size on a per-task basis to ensure a fair and robust evaluation.

To analyze how different sources of information contribute to multi-property prediction. **Hybrid**: our base architecture, which encodes enzymes using ProtT5 and fuses enzyme-substrate representations through direct feature concatenation and a trainable cross-attention module. **Hybrid + ProstT5**: this variant replaces ProtT5 with ProstT5, enabling the model to incorporate structure-aware protein representations and quantify the effect of structural information. **Hybrid++**: on top of the Hybrid design, we further introduce auxiliary physicochemical descriptors: a 22-dimensional vector computed from the primary sequence using Biopython’s ProtParam, including amino acid composition, molecular weight, isoelectric point, and related features. Together, these three variants allow us to disentangle the contributions of sequence encoders, structural priors, and handcrafted physicochemical features. To ensure fair comparison, we first pair our Hybrid architecture with ProtT5 and benchmark it against ProtT5-based baselines, demonstrating that the architectural design itself provides clear advantages. We then present results from the ProstT5 variant, showing that the same architecture can further improve performance when equipped with a more advanced encoder, highlighting its scalability and future extensibility. The complete, task-specific hyperparameter configurations are detailed in Appendix A.

4.3 MAIN RESULTS

We evaluated our ENZYME-UNIFIED hybrid architecture across all five predictive tasks, comparing its performance against a comprehensive suite of established baselines on both public benchmarks and our newly constructed datasets. The results consistently demonstrate the superiority of our hybrid approach, which not only establishes a new state-of-the-art in kinetic prediction but also shows a dominant advantage in modeling complex environmental dependencies.

Our base hybrid model, without any structural enhancements, immediately establishes itself in the top tier of predictors. For k_{cat} prediction, it outperforms the vast majority of prior methods and achieves a new **state-of-the-art PCC of 0.498**. We then investigated whether incorporating explicit

Table 1: Side-by-side performance comparison on the public CataPro benchmark datasets. Best results for each metric are highlighted in bold.

Model	k_{cat} dataset			K_m dataset		
	RMSE ↓	PCC ↑	SCC ↑	RMSE ↓	PCC ↑	SCC ↑
DLKcat	1.426	0.368	0.349	1.052	0.578	0.578
UniKP	1.382	0.448	0.449	1.039	0.599	0.594
ESM2_15B	1.431	0.435	0.467	1.063	0.581	0.567
ESM2_RDKitFP	1.380	0.435	0.424	1.025	0.609	0.607
ESM2_Morgan	1.370	0.446	0.436	1.017	0.617	0.616
ESM2_MACC	1.356	0.463	0.453	1.007	0.625	0.622
ESM2_MolT5	1.361	0.455	0.445	1.011	0.620	0.616
ESM2_MolT5-MACC	1.357	0.464	0.459	1.004	0.627	0.623
ProtT5_3B	1.441	0.443	0.434	1.068	0.603	0.519
ProtT5_RDKitFP	1.389	0.421	0.415	1.029	0.604	0.602
ProtT5_Morgan	1.364	0.465	0.462	1.015	0.621	0.620
ProtT5_Mole-BERT	1.341	0.483	0.481	1.014	0.617	0.611
ProtT5_MACC	1.338	0.486	0.482	1.011	0.621	0.616
ProtT5_MolT5	1.338	0.491	0.484	1.001	0.629	0.625
ProtT5-SaProt_MolT5	1.323	0.490	0.489	1.004	0.626	0.620
CataPro	1.329	0.497	0.495	0.998	0.633	0.629
CataPro+SaProt	1.317	0.493	0.489	0.999	0.631	0.625
ENZYME-UNIFIED (hybrid)	1.304	0.498	0.484	0.996	0.643	0.628
ENZYME-UNIFIED (hybrid+ProstT5)	1.299	0.496	0.494	0.997	0.639	0.631
ENZYME-UNIFIED (hybrid++)	1.301	0.495	0.419	0.999	0.643	0.624

structural information via ProstT5 embeddings could further boost performance. This enhanced model, hybrid+ProstT5, sets another **SOTA record with an RMSE of 1.299**.

Dominance on public kinetic benchmarks. To first validate the effectiveness of our architecture, we benchmarked it against 14 prior models on the public, non-homologous CataPro datasets for catalytic rate (k_{cat}) and the Michaelis constant (K_m). The comprehensive results are presented in Table 1.

Compared to previous large-scale pretrained models (ProtT5-3B and ESM2-15B), our model also demonstrates strong performance on the k_{cat} : its RMSE of 1.299 represents a 9.2% relative improvement over the ESM2-15B baseline (1.431). The advantage is even more pronounced for K_m prediction. Our base hybrid model significantly outperforms all 14 baselines, including the strong CataPro models, on two of the three key metrics, establishing a new SOTA with an **RMSE of 0.996** and a **PCC of 0.643**. The hybrid+ProstT5 variant further pushes the performance boundary, achieving the **best SCC of 0.631**. These results confirm that our hybrid architecture, by balancing global and fine-grained features, provides a more powerful and extensible foundation for enzyme kinetic prediction.

Predicting catalytic efficiency Directly. While predicting k_{cat} and K_m separately is a standard benchmark, the direct prediction of catalytic efficiency (k_{cat}/K_m) is often more relevant for enzyme engineering and represents a more complex learning task. (Fersht, 1985) We evaluated our framework on our newly constructed k_{cat}/K_m dataset, with results shown in Table 2.

On this challenging task, both variants of ENZYME-UNIFIED significantly outperform the strong UniKP and CataPro baselines. Our base hybrid model reduces the **RMSE by over 6.3%** compared to the next-best model (UniKP) and achieves the highest PCC of 0.409. The hybrid+ProstT5 variant demonstrates the best ranking correlation with an SCC of 0.421. This demonstrates our framework’s superior ability to learn the complex, non-linear relationship between sequence and catalytic efficiency on a challenging, rigorously partitioned dataset.

Superior performance on environmental property prediction. Our framework’s advantages are highly pronounced for environmental property prediction (Table 2). For optimal pH, our model achieves an RMSE of 0.840, a remarkable reduction of over 61% compared to the strong CataPro

Table 2: Comprehensive side-by-side performance comparison on our newly constructed datasets. For each model, we report performance across kinetics (k_{cat}/K_m), optimal pH, and optimal temperature (T_{opt}) prediction tasks. Best results within each column are highlighted in bold.

Model	k_{cat}/K_m Dataset			Optimal pH Dataset			Optimal T_{opt} Dataset		
	RMSE ↓	PCC ↑	SCC ↑	RMSE ↓	PCC ↑	SCC ↑	RMSE ↓	PCC ↑	SCC ↑
CataPro	1.716	0.368	0.366	2.167	0.521	0.427	17.553	0.661	0.495
UniKP	1.706	0.399	0.371	0.914	0.576	0.432	14.347	0.670	0.473
ProtT5-3B	1.745	0.359	0.347	0.989	0.523	0.507	16.342	0.686	0.502
ESM2-15B	1.642	0.337	0.381	0.949	0.579	0.512	12.170	0.697	0.582
ENZYME-UNIFIED (hybrid)	1.597	0.409	0.416	0.840	0.657	0.512	14.446	0.666	0.622
ENZYME-UNIFIED (hybrid+ProtT5)	1.601	0.403	0.421	0.901	0.613	0.510	12.433	0.703	0.612
ENZYME-UNIFIED (hybrid++)	1.599	0.408	0.419	0.912	0.626	0.511	10.329	0.783	0.640

Table 3: Side-by-side ablation study on the key architectural components of ENZYME-UNIFIED. The model is evaluated across three distinct datasets: K_{cat} , K_m , and pH. Best results for each metric within each dataset are highlighted in bold.

Model Configuration	K_{cat} Dataset			K_m Dataset			pH Dataset		
	RMSE ↓	PCC ↑	SCC ↑	RMSE ↓	PCC ↑	SCC ↑	RMSE ↓	PCC ↑	SCC ↑
Concatenation-only	1.344	0.461	0.449	1.002	0.630	0.625	2.034	0.562	0.459
Attention-only	1.373	0.443	0.417	1.010	0.633	0.629	1.450	0.591	0.503
Hybrid w/ Simple Average	1.359	0.452	0.433	1.006	0.632	0.627	1.742	0.577	0.481
ENZYME-UNIFIED (hybrid)	1.304	0.498	0.484	0.997	0.639	0.631	0.840	0.657	0.512

baseline, while increasing PCC by over 0.13. Compared to large-scale pretrained models such as ESM2-15B, our model’s PCC (0.657) is 13.5% better than the ESM2-15B baseline (0.579). On the challenging task of optimal temperature prediction, enhancing our model with amino acid physicochemical features (Kawashima et al., 2008; Gasteiger et al., 2005)(hybrid++) yields a dramatic performance leap, reducing the RMSE to an exceptional 10.329 and boosting the PCC to 0.783. Most strikingly, our model’s RMSE (10.329) is 17.5% better than the ESM2-15B baseline (12.170) and an impressive 36.8% better than the ProtT5-3B baseline (16.342). These results highlight our hybrid architecture’s effectiveness in modeling an enzyme’s environmental adaptation.

4.4 ABLATION STUDY

To dissect the contributions of our key architectural components and feature choices, we conducted a series of ablation studies, with the core results presented in Table 3. The study validates three foundational aspects of our design.

First, the results unequivocally demonstrate the **synergy of the hybrid architecture**. Our full ENZYME-UNIFIED model substantially outperforms both the “Concatenation-only” and “Attention-only” pathways when used in isolation, confirming that the fine-grained pathway provides crucial complementary information to the strong global feature baseline. Second, the importance of the learnable **Gated Fusion mechanism** is validated, as it consistently yields better performance than a simple, fixed averaging of the two pathways.

Finally, we confirm the framework’s ability to effectively leverage auxiliary information. While the benefit of some features can be task-dependent, incorporating amino acid physicochemical properties for T_{opt} prediction led to a **dramatic performance leap, reducing RMSE by nearly 41%** (detailed in Table 2). Collectively, these studies affirm that our hybrid design, adaptive fusion mechanism, and feature extensibility are essential for achieving state-of-the-art performance.

Beyond feature engineering, we also observe significant Cross-Task Synergy. As elaborated in Appendix C.2, pre-training the encoder on the Optimal pH task before fine-tuning on T_{opt} reduced the RMSE by 21.6% compared to training from scratch (14.17 → 11.11). This implies that ENZYME-UNIFIED captures fundamental, transferable biochemical determinants of protein stability that transcend individual tasks. Collectively, these studies affirm that our hybrid design, adaptive fusion mechanism, and capacity for transfer learning are essential for achieving state-of-the-art performance. A more detailed discussion is provided in Appendix D.

Table 4: Validating the fine-grained attention on RNase A’s catalytic histidines (His12, His119). Mutation of the critical H119 residue (H119A) specifically disrupts the local attention pattern. In contrast, the negative control mutation (S16A) shows no effect, confirming the model’s focus on functionally critical sites.

RNase A Variant	Region at His12	Region at His119
Wild-Type (WT)	12-12	116-119
H119A (Catalytic Mutant)	12-12	117-117
S16A (Negative Control)	12-12	116-119

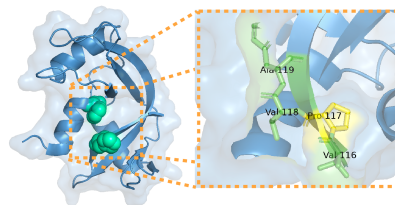


Figure 4: Interpretable attention on RNase A’s catalytic site. **(Left)** Wild-type RNase A with catalytic histidines. **(Right)** Mutating the critical His119 causes attention to collapse from the catalytic region (cf. Table 4) to a narrow focus on the adjacent Pro117, confirming sensitivity to the functional site.

4.5 CASE STUDY: INTERPRETING THE ATTENTION MECHANISM ON RIBONUCLEASE A

To validate that our model learns biochemically relevant principles, we conducted an in silico mutagenesis study on Ribonuclease A (RNase A) to probe the interpretability of our Fine-Grained Interaction Pathway (Rives et al., 2021). We hypothesized that the model’s attention should focus on the known catalytic histidines (His12, His119). (Crestfield et al., 1963). To test this, we compared the high-attention regions for the wild-type (WT) enzyme against a critical catalytic mutant (H119A) and a non-essential control mutant (S16A). As summarized in Table 4, our findings confirm this hypothesis, providing strong evidence that the model learns to identify functional sites.

The Model Accurately Identifies Catalytic Histidines in WT RNase A. For the WT enzyme, the model’s high-attention regions precisely pinpointed the locations of the two catalytic histidines. As shown in Table 4, the identified interaction regions include the segment 12-12 (His12) and the segment 116-119 (containing His119). This demonstrates that the model, without any explicit structural or functional labels, correctly learned to focus on the most important functional sites of the enzyme.

Mutation of a Key Histidine Specifically Disrupts Attention. Next, we assessed the effect of “knocking out” a critical residue (Smith, 1985; Wlodawer & Sjölin, 1982). Upon mutating the key catalytic histidine at position 119 to an alanine (H119A), the model’s attention pattern changed dramatically. The prominent, contiguous attention region spanning residues 116-119 in the WT enzyme was completely disrupted, collapsing to an isolated, single-residue focus at position 117-117. This strongly suggests that the model’s attention is not merely sequence-based but is directly correlated with the functional importance of the catalytic site (Vig et al., 2021).

The Negative Control Confirms Specificity. Finally, to confirm that this effect was specific to functional sites, we analyzed the S16A mutant. Serine 16 is not directly involved in catalysis (Raines, 1998). In this case, the mutation had a negligible impact on the attention patterns at the distant catalytic sites. The high-attention regions at 12-12 and 116-119 remained fully intact, identical to that of the WT enzyme. This result validates that the model’s response is highly specific and not a random artifact of sequence perturbation. [For more examples, please refer to the Appendix F.](#)

5 CONCLUSION

In this work, we introduced ENZYME-UNIFIED, a hybrid learning framework that addresses the fragmented, single-task paradigm by providing a holistic profile of an enzyme’s kinetic and environmental properties. Its novel hybrid architecture, validated by extensive ablation studies, synergistically combines global and fine-grained interaction pathways. Evaluated on three new, large-scale datasets with strict sequence-identity partitioning, our framework establishes a new state-of-the-art on kinetic benchmarks and shows dominant performance on environmental prediction. Crucially, a case study on RNase A confirmed the model’s biochemical interpretability, as its attention mechanism correctly identifies catalytic hotspots. ENZYME-UNIFIED thus provides a robust, extensible, and interpretable foundation for accelerating the design of novel biocatalysts.

REFERENCES

- 540
541
542 Uwe T. Bornscheuer, Gjalt W. Huisman, Romas J. Kazlauskas, Stefan Lutz, Jeffrey C. Moore, and
543 Karen Robins. Engineering enzymes for industrial applications. *Nature*, 485(7397):185–194,
544 2012.
- 545 Antje Chang, Lisa Jeske, Sarah Ulbrich, Johannes Hofmann, Julia Koblitz, Ida Schomburg, Meina
546 Neumann-Schaal, Dieter Jahn, and Dietmar Schomburg. BRENDA, the elixir core data resource
547 in 2021: new developments and updates. *Nucleic Acids Research*, 49(D1):D498–D508, 2021.
548 doi: 10.1093/nar/gkaa1025.
- 549
550 Kyoung-Rok Choi, Quang-Huy Le, and Dong-Yup Kim. Driving metabolic engineering and syn-
551 thetic biology with data-driven design. *Trends in Biotechnology*, 37(6):652–665, 2019.
- 552 Peter J.A. Cock, Tiago Antao, Jeffery T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke,
553 Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J.L. de Hoon.
554 Biopython: freely available Python tools for computational molecular biology and bioinformatics.
555 *Bioinformatics*, 25(11):1422–1423, 2009. doi: 10.1093/bioinformatics/btp163.
- 556
557 Arthur M. Crestfield, William H. Stein, and Stanford Moore. The preparation and enzymatic hy-
558 drolysis of ribonucleic acid from yeast. *The Journal of Biological Chemistry*, 238(7):2413–2420,
559 1963.
- 560 Haiyang Cui, Yufeng Su, Tanner J Dean, Tianhao Yu, Zhengyi Zhang, Jian Peng, Diwakar Shukla,
561 and Huimin Zhao. Enzyme specificity prediction using cross attention graph neural networks.
562 *Nature*, pp. 1–3, 2025.
- 563
564 Joseph L. Durant, B. K. Leland, D. R. Henry, and J. G. Nourse. Reoptimization of MDL keys for use
565 in drug discovery. *Journal of Chemical Information and Computer Sciences*, 42(6):1273–1280,
566 2002.
- 567
568 Carl Edwards, ChengXiang Zhai, and Heng Ji. MolT5: A self-supervised learning framework for
569 molecule-to-text generation. *Transactions on Machine Learning Research*, 2022. ISSN 2835-
570 8856. URL <https://openreview.net/forum?id=wVKjA12iGG>.
- 571 Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rihawi, Yu Wang, Llion Jones,
572 Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, Debsindhu Bhowmik, and
573 Burkhard Rost. Protrants: Toward cracking the language of life’s code through self-supervised
574 learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127,
575 2022.
- 576
577 Alan Fersht. *Enzyme Structure and Mechanism*. W. H. Freeman, New York, 1985.
- 578 Peng Gao, Ziyi Liu, Jingxuan Wang, Jie Zhang, Xiaohui Zheng, and Yan Li. Catpred: a compre-
579 hensive framework for in vitro enzyme kinetic parameter prediction with deep learning. *Nature*
580 *Communications*, 16(1):3281, 2025.
- 581
582 E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch.
583 Protein identification and analysis tools on the expasy server. In *The Proteomics Protocols Hand-*
584 *book*, pp. 571–607. Humana Press, 2005.
- 585
586 Abhishek Guldhe, Brijesh Singh, Taurai Mutanda, Kugen Permaul, and Faizal Bux. Advances in
587 synthesis of biodiesel via enzyme catalysis: Novel and sustainable approaches. *Renewable and*
Sustainable Energy Reviews, 41:1447–1464, 2015.
- 588
589 Kexin Huang, Cao Xiao, Lucas M. Glass, and Jimeng Sun. Moltrans: Molecular interaction trans-
590 former for drug-target interaction prediction. In *Proceedings of the 27th ACM SIGKDD Confer-*
591 *ence on Knowledge Discovery & Data Mining*, pp. 732–742, 2021.
- 592
593 Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by
reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine*
Learning, volume 37 of *PMLR*, pp. 448–456. PMLR, 2015.

- 594 John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger,
595 Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland,
596 Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-
597 Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman,
598 Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Se-
599 bastian Bodenstern, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Push-
600 meet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold.
601 *Nature*, 596(7873):583–589, 2021.
- 602 Shuichi Kawashima, Piotr Pokarowski, Maria Pokarowska, Andrzej Kolinski, Toshiaki Katayama,
603 and Minoru Kanehisa. AAindex: amino acid index database, progress report 2008. *Nucleic Acids*
604 *Research*, 36(Database issue):D202–D205, 2008.
- 605
606 Dennis Kreutter, Philippe Schwaller, and Jean-Louis Reymond. Predicting enzymatic reactions with
607 a molecular transformer. *Chemical Science*, 12(23):8052–8059, 2021.
- 608
609 Alexander Kroll, Martin K. M. Engqvist, David Heckmann, and Martin J. Lercher. Deep learning
610 allows genome-scale prediction of michaelis constants from structural features. *PLOS Biology*,
611 19(10):e3001402, 2021.
- 612 Feiran Li, Le Yuan, Hongzhong Lu, Ruslan Umarov, S. Aljoscha Wahl, Jens H. Nielsen, and Leif A.
613 T. W. Nielsen. Deep learning-based kcat prediction enables improved enzyme-constrained model
614 reconstruction. *Nature Catalysis*, 5(8):662–672, 2022.
- 615
616 Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin,
617 Robert Verkuil, Ory Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi,
618 Tom Sercu, Sal Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level
619 protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.
- 620 Johannes Linder, Vladimir Gligorijević, Konstantinos Tripsianes, Elaine C Saldivar, Tatjana
621 Schütze, Dominika Misiak, Vincent Pogenberg, Miroslava Schaffer, kim confession, Nikos Pinot-
622 sis, et al. A general-purpose protein design language model for stability and activity enhancement.
623 *Science Advances*, 10(26):eadr2641, 2024.
- 624
625 Yong Liu, Chenqing Hua, Tao Zeng, Jiahua Rao, Zhongyue Zhang, Ruibo Wu, Connor W Coley,
626 and Shuangjia Zheng. Enzymecage: a geometric foundation model for enzyme retrieval with
627 evolutionary insights. *bioRxiv*, pp. 2024–12, 2024.
- 628
629 Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Confer-*
ence on Learning Representations (ICLR), 2019.
- 630
631 Hyunjun Park, Gyoung S. Kim, and Sungroh Yoon. How to train your protein model: the effects of
632 data and data splits on protein representation learning. *Bioinformatics*, 38(1):82–89, 2021.
- 633
634 Ronald T. Raines. Ribonuclease A. *Chemical Reviews*, 98(3):1045–1065, 1998.
- 635
636 Roshan Rao, Nicholas Bhattacharya, Neil Thomas, Yan Duan, Xi Chen, John Canny, Pieter Abbeel,
637 and Yun S. Song. Evaluating protein transfer learning with TAPE. In *Advances in Neural Infor-*
mation Processing Systems 32 (NeurIPS 2019), 2019.
- 638
639 Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo,
640 Myle Ott, C. Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function
641 emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the*
National Academy of Sciences, 118(15):e2016239118, 2021.
- 642
643 Burkhard Rost. Twilight zone of protein sequence alignments. *Protein Engineering, Design and*
644 *Selection*, 12(2):85–94, 1999.
- 645
646 Michael Smith. In vitro mutagenesis. *Annual Review of Genetics*, 19(1):423–462, 1985.
- 647
648 Martin Steinegger and Johannes Söding. MMseqs2 enables sensitive protein sequence searching for
the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, 2017.

- 648 Jesse Vig, Ali Madani, Lav R. Varshney, C. peeling, Maruan Al-Shedivat, and T. N. Held. BERTol-
649 ogy meets biology: Interpreting attention in protein language models. In *International Conference*
650 *on Learning Representations (ICLR)*, 2021.
- 651
652 Ulrike Wittig, Ricardo Kania, Martin Golebiewski, Miriam Rey, Maryam Fazel, Samira Yari, Sarah
653 Mir, Evelyn Wetsch, Andreas Junker, Eida Alga, Andreas Weidemann, and Wolfgang Müller.
654 SABIO-RK: update on the ontology, concepts and associated database services. *Nucleic Acids*
655 *Research*, 46(D1):D633–D639, 2018. doi: 10.1093/nar/gkx1014.
- 656 Alexander Wlodawer and Lennart Sjölin. Structure of ribonuclease a at 1.26 Å resolution. *Biochem-*
657 *istry*, 21(12):2820–2832, 1982.
- 658
659 Sandy Wu, Radka Snajdrova, Jeffrey C. Moore, Kai Baldenius, and Uwe T. Bornscheuer. Biocatal-
660 ysis: Enzymatic synthesis for industrial applications. *Angewandte Chemie International Edition*,
661 60(1):88–119, 2021.
- 662 Likun Yu, Yuyang Wu, Yang Liu, and Jin Liu. Robust enzyme discovery and engineering by deep
663 learning with catapro. *Nature Communications*, 16(1):4062, 2025.
- 664
665 Tianhao Yu, Haiyang Cui, Jianan Canal Li, Yunan Luo, Guangde Jiang, and Huimin Zhao. Enzyme
666 function prediction using contrastive learning. *Science*, 379(6639):1358–1363, 2023.
- 667
668 Fanhu Zeng, Haiyang Guo, Fei Zhu, Li Shen, and Hao Tang. Robustmerge: Parameter-efficient
669 model merging for mllms with direction robustness. In *The Thirty-ninth Annual Conference on*
Neural Information Processing Systems, 2025.
- 670
671 Y-H. Percival Zhang, Jian Ge, and Zhaoping Liu. Enhanced activity of immobilized or chemically
672 modified enzymes. *ACS Catalysis*, 5(8):4503–4513, 2015.
- 673
674 Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and*
Data Engineering, 34(12):5586–5609, 2021.
- 675
676 Yu Zheng, Hong-Bin Zhang, Zhi-Chao Li, Zi-Feng Zhang, Yu-Zong Chen, Duo Li, Xiao-Qin Li,
677 and Feng Zhu. A unified framework for enzyme kinetic parameter prediction. *Nature Communi-*
678 *cations*, 14(1):8338, 2023.
- 679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701

APPENDIX

A HYPERPARAMETER AND TRAINING DETAILS

All models were trained on NVIDIA A100 GPUs. The core architecture for the Fine-Grained Interaction Pathway used a hidden dimension of 768 and 1 cross-attention layer. The specific hyperparameters used for each predictive task are detailed in Table 5.

Table 5: Task-specific hyperparameters for model training. All tasks used the AdamW optimizer and an early stopping patience of 25 epochs on the validation RMSE.

Predictive Task	Learning Rate	Batch Size
Kinetic (k_{cat})	1e-5	512
Kinetic (K_m)	1e-5	512
Kinetic (k_{cat}/K_m)	1e-5	512
Optimal Temperature (T_{opt})	1e-3	64
Optimal pH	5e-4	128

B DETAILS OF BASELINE METHODS

In this section, we briefly describe the underlying principles of the key baseline and state-of-the-art methods used for comparison in our main experiments (Table 1).

B.1 ESTABLISHED ENZYME-SPECIFIC FRAMEWORKS

These are published models specifically designed for enzyme kinetic prediction.

- **DLKcat** is a pioneering deep learning framework developed for k_{cat} prediction. It utilizes a Convolutional Neural Network (CNN) to automatically extract features from a numerical representation of the protein’s primary sequence.
- **UniKP** is a multi-task framework designed to jointly predict both k_{cat} and K_m . Its architecture combines representations from a pre-trained protein language model (ESM-1v) for the enzyme with features from a Graph Neural Network (GNN) for the substrate molecule.
- **CataPro** represents a strong, modern baseline that also relies on the direct concatenation of features. It fuses global representations from a state-of-the-art protein language model (ESM-2) with molecular fingerprints derived from the substrate’s SMILES string. The CataPro+SaProt variant further enhances the protein representation by incorporating features from SaProt, a structure-aware protein model.

B.2 GENERAL-PURPOSE ENCODER BASELINES

To establish a comprehensive and fair comparison, we constructed a suite of strong baseline models by combining state-of-the-art protein and molecule encoders. This allows us to assess the performance of a standard "concatenation-only" approach using the best available off-the-shelf components.

Protein Encoders We used two powerful, pre-trained protein language models:

- **ESM-2** is a transformer-based protein language model trained exclusively on tens of millions of protein sequences. It excels at generating context-aware embeddings that capture evolutionary and biophysical information.
- **ProtT5** is a sequence-to-sequence language model trained on a similar scale to ESM-2. Its architecture is particularly effective at capturing functional features from protein sequences. We also test a +SaProt variant, which enriches the sequence representation with structural information.
- **ProstT5** (Protein structure-sequence T5) is a protein language model (pLM) which can translate between protein sequence and structure. It is based on ProtT5-XL-U50, a T5

756 model trained on encoding protein sequences using span corruption applied on billions of
 757 protein sequences. ProstT5 finetunes ProtT5-XL-U50 on translating between protein se-
 758 quence and structure using 17M proteins with high-quality 3D structure predictions from
 759 the AlphaFoldDB. Protein structure is converted from 3D to 1D using the 3Di-tokens in-
 760 troduced by Foldseek.

761 **Substrate Encoders** The global protein embeddings were concatenated with various substrate rep-
 762 resentations to form the final input for the prediction head:

- 764 • **RDKitFP, Morgan, MACC:** These are widely-used, standard molecular fingerprints.
 765 They are fixed-length bit vectors calculated from the substrate’s 2D structure (SMILES
 766 string) and represent the presence or absence of specific chemical substructures.
- 767 • **MolT5 / Mole-BERT:** These are powerful, pre-trained transformer models for small
 768 molecules. Unlike static fingerprints, they generate context-aware embeddings for the sub-
 769 strate, capturing deeper chemical and semantic information.

770 The performance of these various combinations (e.g., ESM2_MACC, ProtT5_MolT5) represents a
 771 robust but standard approach where global features from both modalities are simply concatenated
 772 before prediction.

774 B.3 FINE-GRAINED INTERACTION PATHWAY

776 Importantly, the cross-attention module does not depend on any predefined structural or biochemical
 777 alignment between enzyme residues and substrate atoms. Instead, it learns these correspondences
 778 dynamically. Each substrate token attends over all enzyme tokens, and each enzyme token sym-
 779 metrically attends over all substrate tokens, forming a fully connected interaction graph. Through
 780 end-to-end optimization, the model identifies the residues most relevant to each substrate fragment,
 781 enabling it to capture catalytic signals, binding preferences, and other fine-grained biophysical rela-
 782 tionships directly from data.

783 To supply positional context for enzyme and substrate sequences, the pathway leverages the absolute
 784 positional embeddings already learned by the pre-trained ProtT5 and MolT5 encoders. These em-
 785 beddings provide robust information about sequential order and local neighborhoods, supporting the
 786 discovery of fine-grained interactions across long protein sequences. Although relative positional
 787 embeddings are not used in the current implementation, they represent a natural extension for future
 788 refinement and may offer benefits for modeling long-range residue dependencies.

789 In combination, padding with masked attention, dynamic cross-attention without predefined align-
 790 ments, and pretrained absolute positional embeddings enable the Fine-Grained Interaction Pathway
 791 to model enzyme–substrate interactions at high resolution while remaining architecture-agnostic and
 792 fully end-to-end.

794 C TRAINING PARADIGMS AND CROSS-TASK SYNERGY

796 C.1 TRAINING PARADIGMS

798 To further clarify the role of the unified architecture and address the question of how different tasks
 799 interact within ENZYME-UNIFIED, we conduct two additional analyses: (1) a direct comparison
 800 between specialized and true multi-task training using a shared hybrid trunk, and (2) a transfer-
 801 learning–based evaluation of cross-task synergy. These analyses complement the main results and
 802 provide a deeper understanding of how the unified backbone supports multi-property modeling.

803 Although ENZYME-UNIFIED is designed as a unified architectural framework, the optimal train-
 804 ing paradigm may depend on dataset size, noise, and task heterogeneity. To evaluate this, we com-
 805 pare three strategies under identical training conditions: **Specialized models:** where ENZYME-
 806 UNIFIED is trained separately for each task; **Multi-task ENZYME-UNIFIED:** where a shared
 807 hybrid trunk feeds five task-specific heads trained jointly; **Multi-task CataPro:** where the previous
 808 SOTA is adapted to the same multi-task setting. The results are shown in the Table 6.

809 Across all five prediction tasks, the unified architecture consistently outperforms CataPro under
 the same multi-task regime, confirming the advantage of our hybrid design independent of training

strategy. Meanwhile, specialized training yields the best overall performance on current homology-filtered datasets (e.g., T_{opt} RMSE 10.33 vs 14.51 vs 18.77 for the three strategies above). This indicates that while the architecture is unified, specialized optimization remains empirically favorable on existing benchmarks, likely due to dataset scale and inter-task imbalance.

Table 6: Head-to-head comparison of training paradigms. ENZYME-UNIFIED (Ours) and CataPro are evaluated under both Specialized (one model per task) and Multi-Task (shared trunk, task-specific heads) settings. Best performance within each paradigm is in **bold**.

Task	Metric	Specialized		Multi-Task	
		Ours	CataPro	Ours	CataPro
k_{cat}	RMSE ↓	1.304	1.329	1.306	1.339
	PCC ↑	0.498	0.497	0.493	0.483
	SCC ↑	0.484	0.495	0.495	0.481
K_m	RMSE ↓	0.996	0.998	0.998	1.009
	PCC ↑	0.643	0.633	0.635	0.629
	SCC ↑	0.631	0.629	0.627	0.626
k_{cat}/K_m	RMSE ↓	1.597	1.716	1.664	1.764
	PCC ↑	0.409	0.368	0.405	0.358
	SCC ↑	0.416	0.366	0.395	0.360
Optimal pH	RMSE ↓	0.840	2.167	0.923	1.223
	PCC ↑	0.657	0.521	0.584	0.513
	SCC ↑	0.512	0.427	0.473	0.433
Optimal T_{opt}	RMSE ↓	10.329	17.553	14.508	18.768
	PCC ↑	0.783	0.661	0.717	0.617
	SCC ↑	0.640	0.495	0.616	0.436

C.2 CROSS-TASK SYNERGY

To examine whether the unified architecture captures transferable biochemical features across properties, we conduct an explicit cross-task transfer experiment. We pretrain ENZYME-UNIFIED on optimal pH prediction, freeze the encoder, and fine-tune a new regression head on T_{opt} . This “pH $\rightarrow T_{opt}$ ” transfer setting isolates representational sharing from optimization-level interactions.

The transfer model achieves substantial improvements over training on T_{opt} alone, reducing RMSE from 14.17 to 11.11 (-21.6%) and increasing PCC from 0.617 to 0.706. This demonstrates that the model learns biophysically meaningful and task-general representations e.g., sequence determinants of stability, that benefit multiple properties. The results are shown in the Table 7.

These findings empirically support our holistic motivation: although tasks are optimized independently in the main results for performance reasons, ENZYME-UNIFIED does learn cross-property biochemical signals that are transferable across tasks.

D DETAILED ABLATION STUDY ANALYSIS

This section provides a more detailed discussion of the ablation studies summarized in Section 4.4 and Table 3 of the main text.

D.1 THE SYNERGY OF THE HYBRID ARCHITECTURE

Our central hypothesis is that a hybrid model combining global and fine-grained features outperforms its constituent pathways. To test this, we trained two ablated models: (1) a **Concatenation-only** model and (2) an **Attention-only** model. As shown in Table 3, the full hybrid model substantially outperforms both across all metrics. Between the individual pathways, the “Concatenation-only” model establishes a stronger baseline. Crucially, however, the full hybrid model’s performance surpasses even this strong baseline. This indicates that while the “Attention-only” pathway is less ef-

Table 7: Direct evidence for cross-task synergy on the Optimal T_{opt} prediction task. This table compares the performance of our hybrid model trained via transfer learning against all standard baselines and our fully trained models. The ‘Transfer’ model, pretrained on Optimal pH data, shows a dramatic performance improvement over the ‘Baseline’ trained from scratch, demonstrating that our hybrid model learns transferable biophysical representations.

Model Setting	RMSE on $T_{\text{opt}} \downarrow$	PCC on $T_{\text{opt}} \uparrow$	SCC on $T_{\text{opt}} \uparrow$
<i>General Baselines</i>			
CataPro	17.553	0.661	0.495
UniKP	14.347	0.670	0.473
ProtT5-3B	16.342	0.686	0.502
ESM2-15B	12.170	0.697	0.582
<i>Our Models (Specialized Training)</i>			
ENZYME-UNIFIED (hybrid)	14.446	0.666	0.622
ENZYME-UNIFIED (hybrid+ProtT5)	12.433	0.703	0.612
ENZYME-UNIFIED (hybrid++)	10.329	0.783	0.640
<i>Synergy Experiment: Transfer vs. From Scratch</i>			
Baseline (Train on T_{opt} from scratch)	14.17	0.617	0.607
Transfer (Pre-train on pH \rightarrow Fine-tune on T_{opt})	11.11	0.706	0.640

fective in isolation, it captures distinct, complementary features that are vital for refining the model’s predictions. The synergy of combining both pathways is therefore essential.

D.2 THE IMPORTANCE OF THE GATED FUSION MECHANISM

We investigated the importance of our trainable gating parameter, α , by comparing our standard model against a variant using a simple, fixed average (i.e., $\sigma(\alpha)$ is fixed to 0.5). As shown in Table 3, the learnable gate provides a consistent performance improvement. This suggests that allowing the model to learn the optimal, task-specific balance between the global and fine-grained pathways is a beneficial architectural refinement that contributes to the final performance.

D.3 QUANTIFYING THE IMPACT OF ENHANCED FEATURES

Here we provide a more detailed analysis of the performance gains from incorporating advanced features, referencing results from our main experiments (Table 2).

Kinetic Prediction Incorporating structural embeddings from ProtT5 yielded only a marginal improvement. We hypothesize that this is due to a high degree of informational redundancy for this task. With over 90% of the enzymes in our training set having structures predicted by AlphaFold2, the 3D tokens from ProtT5 may exhibit strong collinearity with the primary sequence features they are derived from, limiting the amount of novel, independent information the model can extract. It is possible that the global structural representations from ProtT5, while informative, lack the fine-grained resolution to capture the subtle conformational dynamics crucial for kinetics.

Environmental Prediction In stark contrast, the impact was significantly more pronounced for environmental prediction. Supplementing our model with amino acid physicochemical features for T_{opt} prediction resulted in a dramatic performance leap, **reducing the RMSE by nearly 41%** (from 17.55 to 10.33) and **increasing the PCC by over 0.12**. We attribute this substantial improvement to the fact that these features, such as the aliphatic index, provide crucial biophysical information directly related to thermal stability—information that is not explicitly or easily derivable from sequence or global structure alone. This suggests they offer a new, orthogonal source of information that is highly relevant for modeling an enzyme’s adaptation to temperature.

E DATA PROCESSING

To guarantee an unbiased and generalizable evaluation, we constructed our dataset using a rigorous clustering-based splitting protocol: all enzyme sequences were first clustered with MMseqs2 at a 40% identity threshold, and then entire clusters—rather than individual samples—were distributed into non-overlapping training, validation and test folds. This ensures that the maximum sequence

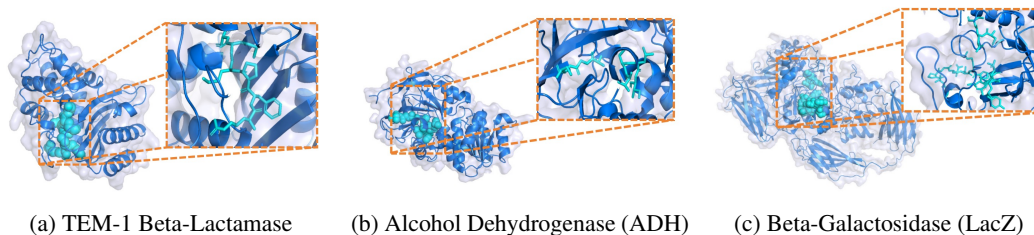


Figure 5: Cross-attention visualizations for three representative enzymes across different EC classes. ENZYME-UNIFIED consistently highlights known catalytic residues without structural supervision.

similarity between any enzyme in different folds never exceeds 40%, forcing the model to generalize across protein families. Substrate or assay meta-information were retained in their natural distribution across folds, were not used as splitting criteria, and received no additional annotation, eliminating any implicit leakage. The procedure follows best practices in protein learning, and Table 8 is a detailed description of each dataset. Consequently, our dataset has been strictly processed and is free of data-leakage issues.

Table 8: The k_{cat}/K_m data contain 10 folds; the first fold comprises 2,637 samples, and each of the remaining nine folds contains 2,636 samples. The pH dataset consists of 5 folds, each with 104 samples. The temp dataset has 5 folds; the first fold includes 142 samples, and the other four folds each contain 141 samples.

Dataset	Total Samples	Seq. Id. Threshold	Clusters	Folds	Samples per Fold Distribution
k_{cat}/K_m	26,361	0.4	3329	10	All Folds 1: 2,637; Folds 2–10: 2,636
pH	520	0.4	205	5	All Folds: 104
Temperature	706	0.4	251	5	Fold 1: 142; Folds 2–5: 141

F CASE STUDY

This appendix provides additional interpretability experiments to evaluate the robustness and generality of the ENZYME-UNIFIED cross-attention mechanism across different enzyme classes, catalytic mechanisms, and sequence scales. We selected three representative enzymes: **TEM-1 Beta-Lactamase**, **Alcohol Dehydrogenase (ADH)**, and **Beta-galactosidase (LacZ)**. All analyses use raw cross-attention weights without any structural labels or task-specific fine-tuning.

TEM-1 Beta-Lactamase (EC 3.5.2.6): The model consistently highlights residues 65–70, including the catalytic nucleophile Ser70. This indicates that ENZYME-UNIFIED can identify the beta-lactam hydrolysis motif directly from sequence. **Alcohol Dehydrogenase (ADH, EC 1.1.1.1):** For this zinc-dependent oxidoreductase, the model focuses on residues 70–72 and 169–172, aligning with the catalytic zinc-binding residues His67 and Cys174. **Beta-Galactosidase (LacZ, EC 3.2.1.23):** For this large (>1000 aa) enzyme, ENZYME-UNIFIED highlights regions 460–462 and 525–538, capturing the key catalytic residues Glu461 and Glu537.

Across these diverse enzymes, ENZYME-UNIFIED reliably identifies catalytic centers, demonstrating generalizable interpretability beyond RNase A.

ETHICS STATEMENT

The work presented in this paper adheres to the ICLR Code of Ethics. All datasets constructed and used in this study were aggregated exclusively from publicly available scientific databases, namely BRENDA and SABIO-RK, which contain data from peer-reviewed scientific literature. Our research does not involve any human subjects, personally identifiable information, or sensitive data. The intended application of this work is to accelerate scientific research and engineering in biocatalysis for positive societal impact, such as in green chemistry and therapeutics. To promote transparency and reproducibility, we will make our curated datasets and source code publicly available upon acceptance of this manuscript.