Supplementary Material for Multimodal Protein Language Models for Enzyme Kinetic Parameters: From Substrate Recognition to Conformational Adaptation

Fei Wang[1,2], Xinye Zheng[1], Kun Li[3], Yanyan Wei[1,4,*], Yuxin Liu[2], Ganpeng Hu[5], Tong Bao[5], Jingwen Yang[5]

[1] School of Computer Science and Information Engineering, Hefei University of Technology

[2] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center

[3] CVLab, College of Information Technology, United Arab Emirates University

[4] Intelligent Interconnected Systems Laboratory of Anhui Province

[5] School of Food Biological Engineering, Hefei University of Technology

Overview

The supplementary materials provide additional details and elaborations on the datasets used, quantitative analysis experiments, and ablation studies presented in the main manuscript. These topics are organized as follows:

Contents

*Corresponding authors

A. Enzyme Kinetic Parameter Dataset

A.1. Details Dataset from [3]

In the main experiments of our manuscript, we employ three specially curated kinetic endpoint datasets, as shown in Table S1, to construct and evaluate learning-based models for predicting enzyme kinetic parameters. The raw data underlying the three endpoint datasets were curated from two authoritative biochemical databases: BRENDA [12] release 2022_2 and SABIO-RK [18] as of November 2023. This construction aims to provide an expanded and standardized resource for the field of enzyme kinetic parameter prediction, thereby promoting the systematic development and fair benchmarking of enzyme kinetic prediction models.

The datasets focus on three primary enzyme kinetic parameters: the turnover number (k_{cat}), Michaelis constant (K_m), and inhibition constant (K_i), based on in vitro measurements of wild-type enzymes. Specifically, the datasets include:

- 23,151 entries for Kinetic Parameter k_{cat};
- 41,174 entries for Kinetic Parameter K_m;
- 11,929 entries for Kinetic Parameter K_i.

The construction of these datasets introduces several improvements:

❶ **Data Selection and Standardization:** When multiple experimental measurements exist for the same enzyme-substrate pair, the highest value is retained for k_{cat} to capture optimal reaction conditions. For K_m and K_i, the geometric mean is used, reflecting better consistency and stability for

Class	Kinetic Dataset k_{cat}	K_m	K_i	Total	Details
Dataset I (Main Test): k_{cat}, K_m and K_i from [3]					
Entries	23,151	41,174	11,929	76,254	
EC 1 Oxidoreductases	7,756	13,931	2,744	24,431	Oxidoreductases catalyze diverse redox reactions by transferring electrons between molecules and substrates, and represent the largest EC class with broad kinetic and functional diversity.
EC 2 Transferases	4,155	10,182	2,706	17,043	Transferases catalyze the transfer of functional groups between diverse substrates, enabling the formation of new chemical bonds and the rapid and efficient interconversion of molecules.
EC 3 Hydrolases	8,065	11,025	4,083	23,173	Hydrolases catalyze the hydrolysis of diverse chemical bonds in biological molecules. They efficiently cleave covalent linkages in various substrates with variable enzyme kinetic parameters.
EC 4 Lyases	1,620	2,962	1,525	6107	Lyases catalyze the cleavage of chemical bonds without hydrolysis or oxidation. They form double bonds or ring structures in substrates.
EC 5 Isomerases	928	1,311	434	2,673	Isomerases catalyze the conversion of molecules into their isomeric forms by reorganizing atomic connectivity within the same substrate.
EC 6 Ligases	590	1,363	254	2,207	Ligases catalyze the formation of covalent bonds between two different molecules using chemical energy from nucleotide triphosphates.
EC 7 Translocases	16	98	23	137	Translocases catalyze the active transport of ions or molecules across biological membranes, efficiently coupled with chemical reaction energy.
Unclassified/Invalid	21	302	160	483	Entries that are unclassified or contain format errors, excluded from standard EC categorization.
Unique Sequences	7,183	12,355	2,829	22,367	
Unique EC Classes	2,657	3,550	1,306	7,513	
Unique Organisms	1,685	2,419	652	4,756	
Dataset II (OOD Test): k_{cat} and K_m from [14]					
Entries	35,147	29,321	-	64,468	-
EC 1 Oxidoreductases	12,197	9,770	-	21,967	Oxidoreductases catalyze diverse redox reactions by transferring electrons between molecules and substrates, and represent the largest EC class with broad kinetic and functional diversity.
EC 2 Transferases	8,148	7,820	-	15,968	Transferases catalyze the transfer of functional groups between diverse substrates, enabling the formation of new chemical bonds and the rapid and efficient interconversion of molecules.
EC 3 Hydrolases	8,372	6,537	-	14,909	Hydrolases catalyze the hydrolysis of diverse chemical bonds in biological molecules. They efficiently cleave covalent linkages in various substrates with variable enzyme kinetic parameters.
EC 4 Lyases	3,264	2,510	-	5,774	Lyases catalyze the cleavage of chemical bonds without hydrolysis or oxidation. They form double bonds or ring structures in substrates.
EC 5 Isomerases	1,854	1,367	-	3,221	Isomerases catalyze the conversion of molecules into their isomeric forms by reorganizing atomic connectivity within the same substrate.
EC 6 Ligases	1,252	1,187	-	2,439	Ligases catalyze the formation of covalent bonds between two different molecules using chemical energy from nucleotide triphosphates.
EC 7 Translocases	60	130	-	190	Translocases catalyze the active transport of ions or molecules across biological membranes, efficiently coupled with chemical reaction energy.

Table S1. Detailed breakdown of the enzyme kinetic datasets from [3, 14] used for training and out-of-distribution (OOD) testing. We provide the number of entries, unique sequences, EC classes, organisms, and additional information for each enzyme class in the main and OOD test datasets.

binding affinity measurements.

❷ **Diversity of Enzyme Sequences:** These datasets significantly expand enzyme sequence diversity compared to previous datasets. They cover a broader range of enzyme families, particularly at the first Enzyme Commission (EC) level, without overrepresenting any specific categories. This is an important step towards fair and comprehensive benchmarking.

❸ **Link to Structural Data:** Each entry is linked to a predicted three-dimensional structure of the corresponding enzyme. In cases where structures are not available in AlphaFold-2.0 [6], ESMFold [11] is used to generate these models. This ensures that structural context is available for all entries, further enriching the dataset for model training.

❹ **Supporting Fair Benchmarking:** By standardizing and expanding the sequence and structure coverage, the dataset offers a more representative resource for enzyme kinetic pa-

rameter prediction. This makes it well-suited for developing and fairly evaluating learning-based models in the field.

Overall, this expanded dataset not only enhances the diversity of enzyme sequences but also ensures consistency and stability in model training by using a representative value for each enzyme-substrate pair. Compared to prior works like DLKcat [10] and TurNup [8], our dataset covers a broader range of enzyme families and provides a more comprehensive foundation for large-scale enzyme kinetic predictions.

A.2. Details OOD Dataset from [14]

To evaluate the out-of-distribution (OOD) generalization performance of our model, we utilize two enzyme kinetic datasets focused on k_{cat} and K_{m}, compiled from the BRENDA and SABIO-RK databases as collected by [14] in Table S1. Each data entry contains enzyme sequences, substrate information, and experimentally measured kinetic parameters.

❶ **Data Curation and Filtering:** The initial collections contained 143,473 entries for k_{cat} and 75,149 entries for K_{m}. To ensure biological relevance, every protein sequence was mapped to a unique UniProt ID, and sequences lacking a valid or unambiguous ID were excluded. Substrate names were cross-referenced with KEGG [7] and PubChem [16] to obtain standardized 2D molecular structures. These substrates were then converted into canonical SMILES using RDKit [9]. Any substrates with ambiguous or unresolvable identities were discarded to maintain chemical consistency.

❷ **Inclusion of Mutant Enzymes:** Mutant enzyme variants were retained in the dataset. While K_{m} showed minimal differences between wild-type and mutant enzymes, k_{cat} distributions revealed that mutant enzymes typically exhibit significantly reduced catalytic activity and efficiency compared to wild-type counterparts.

❸ **Final dataset composition:** After rigorous screening, the final dataset comprised 35,147 k_{cat} records and 29,321 K_{m} records, with one-tenth allocated as a test set for evaluation.

❹ **3D Structure Generation for OOD Testing:** Notably, we use the k_{cat} and K_{m} test sets from [14] for out-of-distribution generalization testing. The corresponding 3D structures for the enzymes are generated using OpenFold [1] or ESMFold [11], ensuring that the structural data is consistent with the latest advancements in protein structure prediction.

B. Enzyme Commission Distribution

This section explores the performance of various models across different Enzyme Commission (EC) classes, focusing on the influence of model architecture and the impact of our **ERBA (Enzyme-Reaction Bridging Adapter)** in improving predictions for enzyme kinetic parameters.

B.1. Overview of EC Class Distributions

Enzyme Commission (EC) classes group enzymes based on their catalytic functions, and understanding the distribution of these enzymes across various EC classes is crucial for predicting their kinetic parameters accurately. The datasets used in this experiment cover a wide range of EC classes, from EC 1 (Oxidoreductases) to EC 6 (Ligases), and include kinetic data for k_{cat}, K_{m}, and K_{i}. Each class includes different numbers of entries, with EC 3 (Hydrolases) containing the most data, and EC 6 (Ligases) containing fewer. The purpose of this analysis is to evaluate how different model backbones and ERBA impact the prediction performance across these classes. It is notable that EC 7 (Translocases) is excluded from the analysis due to insufficient data.

B.2. Effect of Different Model Architectures Across EC Classes

Table S2 presents a comparison of various models across six EC classes, with a focus on the performance metrics R^2, PCC, RMSE, and MAE for k_{cat}, K_{m}, and K_{i}. We observe that larger backbone models, such as Ankh3-5.7B [2] and ProtT5-3B [5], outperform smaller models, like ESM2-650M, across all EC classes, suggesting that larger models are better equipped to handle the complexities of enzyme-substrate interactions. In EC 1 (Oxidoreductases), Ankh3-5.7B achieves a higher R^2 score for k_{cat} (0.52) compared to ESM2-650M (R^2 = 0.45). This indicates that larger backbone models are better at capturing the intricate relationships between enzyme sequences and substrate properties.

B.3. Impact of Model Size within the ESM Architecture

Focusing specifically on the ESM2 architecture, we examine how different model sizes affect performance. As shown in (c) and (d), increasing the model size from ESM2-650M to ESM2-3B leads to noticeable improvements in performance across EC classes. For instance, in EC 5 (Isomerases), the R^2 score for k_{cat} improves from 0.48 with ESM2-650M to 0.50 with ESM2-3B. This performance increase reflects the enhanced capacity of larger models to process more complex enzyme-substrate interactions, which are particularly important in enzyme families with diverse substrates like EC 5 and EC 6.

B.4. Conclusion

The results demonstrate that larger backbone models, such as Ankh3-5.7B and ProtT5-3B, perform better across diverse EC classes, highlighting the importance of model size in capturing complex enzymatic processes. In the case of ESM2, increasing the model size from ESM2-650M to

Index	Class Name	Kinetic Endpoint: k_{cat}				Kinetic Endpoint: K_m				Kinetic Endpoint: K_i			
		R^2↑	PCC↑	RMSE↓	MAE↓	R^2↑	PCC↑	RMSE↓	MAE↓	R^2↑	PCC↑	RMSE↓	MAE↓
(a) ***Ankh3-5.7B [2] w/ ERBA***													
EC-1	Oxidoreductases	0.39	0.68	1.25	0.89	0.51	0.78	0.83	0.61	0.53	0.73	1.29	0.96
EC-2	Transferases	0.44	0.65	1.23	0.87	0.55	0.74	0.78	0.58	0.39	0.65	1.27	0.97
EC-3	Hydrolases	0.45	0.70	1.13	0.85	0.53	0.73	0.85	0.60	0.29	0.60	1.72	1.27
EC-4	Lyases	0.52	0.73	1.20	0.91	0.60	0.79	0.73	0.55	0.54	0.74	1.36	0.99
EC-5	Isomerases	0.42	0.62	1.41	0.95	0.60	0.76	0.85	0.67	0.65	0.82	1.15	0.91
EC-6	Ligases	0.50	0.71	1.01	0.79	0.30	0.55	1.12	0.83	0.27	0.51	1.38	1.08
(b) ***ProtT5-3B [4] w/ ERBA***													
EC-1	Oxidoreductases	0.40	0.71	1.44	0.99	0.60	0.79	0.77	0.58	0.55	0.73	1.25	0.98
EC-2	Transferases	0.48	0.73	1.25	0.90	0.51	0.69	0.79	0.58	0.42	0.69	1.39	1.03
EC-3	Hydrolases	0.42	0.69	1.20	0.89	0.50	0.71	0.94	0.69	0.46	0.68	1.41	1.04
EC-4	Lyases	0.48	0.70	1.21	0.85	0.53	0.71	0.87	0.65	0.69	0.83	1.07	0.77
EC-5	Isomerases	0.46	0.70	1.29	0.90	0.63	0.80	0.80	0.61	0.65	0.81	1.10	0.89
EC-6	Ligases	0.54	0.79	0.92	0.67	0.40	0.65	0.97	0.71	0.41	0.65	1.30	0.97
(c) ***ESM2-650M [11] w/ ERBA***													
EC-1	Oxidoreductases	0.35	0.62	1.37	1.05	0.43	0.67	1.01	0.77	0.51	0.72	1.28	0.98
EC-2	Transferases	0.39	0.66	1.34	1.02	0.34	0.62	0.96	0.74	0.35	0.60	1.39	1.05
EC-3	Hydrolases	0.40	0.65	1.32	1.01	0.33	0.60	1.05	0.80	0.31	0.65	1.71	1.29
EC-4	Lyases	0.44	0.69	1.28	0.98	0.31	0.58	1.00	0.77	0.46	0.74	1.47	1.10
EC-5	Isomerases	0.43	0.65	1.32	0.99	0.53	0.77	0.90	0.81	0.71	0.87	1.12	0.88
EC-6	Ligases	0.50	0.71	0.99	0.79	0.33	0.57	1.09	0.85	0.32	0.60	1.41	1.08
(d) ***Ours (ESM2-3B [11] w/ ERBA)***													
EC-1	*Oxidoreductases*	0.37	0.62	1.32	0.95	0.55	0.74	0.90	0.69	0.58	0.77	1.19	0.92
EC-2	*Transferases*	0.46	0.69	1.22	0.87	0.53	0.73	0.81	0.61	0.38	0.65	1.37	0.99
EC-3	*Hydrolases*	0.43	0.68	1.18	0.88	0.51	0.72	0.90	0.65	0.48	0.71	1.50	1.10
EC-4	*Lyases*	0.50	0.71	1.27	0.92	0.56	0.75	0.81	0.61	0.65	0.81	1.19	0.84
EC-5	*Isomerases*	0.46	0.73	1.39	0.94	0.61	0.78	0.84	0.64	0.67	0.83	1.20	0.95
EC-6	*Ligases*	0.53	0.77	0.96	0.71	0.34	0.59	1.02	0.76	0.34	0.62	1.39	1.03

Table S2. Performance Comparison Across Enzyme Commission Classes. The table compares the performance of different backbone models and the impact of ERBA on kinetic parameter prediction for six EC classes: EC 1 (Oxidoreductases), EC 2 (Transferases), EC 3 (Hydrolases), EC 4 (Lyases), EC 5 (Isomerases), and EC 6 (Ligases). Metrics include R^2, PCC, RMSE, and MAE for k_{cat}, K_m, and K_i, showcasing the effect of model size and architecture on the predictive accuracy across enzyme classes.

ESM2-3B significantly enhances performance, especially in EC classes with diverse substrates, such as EC 5 (Isomerases). Furthermore, the introduction of ERBA provides a substantial boost to model performance by incorporating both substrate recognition and structural adaptation. This underscores the effectiveness of staged conditioning in improving enzyme kinetic predictions and enhancing the model's generalization capabilities.

C. Error Distribution Analysis

C.1. Overview of Error Distributions

The error distributions across different models (Ankh3-5.7B, ProtT5-3B, ESM2-650M, and ESM2-3B) and their corresponding performance with ERBA are analyzed in terms of absolute prediction error. As shown in Figure S1, all models show an overall improvement when enhanced with ERBA, particularly in terms of 1-Radio$_{AE}$ values, which represent the percentage of predictions with an absolute error smaller than 1.

C.2. Performance of Ankh3-5.7B and Its Unique Behavior on K_m

Notably, the Ankh3-5.7B model, while showing lower overall performance compared to ESM2-3B (as observed in the manuscript), performs significantly better in terms of K_m predictions. Specifically, the Ankh3-5.7B model achieves a high 1-Radio$_{AE}$ value of 78.82% for K_m predictions, indicating a higher proportion of predictions with small errors despite having lower overall R^2 and PCC scores. This suggests that Ankh3-5.7B may be better at making precise predictions within a narrower range of K_m values, while other models may struggle with broader generalization across the K_m space.

C.3. ESM2-3B: Best Overall Performance with ERBA

The ESM2-3B model consistently outperforms all other backbones, particularly for k_{cat} and K_m. The error distributions show that the ESM2-3B with ERBA achieves the highest 1-RadioAE values (67.67% for k_{cat}, 79.53% for K_m,

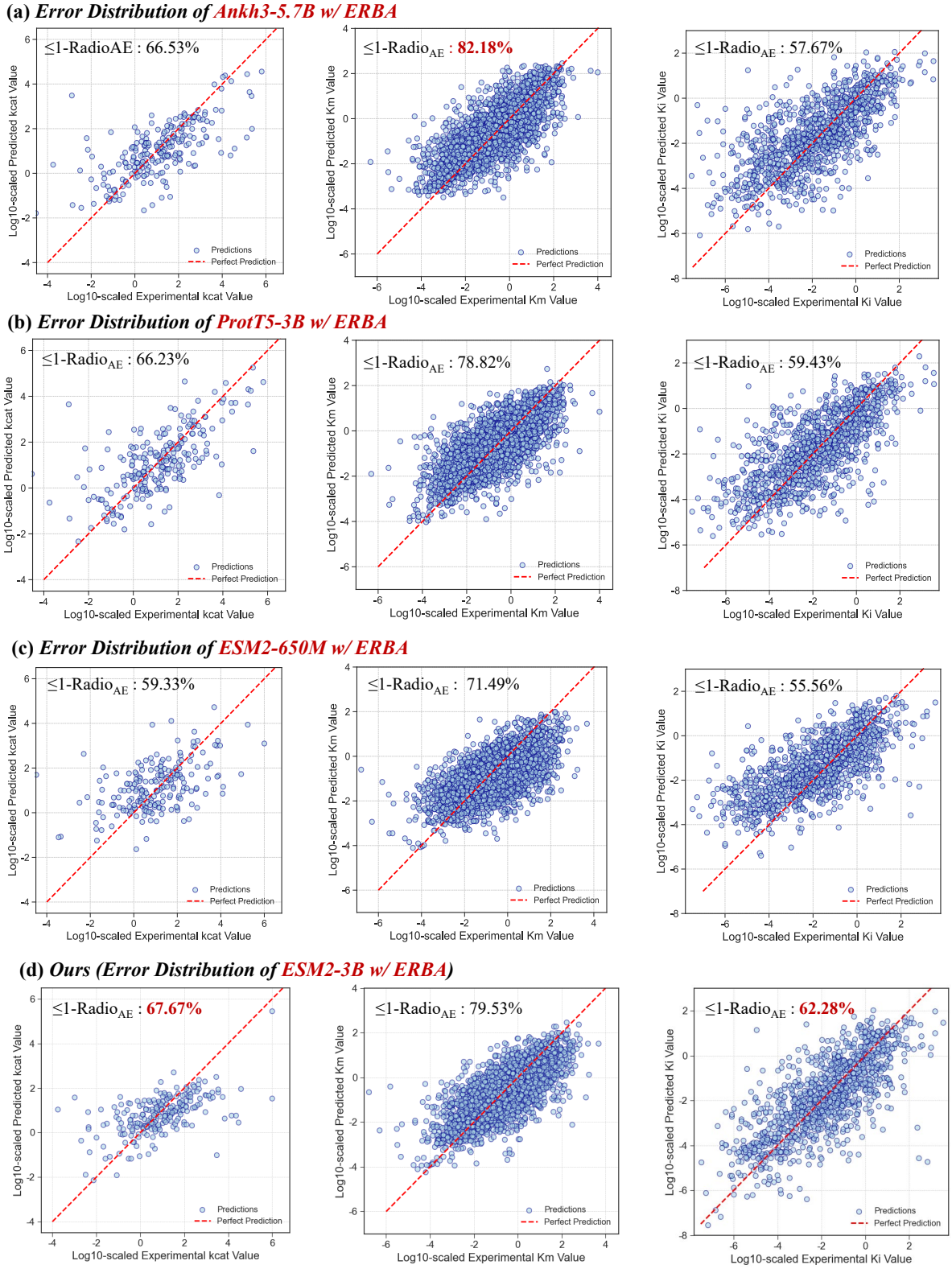

Figure S1. Error Distribution Comparison Across Different Backbone Models and ESM Sizes. It shows the error distribution of predicted versus experimental values for three kinetic parameters (k_{cat}, K_m, and K_i) across four backbone models: (a) Ankh3-5.7B [2], (b) ProtT5-3B [4], (c) ESM2-650M [11], and (d) ESM2-3B [11], each augmented with ERBA. The plots show the proportion of predictions with absolute error less than or equal to 1, denoted as 1-Radio$_{AE}$.

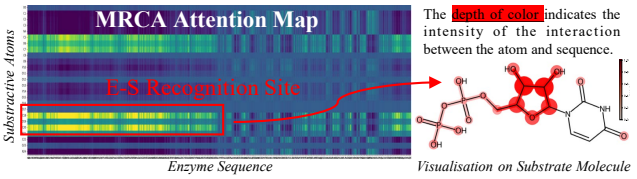

Figure S2. Visualization of MRCA.

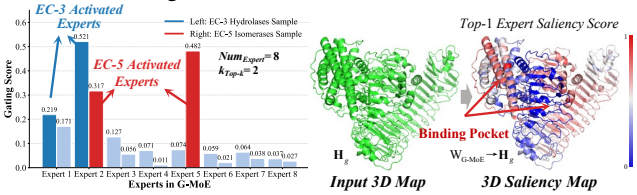

Figure S3. Visualization of G-MoE.

	Method	Kinetic Endpoint: K_i			
		$R^2\uparrow$	PCC↑	RMSE↓	MAE↓
Stage 1	*w/* Concat & Self-Atention [15]	0.47	0.69	1.44	1.10
	***w/* MRCA (Ours)**	**0.61**	**0.78**	**1.26**	**0.91**

Table S3. Stage 1 (Recognition) ablation on K_i: Replacing *Concat & Self-Attention* with our **MRCA** markedly improves R^2 and PCC while reducing RMSE/MAE under identical settings.

and 62.28% for K_i), demonstrating its robust performance across all three kinetic endpoints. This confirms the effectiveness of the ESM2-3B backbone combined with ERBA for enzyme kinetic parameter prediction.

C.4. Comparison with Smaller ESM2 Variants

Interestingly, while the larger models like ESM2-3B excel overall, the smaller versions (such as ESM2-650M) show a reduction in the percentage of predictions with smaller errors, especially for K_m. This highlights the trade-off between model size and accuracy for out-of-distribution generalization. Larger models tend to perform better overall but may exhibit more variability in specific kinetic parameter predictions, particularly in smaller enzyme classes or in scenarios with less data.

D. Explainable Analysis

We further analyze the interpretability of our method through the visualization of module behaviors. We further analyze the interpretability of our method through the visualization of module behaviors. In Figure S2, MRCA produces highly localized attention regions on the sequence and assigns the highest scores to chemically functional groups on the substrate. In Figure S3, the gating distribution is sparse and dynamically relies on enzyme characteristics, while 3D saliency map shows that top-1 expert responses cluster around the adaptive binding pocket.

E. Additional Ablation Studies

To isolate the contribution of each component, we ablate on K_i prediction under identical preprocessing, backbone, and training schedule. Across all studies, each module provides measurable gains and the full model is consistently best.

	Method	Kinetic Endpoint: K_i			
		$R^2\uparrow$	PCC↑	RMSE↓	MAE↓
Stage 2	*w/* MoE [13]	0.50	0.71	1.39	1.05
	***w/* G-MoE (Ours)**	**0.61**	**0.78**	**1.26**	**0.91**

Table S4. Stage 2 (Adaptation) ablation on K_i: *standard MoE vs.* our **G-MoE** that routes by pocket geometry; geometry-aware experts yield consistent gains across all metrics.

	Method	Kinetic Endpoint: K_i			
		$R^2\uparrow$	PCC↑	RMSE↓	MAE↓
Optimization	*w/* $\mathcal{L}_2$ Loss [17]	0.48	0.69	1.42	1.08
	***w/* ESDA (Ours)**	**0.61**	**0.78**	**1.26**	**0.91**

Table S5. Objective ablation on K_i: *plain* $\mathcal{L}_2$ *vs.* **ESDA**-regularized objective. Distribution alignment in RKHS stabilizes fine-tuning and improves R^2/PCC while lowering RMSE/MAE.

E.1. Stage 1: Does MRCA Help Recognition?

We replace the common *Concat & Self-Attention* [15] fusion with our **MRCA**, as shown in Table S3, which explicitly conditions enzyme tokens on substrate tokens in the PLM latent space before any structural input. MRCA improves R^2 from 0.47 to **0.61** (+0.14 / ↑29.8%), PCC from 0.69 to **0.78** (+0.09 / ↑13.0%), and reduces RMSE from 1.44 to **1.26** (↓12.5%) and MAE from 1.10 to **0.91** (↓17.3%). Results indicate that conditioning the PLM on substrate identity early yields a cleaner recognition signal than shallow concatenation, which underuses PLM priors.

E.2. Stage 2: Is Geometry-aware MoE Necessary?

As shown in Table S4, we compare a *standard MoE* [13] against our **G-MoE**, which routes by pocket geometry and applies pocket-local, low-rank adaptation. G-MoE raises R^2 0.50→**0.61** (+0.11 / ↑22.0%) and PCC 0.71→**0.78** (+0.07 / ↑9.9%), while lowering RMSE 1.39→**1.26** (↓9.4%) and MAE 1.05→**0.91** (↓13.3%). This validates that the geometry-routed experts specialize to distinct conformational regimes (pocket size/shape/flexibility), outperforming a single shared adaptation rule.

E.3. Does ESDA Stabilize Fine-tuning?

We replace the plain $\mathcal{L}_2$ regression with our **ESDA** objective, reported in Table S5, which aligns the distributions of $\mathbf{H}^{(0)}$ (sequence), $\mathbf{H}^{(1)}$ (seq+substrate), and $\mathbf{H}^{(2)}$ (seq+substrate+structure) to the PLM manifold via RKHS MMD. ESDA improves R^2 0.48→**0.61** (+0.13 / ↑27.1%) and PCC 0.69→**0.78** (+0.09 / ↑13.0%), and reduces RMSE 1.42→**1.26** (↓11.3%) and MAE 1.08→**0.91** (↓15.7%). Thus, the distribution alignment curbs feature flooding from strong structural cues and preserves pretrained biochemical semantics during adaptation.

E.4. Overall

Gains accumulate across stages: MRCA supplies substrate-aware recognition, G-MoE injects pocket-specific adaptation, and ESDA keeps both anchored to the PLM's semantic space, yielding the best accuracy and lowest errors on K_{i}.

References

[1] Gustaf Ahdritz, Nazim Bouatta, Christina Floristean, Sachin Kadyan, Qinghui Xia, William Gerecke, Timothy J O'Donnell, Daniel Berenberg, Ian Fisk, Niccolò Zanichelli, et al. Openfold: Retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature methods*, 21(8):1514–1524, 2024. 3

[2] Hazem Alsamkary, Mohamed Elshaffei, Mohamed Elkerdawy, and Ahmed Elnaggar. Ankh3: Multi-task pretraining with sequence denoising and completion enhances protein representations. *arXiv preprint arXiv:2505.20052*, 2025. 3, 4, 5

[3] Veda Sheersh Boorla and Costas D Maranas. Catpred: a comprehensive framework for deep learning in vitro enzyme kinetic parameters. *Nature communications*, 16(1):2072, 2025. 1, 2

[4] Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, et al. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7112–7127, 2021. 4, 5

[5] Michael Heinzinger, Konstantin Weissenow, Joaquin Gomez Sanchez, Adrian Henkel, Milot Mirdita, Martin Steinegger, and Burkhard Rost. Bilingual language model for protein sequence and structure. *NAR Genomics and Bioinformatics*, 6(4):lqae150, 2024. 3

[6] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021. 2

[7] Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1): 27–30, 2000. 3

[8] Alexander Kroll, Yvan Rousset, Xiao-Pan Hu, Nina A Liebrand, and Martin J Lercher. Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nature communications*, 14(1):4139, 2023. 3

[9] Greg Landrum. Rdkit documentation. *Release*, 1(1-79):4, 2013. 3

[10] Feiran Li, Le Yuan, Hongzhong Lu, Gang Li, Yu Chen, Martin KM Engqvist, Eduard J Kerkhoven, and Jens Nielsen. Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis*, 5(8):662–672, 2022. 3

[11] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. 2, 3, 4, 5

[12] Ida Schomburg, Antje Chang, Christian Ebeling, Marion Gremse, Christian Heldt, Gregor Huhn, and Dietmar Schomburg. Brenda, the enzyme database: updates and major new developments. *Nucleic acids research*, 32(suppl_1):D431–D433, 2004. 1

[13] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*, 2017. 6

[14] Xiaowei Shen, Ziheng Cui, Jianyu Long, Shiding Zhang, Biqiang Chen, and Tianwei Tan. Eitlem-kinetics: A deep-learning framework for kinetic parameter prediction of mutant enzymes. *Chem Catalysis*, 4(9), 2024. 1, 2, 3

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 6

[16] Yanli Wang, Jewen Xiao, Tugba O Suzek, Jian Zhang, Jiyao Wang, and Stephen H Bryant. Pubchem: a public information system for analyzing bioactivities of small molecules. *Nucleic acids research*, 37(suppl_2):W623–W633, 2009. 3

[17] Bernard Widrow and Marcian E Hoff. Adaptive switching circuits. In *Neurocomputing: foundations of research*, pages 123–134. 1988. 6

[18] Ulrike Wittig, Renate Kania, Martin Golebiewski, Maja Rey, Lei Shi, Lenneke Jong, Enkhjargal Algaa, Andreas Weidemann, Heidrun Sauer-Danzwith, Saqib Mir, et al. Sabiork—database for biochemical reaction kinetics. *Nucleic acids research*, 40(D1):D790–D796, 2012. 1