On *de novo* Enzyme Discovery with ReactZyme and EnzymeFlow: Towards AI-driven Enzyme Design Platform

(Presented by McGill University, Mila-Quebec AI Institute, Shanghai Jiao Tong University, University of Montreal, Hong Kong University of Science and Technology, University of Washington, Institue for Protein Design, Microsoft Research, DeepMind)

ReactZyme: A Benchmark for Enzyme-Reaction Prediction

Chenqing Hua[1,3*] **Bozitao Zhong**[2*] **Sitao Luan**[1,3]

Liang Hong [2] **Guy Wolf** [3,4] **Doina Precup** [1,3,5] **Shuangjia Zheng**[2†]
[1]McGill; [2]SJTU; [3]Mila; [4]UdeM; [5]DeepMind

ReactZyme – In proceedings of 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks.

ReactZyme paper: https://arxiv.org/pdf/2408.13659

ReactZyme code: https://github.com/WillHua127/ReactZyme

ENZYMEFLOW: GENERATING REACTION-SPECIFIC ENZYME CATALYTIC POCKETS THROUGH FLOW MATCHING AND CO-EVOLUTIONARY DYNAMICS

Chenqing Hua[1,3] **Yong Liu**[5] **Dinghuai Zhang**[3,4] **Odin Zhang**[6]

Sitao Luan [3] **Kevin K. Yang** [7] **Guy Wolf** [3,4] **Doina Precup** [1,3,8] **Shuangjia Zheng**[2*]
[1]McGill; [2]SJTU; [3]Mila-Quebec AI Institute; [4]UdeM; [5]HKUST;
[6]Institute for Protein Design, UW; [7]Microsoft Research; [8]DeepMind

EnzymeFlow – In Submission.

EnzymeFlow paper: https://arxiv.org/pdf/2410.00327

EnzymeFlow code: https://github.com/WillHua127/EnzymeFlow

Contribution

We focus on the major contributions; additional minor ones can be found in our papers.
In ReactZyme, we (1) introduced ReactZyme dataset consisting of 178, 463 enzyme-reaction pairs of sequences and SMILES, (2) propose to compute the transition state from substrates to products using cross-attention, (3) developed a retrieval-based evaluation system for enzyme-reaction prediction, (4) conducted comprehensive benchmarking.

In EnzymeFlow, we (1) proposed EnzymeFlow model that operates on reaction/substrate-conditioned enzyme catalytic pocket design, (2) introduced EnzymeFill dataset consisting of 328,192 enzyme-reaction pairs of valid catalytic pocket structures, (3) introduced enzyme-reaction co-evolution to account for dynamic changes in catalytic process.

Motivation

For more in-depth reasoning, please refer to our publications. In general, (1) Current datasets only contain enzyme sequences and SMILES representations. There is a need for datasets that include enzyme structures and their functional sites. (2) Current protein design models focus on static protein-ligand interactions. However, enzymatic catalysis is a dynamic process involving chemical transformations of substrates into products. (3) Current enzyme function annotation evaluations may be biased, leading to unfair comparisons and validations of methods. The current limitations lead to the introduction of (1) new datasets, (2) new models, and (3) new evaluation metrics.

(1) Introduction to ReactZyme Dataset

Table 1: Comparison of ESP, EnzymeMap, and ReactZyme

Dataset	#Pair	#Enzyme	#Molecule/Reaction	Substrate Info	Product Info	Reaction Info	Atom-Mapping
ESP	18,351	12,156	1,379	✓	✗	✗	✗
EnzymeMap	46,356	12,749	16,776	✓	✓	✓	✓
ReactZyme	178,463	178,327	7,726	✓	✓	✓	✗

The ReactZyme dataset bypasses the reliance on traditional labels like EC, GO, or KEGG annotations, focusing purely on enzyme-reaction interactions. By directly using network logits, it evaluates catalytic capabilities. However, the dataset is still limited by the lack of atom-atom mapping and reduced reaction diversity, which may hinder the exploration of different functional groups.

(2) Introduction to ReactZyme Evaluation System

Table 2: Average results of baseline models of *time-based split*. Top results are highlighted in green, orange, and purple, respectively.

(a) Given the enzyme, the list of candidate reactions is evaluated (`#enzymes, #reactions`).

Time/enzyme-reaction	GNN Encoding	Top1	Top2	Top3	Top4	Top5	Top10	Top20	Top1-N	Top2-N	Top3-N	Top4-N	Top5-N	Top10-N	Top20-N	Mean Rank	MRR
Data(Ground-truth)		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5004	0.5336	0.2502	0.2002	0.1001	0.0500	1.0004	0.9998
MAT-2D + ESM	✗	0.3246	0.4526	0.5255	0.5700	0.6044	0.7079	0.7972	0.3246	0.2263	0.1752	0.1425	0.1209	0.0708	0.0399	40.4756	0.4549
MAT-2D + SaProt	✗	0.2073	0.2945	0.3408	0.3678	0.4020	0.5004	0.6120	0.2073	0.1472	0.1136	0.0937	0.0804	0.0499	0.0306	75.3546	0.2898
UniMol-2D + ESM	✗	0.2827	0.4024	0.4335	0.4889	0.5210	0.6508	0.7612	0.2827	0.2012	0.1443	0.1221	0.1041	0.0651	0.0380	53.4261	0.4011
UniMol-2D + SaProt	✗	0.1957	0.2863	0.3066	0.3622	0.3855	0.4380	0.6021	0.1957	0.1431	0.1022	0.0905	0.0771	0.0438	0.0301	79.8460	0.2788
UniMol-2D + ESM	✓	0.2948	0.4494	0.5067	0.5252	0.5866	0.6912	0.7831	0.2948	0.2247	0.1689	0.1313	0.1173	0.0691	0.0391	45.0611	0.4289
UniMol-2D + SaProt	✓	0.2512	0.3635	0.4052	0.4336	0.4329	0.6474	0.6879	0.2512	0.1818	0.1351	0.1084	0.0866	0.0647	0.0344	63.1455	0.3176
MAT-3D + ESM	✗	0.2858	0.4005	0.4344	0.4852	0.4955	0.6548	0.7405	0.2858	0.2001	0.1448	0.1213	0.0991	0.6550	0.0371	60.3628	0.4041
MAT-3D + SaProt	✗	0.1210	0.1768	0.2084	0.2226	0.2265	0.3108	0.4015	0.1210	0.0884	0.0695	0.5565	0.0453	0.0311	0.0201	150.0301	0.1862
UniMol-3D + ESM	✗	0.2905	0.4007	0.4563	0.4984	0.5365	0.6586	0.7639	0.2905	0.2004	0.1522	0.1247	0.1074	0.0659	0.0382	46.0553	0.4104
UniMol-3D + SaProt	✗	0.0916	0.1328	0.1650	0.1908	0.2134	0.2923	0.3882	0.0916	0.0664	0.0550	0.0477	0.0426	0.0292	0.0194	168.8244	0.1591
UniMol-3D + ESM	✓	0.3588	0.5158	0.5919	0.6044	0.6545	0.7815	0.8126	0.3588	0.2579	0.1973	0.1511	0.1309	0.0781	0.0406	32.7443	0.4952
UniMol-3D + SaProt	✓	0.2508	0.3528	0.3995	0.4016	0.4075	0.5448	0.6421	0.2508	0.1764	0.1331	0.1004	0.0815	0.0546	0.0321	59.8345	0.3453

(b) Given the reaction, the list of candidate enzymes is evaluated (`#reactions, #enzymes`).

Time/reaction-enzyme	GNN Encoding	Top1	Top2	Top3	Top4	Top5	Top10	Top20	Top1-N	Top2-N	Top3-N	Top4-N	Top5-N	Top10-N	Top20-N	Mean Rank	MRR
Data(Ground-truth)		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7775	0.6377	0.5420	0.4718	0.2895	0.1677	2.8324	0.7497
MAT-2D + ESM	✗	0.2175	0.2733	0.3144	0.3493	0.3815	0.4924	0.6033	0.2175	0.2001	0.1817	0.1688	0.1570	0.1206	0.0871	165.3066	0.1789
MAT-2D + SaProt	✗	0.1260	0.1537	0.1791	0.1943	0.2153	0.2921	0.3778	0.1260	0.1126	0.1035	0.0943	0.0886	0.0716	0.0546	281.2419	0.0981
UniMol-2D + ESM	✗	0.1435	0.1773	0.1977	0.2239	0.2299	0.3554	0.4367	0.1435	0.1299	0.1143	0.1087	0.0946	0.0871	0.0631	270.9385	0.1233
UniMol-2D + SaProt	✗	0.0912	0.1194	0.1342	0.1444	0.1494	0.2252	0.3488	0.0912	0.0875	0.0776	0.0701	0.0615	0.0552	0.0504	536.5624	0.0805
UniMol-2D + ESM	✓	0.1486	0.1788	0.2092	0.2250	0.2294	0.3529	0.4865	0.1486	0.1310	0.1209	0.1092	0.0944	0.0865	0.0703	254.1982	0.1257
UniMol-2D + SaProt	✓	0.0988	0.1284	0.1458	0.1572	0.1587	0.2273	0.3536	0.0988	0.0941	0.0843	0.0763	0.0653	0.0557	0.0511	504.2854	0.0934
MAT-3D + ESM	✗	0.2281	0.3041	0.3518	0.3945	0.4240	0.5502	0.5879	0.2281	0.2097	0.1933	0.1818	0.1703	0.1393	0.0852	152.1328	0.1931
MAT-3D + SaProt	✗	0.1037	0.1372	0.1629	0.1738	0.1800	0.2603	0.3671	0.1037	0.0946	0.0895	0.0801	0.0723	0.0659	0.0532	411.5762	0.1056
UniMol-3D + ESM	✗	0.1678	0.2240	0.2631	0.2938	0.3155	0.3960	0.5011	0.1678	0.1543	0.1443	0.1349	0.1267	0.1002	0.0748	177.4881	0.1400
UniMol-3D + SaProt	✗	0.0558	0.0721	0.0815	0.0883	0.0979	0.1359	0.1918	0.0558	0.0497	0.0448	0.0407	0.0393	0.0344	0.0278	700.9714	0.0538
UniMol-3D + ESM	✓	0.2045	0.2835	0.3398	0.3722	0.3792	0.4475	0.5168	0.2045	0.1955	0.1867	0.1715	0.1523	0.1133	0.0749	167.5862	0.1628
UniMol-3D + SaProt	✓	0.1331	0.1750	0.1886	0.1979	0.2044	0.3365	0.4119	0.1331	0.1207	0.1036	0.0912	0.0821	0.0852	0.0597	322.5755	0.1122

The strength of ReactZyme lies in its design as a ranking-based task, which avoids the traditional reliance on metrics like accuracy or precision, or the engagement of negative samples. By focusing on ranking, this system sidesteps common issues found in discriminative tasks and adopts metrics that better capture the model's effectiveness. This ranking-based

approach provides a more accurate and fair evaluation by removing biases that often stem from using binary labels (i.e., 0s and 1s). As a result, the evaluation process becomes more objective, allowing for meaningful comparisons and ensuring fairness in benchmarking.

(3) Introduction to ReactZyme Benchmarking

ReactZyme is benchmarked on a wide range of baselines, from classical annotation methods to deep learning models. These are protein language models (ESM and SaProt), molecule foundation models (UniMol-2D, UniMol-3D, MAT-2D, MAT-3D), MLP, transformer, bi-RNN, BLASTp, contrastive learning techniques, pseudo-graph methods, fingerprint representations, and many others. ReactZyme benchmarking is comprehensive.

(4) Introduction to EnzymeFill Dataset

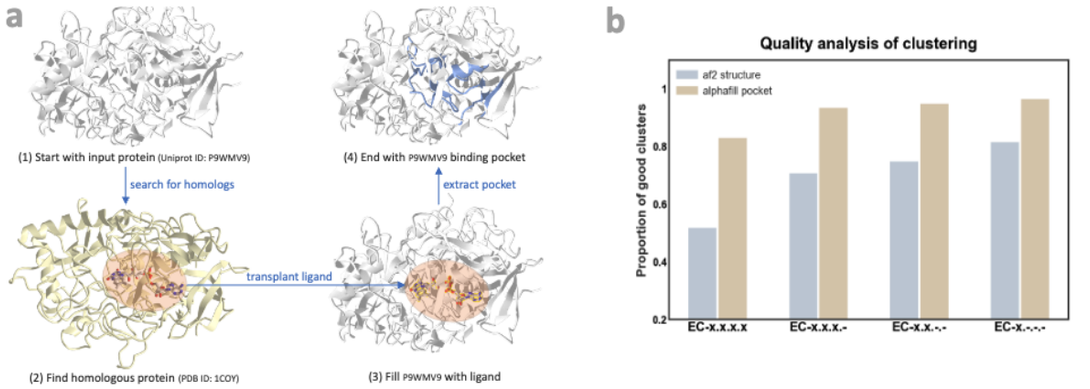


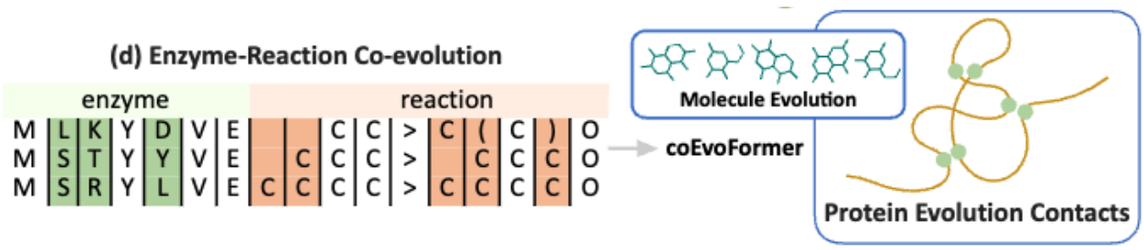
Figure 4: (a) Enzyme pocket extraction workflow with AlphaFill. (b) Quality analysis of clustering between enzyme pockets and full structures; good clusters have high functional concentration.

Data	Reaction	Enzyme	Substrate		Product		Enzyme Commission Class						
	#reaction	#enzyme	#substrate	#avg atom	#product	#avg atom	EC1	EC2	EC3	EC4	EC5	EC6	EC7
Rawdata	232520	97912	7259	30.81	7664	30.34	44881 (19.30%)	75944 (32.66%)	37728 (16.23%)	47242 (20.32%)	8315 (3.58%)	18281 (7.86%)	129 (0.06%)
40% Homo	19379	6922	4798	31.06	4897	30.24	4754 (24.53%)	5857 (30.22%)	4839 (24.97%)	1764 (9.10%)	759 (3.92%)	1379 (7.12%)	27 (0.14%)
50% Homo	34750	13442	5675	31.45	5871	30.75	8184 (23.55%)	11174 (32.16%)	8050 (23.17%)	3203 (9.22%)	1357 (3.91%)	2752 (7.92%)	30 (0.09%)
60% Homo	53483	22350	6112	30.95	6331	30.34	11674 (21.83%)	18419 (34.44%)	11394 (21.30%)	5555 (10.39%)	2194 (4.10%)	4200 (7.85%)	47 (0.09%)
80% Homo	100925	43458	6619	30.46	6943	29.95	21308 (21.11%)	34344 (34.03%)	18925 (18.75%)	14010 (13.88%)	3901 (3.87%)	8371 (8.29%)	66 (0.07%)
90% Homo	132047	55697	6928	30.32	7298	29.81	28833 (21.84%)	43287 (32.78%)	23989 (18.17%)	20070 (15.20%)	5015 (3.80%)	10766 (8.15%)	87 (0.07%)

EnzymeFill builds upon the ReactZyme dataset by not only expanding the number of enzyme-reaction pairs but also providing atom-atom mappings and catalytic pocket structures for pocket design. EnzymeFill consists of 328,192 enzyme-reaction pairs, where catalytic pocket structures are determined using AlphaFill with ligand transplantation. The catalytic pockets are defined by selecting the 32 residues closest to the ligand within a 10-angstrom radius. Pockets containing fewer than 32 residues are excluded, resulting in a total of 232,520 valid catalytic pockets. Additionally, EnzymeFill is debiased through enzyme homology filtering, where sequence alignment is used to categorize enzymes into five homology levels.

(5) Introduction to Enzyme-Reaction Co-Evolution

> **In EnzymeFlow, co-evolutionary relationships are computed by aligning enzyme sequences and reaction SMILES.**



EnzymeFlow is the first work to introduce the high-level concept of co-evolution in synthetic and computational biology. We leverage this idea to explore evolutionary relationships between enzymes and reactions. For example, if both Enzyme A and Enzyme B can catalyze Reaction 1, there may be evolutionary information linking these enzymes. Similarly, if Enzyme A can catalyze both Reaction 1 and Reaction 2, there may be evolutionary connections between these reactions. This concept extends to a broader idea of enzyme-reaction co-evolutionary information. In EnzymeFlow, co-evolutionary relationships are computed by aligning enzyme sequences and reaction SMILES. This co-evolution framework is key to capturing and defining the functional roles of enzymes in catalytic processes, forming the foundation of the EnzymeFlow model.

(6) Introduction to EnzymeFlow

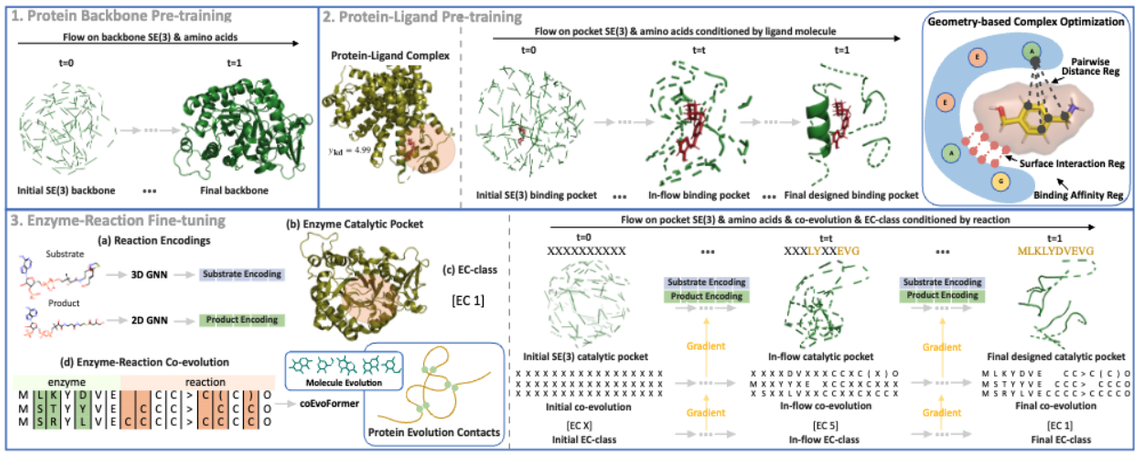


Figure 2: Overview of EnzymeFlow with hierarchical pre-training and enzyme-reaction co-evolution. **(1)** Flow model pre-trained on protein backbones and amino acid types. **(2)** Flow model further pre-trained on protein binding pockets, conditioned on ligand molecules with geometry-specific optimization. **(3)** Flow model fine-tuned on enzyme catalytic pockets, and conditioned on substrate and product molecules, with enzyme-reaction co-evolution and EC-class generation.

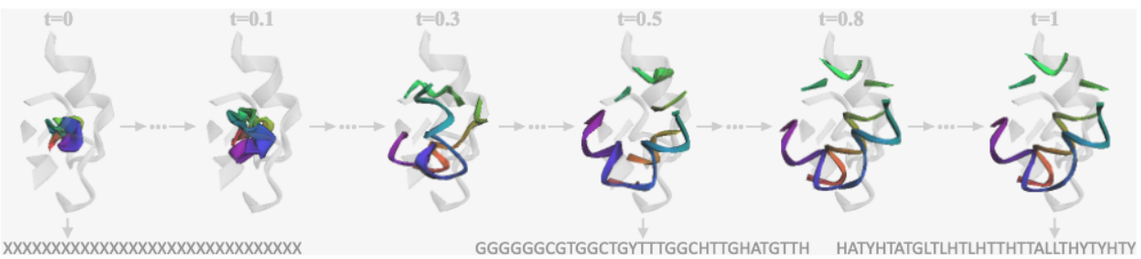


Figure 3: Catalytic pocket design example using EnzymeFlow (UniProt: Q7U4P2). The pocket generation is conditioned on reaction `CN[C@H](C(=O)C)CS.C/C=C\\1/C(=C/c2[nH]c(c(c2C)CCC(=O)O)/C=C/2\\N=C(C(=C2CCC(=O)O)O)C)C[C@H]2NC(=O)C(=C2C)C=C)/NC(=O)[C@@H]1C` → `CN[C@H](C(=O)C)CSC(C1=C(C)C(=O)N[C@H]1Cc1[nH]c(c(c1C)CCC(=O)O)/C=C/1\\N=C(C(=C1CCC(=O)O)O)C)C[C@H]1NC(=O)C(=C1C)C=C)C` of EC4 (ligase enzyme), from $t = 0$ to $t = 1$.

EnzymeFlow is a flow matching model designed for reaction-conditioned enzyme catalytic pocket generation. It operates on the catalytic pocket backbone, amino acid types, enzyme commission numbers, and enzyme-reaction co-evolution. Additionally, EnzymeFlow is pretrained on protein backbone structures and protein-ligand complexes for generalizability.

(6) EnzymeFlow vs. RFDiffusionAA

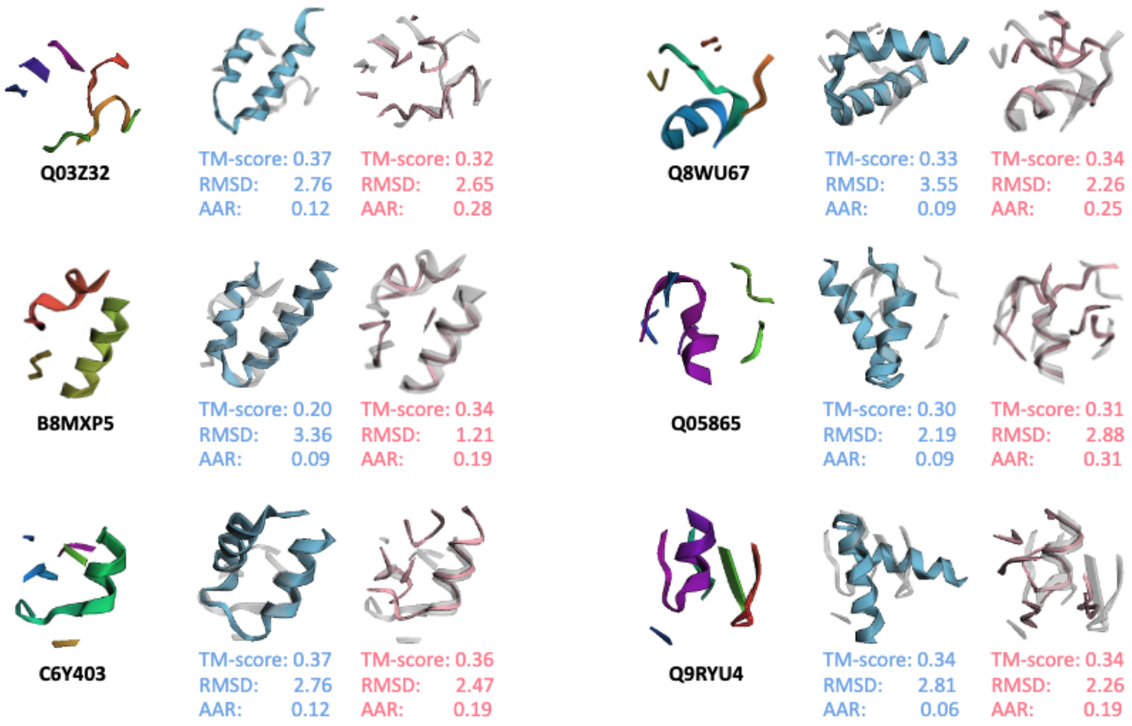


Figure 14: Visualization and comparison between RFDiffusionAA-designed pockets and EnzymeFlow-designed pockets after superimposition with ground-truth pockets. Light color represents EnzymeFlow-designed pockets, blue color represents RFDiffusionAA-designed pockets, spectral color represents the ground-truth reference pockets. `TM-score`, `RMSD`, `AAR` are reported.

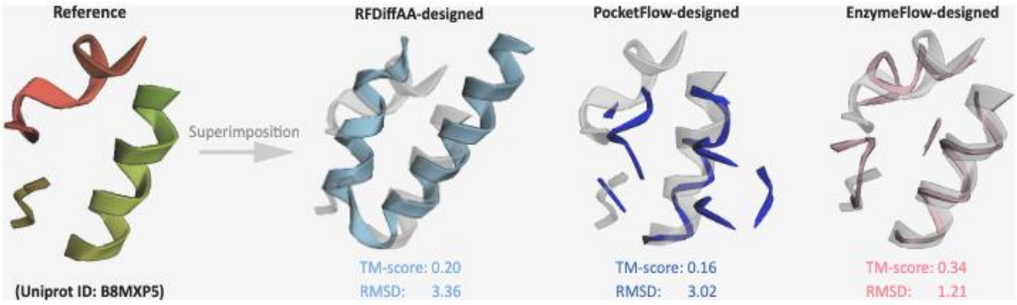


Figure 5: Case study of catalytic pocket design (UniProt: B8MXP5). We show the reference and designed pockets of different models. The pocket generation is conditioned on reaction `OC[C@H]1O[C@@H](Oc2ccccc2/C=C\\C(=O)O)[C@@H]([C@H]([C@@H]1O)O)O` → `OC(=O)/C=C\\c1ccccc1O` of EC3.

Table 2: Evaluation of structural validity of EnzymeFlow- and baseline-generated catalytic pockets. The binding affinities (`Kd`) and structural confidence (`chai`) are computed by performing docking on the catalytic pocket and substrate conformation using Vina (Trott & Olson, 2010) and Chai (Chai, 2024), respectively. We highlight top three results in **bold**, underline, and *italic*, respectively.

Model	cRMSD (↓)			TM-score (↑)			Kd (↓)	chai (↑)	AAR (↑)	ECacc (↑)
	Top1	Top10	Median	Top1	Top10	Median				
Eval Data		-			-		-4.65	-	-	-
DEPACT	9.25	9.75	11.16	0.238	0.206	0.149	-5.46	0.125	0.112	0.149
PocketGen	7.65	8.14	10.45	0.260	0.233	0.193	-5.01	0.121	0.176	0.152
RFDiffusionAA	9.13	9.77	11.92	0.269	0.245	0.198	**-12.71**	**0.232**	0.153	0.170
PocketFlow	7.42	8.09	10.01	0.268	*0.260*	0.197	-4.93	0.123	*0.207*	0.166
EnzymeFlow (T=50)	**6.94**	**7.57**	9.04	**0.290**	**0.262**	**0.209**	-5.03	0.129	**0.216**	**0.280**
w/o coevo	7.02	7.60	*9.15*	0.288	*0.260*	*0.205*	-4.86	0.123	0.196	0.246
w/o pretraining	*7.01*	*7.69*	9.29	*0.286*	0.261	0.207	-4.33	*0.134*	0.202	*0.255*
w/o coevo+pretraining	7.05	7.81	9.43	0.278	0.255	0.204	-4.72	0.125	0.154	0.221
EnzymeFlow (T=100)	6.97	**7.57**	**9.02**	0.283	0.258	0.207	*-5.31*	0.135	0.215	0.273

We visualize comparison of EnzymeFlow-generated pockets and RFDiffusionAA-generated pockets. Analysis of enzyme functions can be found in the manuscript. From both functional and structural perspectives, the *function-based, reaction-conditioned* EnzymeFlow outperforms current *structure-based, substrate-conditioned* protein design models in both structural validity and intended function design (catalytic ability).

(7) What's Next?

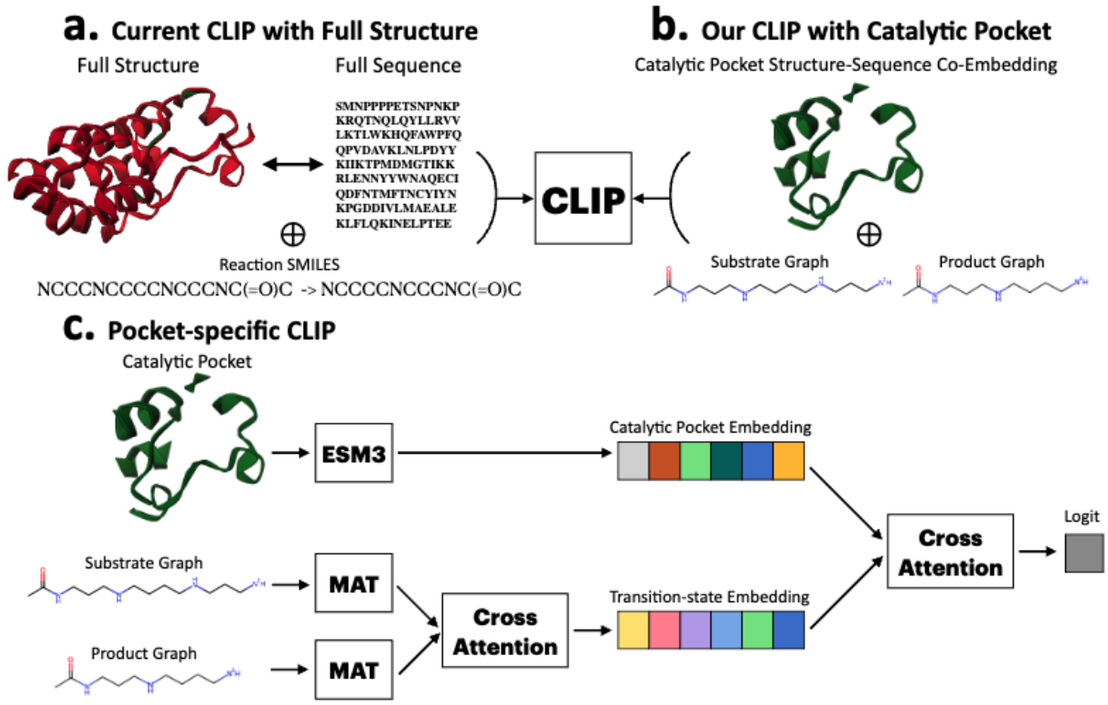


Figure 11: Enzyme-Reaction CLIP model comparison. (a) Existing CLIP models use the full enzyme structure or full enzyme sequence, paired with reaction SMILES as input. (b) Our pocket-specific CLIP model focuses on catalytic pockets, using both their structures and sequences paired with molecular graphs of reactions. The pocket-specific CLIP approach learns from enzyme active sites, which exhibit higher functional concentration. (c) Overview of Pocket-specific CLIP model.

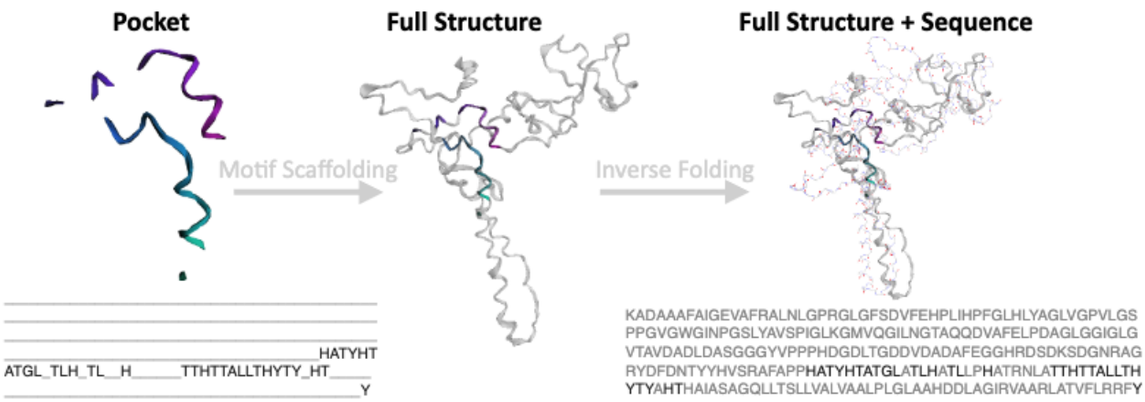


Figure 12: Inpainting catalytic pocket using ESM3.

We are actively working on (1) pocket-specific enzyme-reaction CLIP model, and (2) fine-tuned large protein model for catalytic pocket inpainting (functional (multi-)motif scaffolding). Our goal is to develop an end-to-end automated AI-driven enzyme discovery system:

1. Catalytic Pocket Design: The system will first design enzyme catalytic pockets.

2. Scaffolding Functional Motifs: Next, it will scaffold the functional motifs to generate full enzyme structures.

3. Substrate Binding: The system will bind substrates to the catalytic pockets.

4. Inverse Folding: The enzyme-substrate complex will undergo inverse folding.

5. Computational Screening: Finally, the system will perform computational screening through enzyme CLIP to select the best-generated enzymes.