

Table 1: Evaluation on **cumulative concentrations** while using **AdaBN**

Method	Explicit Concentration	Modality	Unseen Images	Unseen Images + Unseen Molecules	Unseen Dataset (0-shot)	Avg.	Unseen Images	Unseen Images + Unseen Molecules	Unseen Dataset (0-shot)	Avg.
			top-1% active molecules				top-5% active molecules			
MolPhenix	-	AdaBN	.8568	.5336	.3525	.581	.9562	.7603	.5772	.7646
MolPhenix	sigmoid	AdaBN	.911	.5858	.4	.6323	.971	.7997	.6203	.797
MolPhenix	logarithm	AdaBN	.9155	.6106	.4242	.6501	.9729	.8332	.6503	.8188
MolPhenix	one-hot	AdaBN	.9187	.6125	.4225	.6512	.9744	.8302	.6419	.8155
			top-1% all molecules				top-5% all molecules			
MolPhenix	-	AdaBN	.4593	.2409	.1599	.2867	.5983	.4081	.285	.4305
MolPhenix	sigmoid	AdaBN	.5104	.3142	.1957	.3401	.6496	.5165	.331	.499
MolPhenix	logarithm	AdaBN	.5379	.3393	.2071	.3614	.6867	.5561	.3606	.5345
MolPhenix	one-hot	AdaBN	.5476	.3425	.2082	.3661	.7007	.5641	.3603	.5417

Table 2: Evaluation on **cumulative concentrations** while combining **Mol1, ECFP, rdKit, and MACCS**.

Method	Explicit Concentration	Modality	Unseen Images	Unseen Images + Unseen Molecules	Unseen Dataset (0-shot)	Avg.	Unseen Images	Unseen Images + Unseen Molecules	Unseen Dataset (0-shot)	Avg.
			top-1% active molecules				top-5% active molecules			
MolPhenix	-	Ph1 & Mol1 & 3 fps	.9185	.7212	.4717	.7038	.9784	.8805	.718	.859
MolPhenix	sigmoid	Ph1 & Mol1 & 3 fps	.9395	.7408	.5119	.7307	.9817	.8932	.7458	.8736
MolPhenix	logarithm	Ph1 & Mol1 & 3 fps	.9454	.7798	.5658	.7637	.9815	.9163	.7849	.8942
MolPhenix	one-hot	Ph1 & Mol1 & 3 fps	.9419	.7687	.5526	.7544	.9807	.9113	.7681	.8867
			top-1% all molecules				top-5% all molecules			
MolPhenix	-	Ph1 & Mol1 & 3 fps	.4764	.3011	.2068	.3281	.604	.4647	.3415	.4701
MolPhenix	sigmoid	Ph1 & Mol1 & 3 fps	.5076	.342	.2382	.3626	.6383	.521	.3769	.512
MolPhenix	logarithm	Ph1 & Mol1 & 3 fps	.525	.379	.2648	.3896	.658	.5743	.411	.5478
MolPhenix	one-hot	Ph1 & Mol1 & 3 fps	.5355	.3845	.265	.395	.6862	.5916	.4233	.567

Table 3: Evaluation on **heldout concentrations** while combining **Mol1, ECFP, rdKit, and MACCS**.

Method	Explicit Concentration	Modality	Unseen Images	Unseen Images + Unseen Molecules	Unseen Dataset (0-shot)	Avg.	Unseen Images	Unseen Images + Unseen Molecules	Unseen Dataset (0-shot)	Avg.
			top-1% active molecules				top-5% active molecules			
MolPhenix	-	Ph1 & Mol1 & 3 fps	.8364	.5115	.4263	.5914	.9640	.7363	.6850	.7951
MolPhenix	sigmoid	Ph1 & Mol1 & 3 fps	.8300	.5021	.4363	.5895	.9640	.7409	.6931	.7993
MolPhenix	logarithm	Ph1 & Mol1 & 3 fps	.8112	.5107	.4376	.5865	.9544	.7406	.6866	.7939
MolPhenix	one-hot	Ph1 & Mol1 & 3 fps	.7467	.4409	.3830	.5235	.9320	.6827	.6520	.7556
			top-1% all molecules				top-5% all molecules			
MolPhenix	-	Ph1 & Mol1 & 3 fps	.5339	.1980	.1966	.3095	.6968	.2909	.4274	.4717
MolPhenix	sigmoid	Ph1 & Mol1 & 3 fps	.5463	.2026	.2066	.3185	.7179	.3116	.4359	.4885
MolPhenix	logarithm	Ph1 & Mol1 & 3 fps	.5247	.2009	.2078	.3111	.7067	.3133	.4319	.4840
MolPhenix	one-hot	Ph1 & Mol1 & 3 fps	.4690	.1653	.1756	.2700	.6635	.2592	.4118	.4448

Table 4: Comparison of MolPhenix molecular embedding with traditional fingerprints on different tasks of TDC and Polaris datasets. Mean results for TDC, Polaris and together are available in the last three columns.

			concentration																																					
			adme-fang-HCLint-1	adme-fang-HPPB-1	adme-fang-PERM-1	adme-fang-RCLint-1	adme-fang-RPPB-1	adme-fang-SOLL-1	ames	bbb_martins	bioavailability_ma	caco2_wang	clearance_hepatocyte_az	clearance_microsome_az	cyp2d6_substrate_carbonmangels	cyp2d6_veith	cyp2d6_substrate_carbonmangels	cyp2d6_veith	cyp3a4_substrate_carbonmangels	cyp3a4_veith	dili	half_life_obach	herg	hla_bou	ld50_zhu	lipophilicity_astrazeneca	pdp_brocetelli	pka2-egfr-wt-e-1	pka2-egfr-wt-e-1	pka2-kt-wt-e-1	pka2-kt-wt-e-1	pka2-ret-wt-e-1	pka2-ret-wt-e-1	ppbr_az	solubility_aqso1b	vdss_lombardo	TDC Mean	Polaris Mean	Mean	
		metric	pearson	pearson	pearson	pearson	pearson	pearson	auroc	auroc	auroc	mae	spearman	spearman	auroc	auroc	auroc	auroc	auroc	auroc	spearman	auroc	auroc	auroc	mae	mae	auroc	auroc	pearson	auroc	auroc	auroc	auroc	mae	mae	spearman				
		rdkit	0.32	0.34	0.48	0.23	0.38	0.29	0.69	0.72	0.58	-0.54	0.25	0.45	0.31	0.45	0.45	0.29	0.54	0.59	0.71	0.26	0.61	0.71	-0.70	-0.84	0.76	0.20	0.42	0.33	0.53	0.36	0.45	-13.03	-1.63	0.22	0.21	0.36	0.26	
		ecfp	0.46	0.60	0.49	0.43	0.60	0.39	0.69	0.75	0.48	-0.43	0.37	0.50	0.32	0.52	0.44	0.33	0.60	0.64	0.67	0.47	0.73	0.65	-0.73	-0.78	0.79	0.41	0.57	0.33	0.51	0.40	0.55	-9.91	-1.27	0.47	0.28	0.48	0.35	
		maccs	0.37	0.56	0.52	0.22	0.43	0.44	0.71	0.77	0.53	-0.47	0.35	0.42	0.32	0.49	0.45	0.32	0.62	0.61	0.75	0.43	0.66	0.70	-0.66	-0.83	0.79	0.21	0.35	0.25	0.32	0.44	0.49	-10.13	-1.47	0.46	0.26	0.38	0.30	
		Concatnated fps	0.41	0.66	0.58	0.33	0.40	0.37	0.70	0.77	0.58	-0.43	0.38	0.52	0.33	0.54	0.42	0.33	0.57	0.62	0.74	0.45	0.70	0.72	-0.67	-0.80	0.84	0.36	0.56	0.34	0.57	0.44	0.57	-10.94	-1.46	0.48	0.28	0.47	0.35	
		Molphenix fingerprint	0.57	0.63	0.61	0.51	0.70	0.55	0.70	0.76	0.47	-0.47	0.28	0.49	0.35	0.56	0.50	0.36	0.58	0.63	0.72	0.40	0.69	0.68	-0.71	-0.73	0.82	0.28	0.46	0.31	0.39	0.43	0.50	-9.36	-1.11	0.53	0.29	0.50	0.36	
		Molphenix fingerprint	1	0.57	0.75	0.57	0.55	0.72	0.57	0.70	0.74	0.54	-0.48	0.29	0.46	0.32	0.57	0.47	0.38	0.59	0.64	0.77	0.55	0.67	0.69	-0.71	-0.70	0.80	0.20	0.41	0.30	0.43	0.31	0.39	-8.93	-1.10	0.55	0.30	0.48	0.36
		Molphenix fingerprint	10	0.63	0.78	0.60	0.62	0.68	0.60	0.69	0.75	0.55	-0.45	0.32	0.52	0.34	0.58	0.51	0.41	0.62	0.66	0.73	0.44	0.70	0.77	-0.68	-0.68	0.83	0.19	0.40	0.31	0.41	0.42	0.46	-8.00	-1.03	0.51	0.31	0.51	0.38
		Molphenix fingerprint	25	0.64	0.71	0.65	0.62	0.67	0.58	0.69	0.78	0.54	-0.42	0.30	0.43	0.32	0.56	0.49	0.42	0.60	0.67	0.77	0.38	0.69	0.74	-0.67	-0.66	0.84	0.17	0.42	0.32	0.39	0.37	0.46	-8.43	-1.02	0.50	0.30	0.50	0.37