Supplementary Materials

AlloFusion: Allosteric Site Prediction Based on Language Models and Multi-Feature Fusion

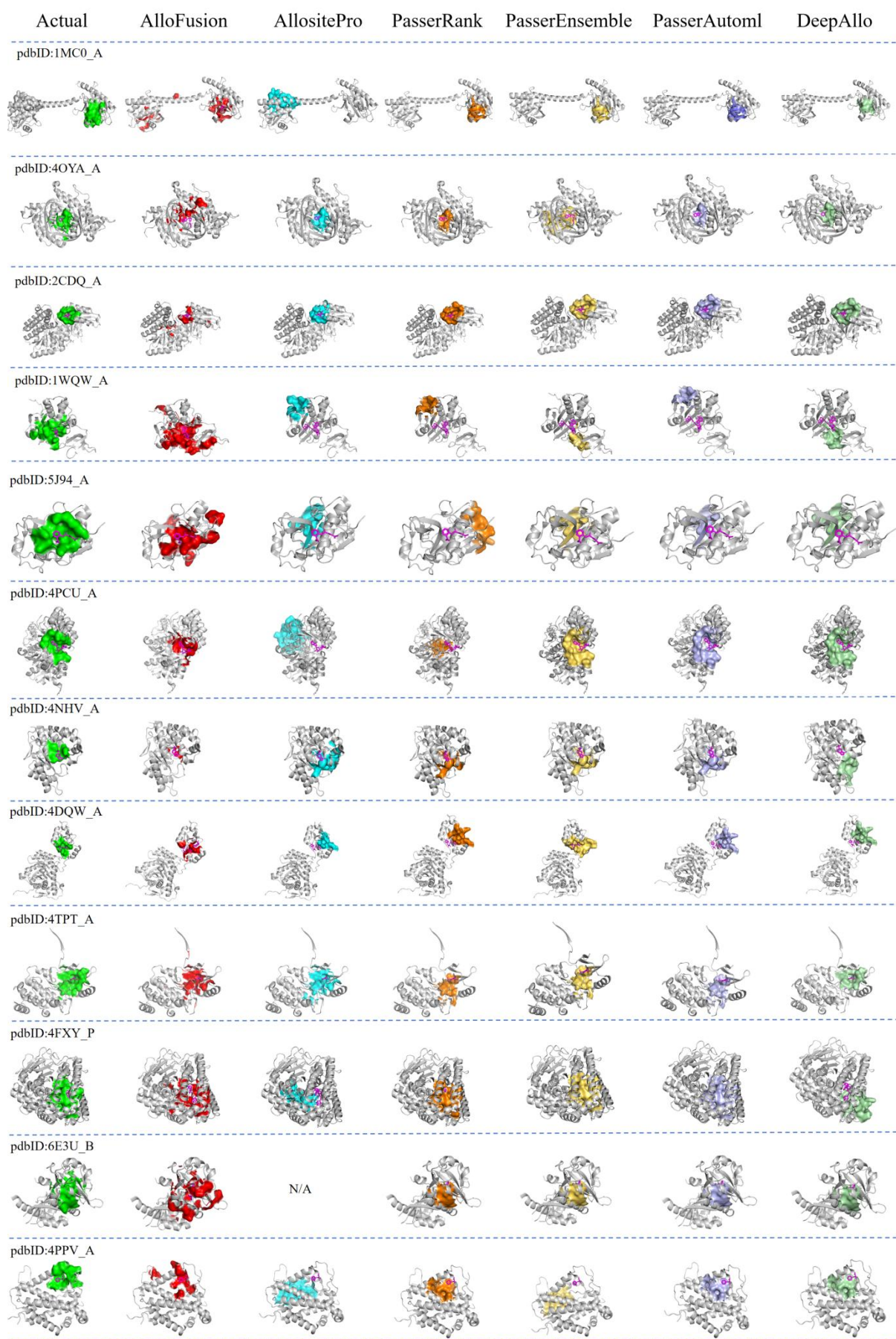**Jiabin Huang[†], Dongliang Guo[‡], Yapeng Liu, Yanfen Wang, Mengya Lv**

[†]*School of Information Science and Engineering, Yanshan University, Qinhuangdao 066000, China*

[‡]Corresponding authors: dongliangguo@ysu.edu.cn

Table S1. The independent test set D24 of 24 allosteric proteins

No.	Protein Name	PDB ID	Chain	Residue Num.
1	3',5'-cyclic nucleotide phosphodiesterase 2A	1MC0	A	369
2	Adenylate cyclase type 10	4OYA	A	470
3	Aspartokinase 1, chloroplastic	2CDQ	A	510
4	BirA(Pyrococcus horikoshii)	1WQW	A	235
5	Cathepsin K	5J94	A	223
6	Cystathionine beta-synthase	4PCU	A	549
7	Dihydropteroate synthase	4NHV	A	297
8	Inosine-5'-monophosp hate dehydrogenase	4DQW	A	509
9	LIM domain kinase 2	4TPT	A	303
10	Neurolysin, mitochondrial	4FXY	P	693
11	Heterodimeric HIF-2	6E3U	B	642
12	Chorismate mutase 1, chloroplastic	4PPV	A	280
13	NAD-dependent protein deacetylase sirtuin-1	5BTR	A	397
14	Threonine synthase 1, chloroplastic	2C2B	A	486

15	Ubiquitin-conjugating enzyme E2 D1	4QPL	A	155
16	Serine/threonine-prote in kinase PAK 1	4ZLO	A	298
17	Metabotropic glutamate receptor 5,Endolysin,Metabotr opic glutamate receptor 5	5CGC	A	444
18	Phosphoenolpyruvate carboxykinase, cytosolic [GTP]	4YW8	A	624
19	Isocitrate dehydrogenase [NADP] cytoplasmic	5DE1	A	413
20	Nuclear receptor ROR-gamma	5C4T	A	241
21	Thyroid hormone receptor alpha	3UVV	B	244
22	GTP pyrophosphokinase YjbM	5DED	A	218
23	DynE7	2XEM	B	150
24	TmaDAH7PS	3PG9	B	338

Actual	AlloFusion	AllositePro	PasserRank	PasserEnsemble	PasserAutoml	DeepAllo

pdbID:1MC0_A

pdbID:4OYA_A

pdbID:2CDQ_A

pdbID:1WQW_A

pdbID:5J94_A

pdbID:4PCU_A

pdbID:4NHV_A

pdbID:4DQW_A

pdbID:4TPT_A

pdbID:4FXY_P

pdbID:6E3U_B

N/A

pdbID:4PPV_A

pdbID:5BTR_A

pdbID:2C2B_A

pdbID:4QPL_A

pdbID:4ZLO_A

pdbID:5CGC_A

pdbID:4YW8_A

pdbID:5DE1_A

pdbID:5C4T_A

pdbID:3UVV_B

pdbID:5DED_A

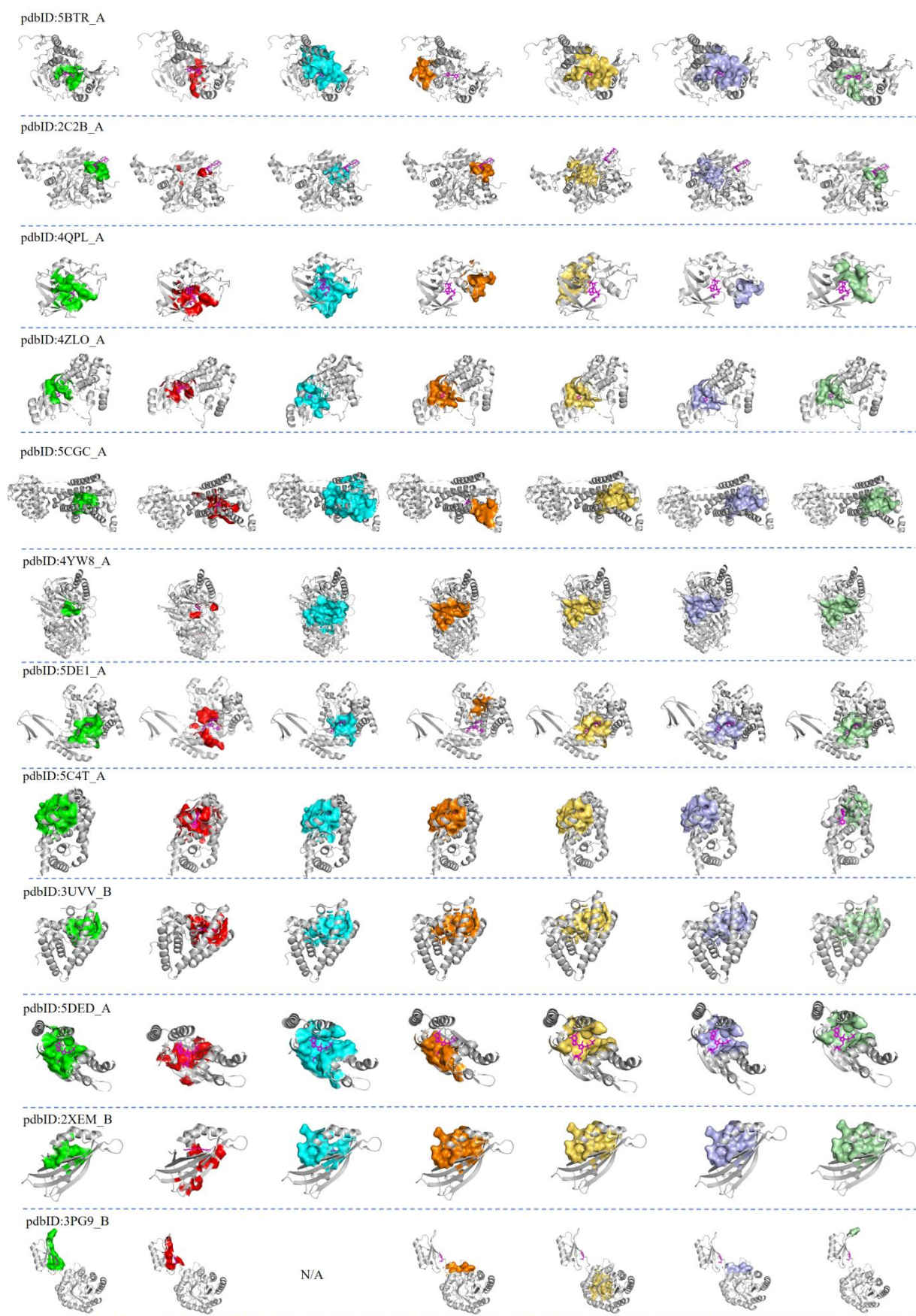pdbID:2XEM_B

pdbID:3PG9_B

N/A

Figure S1. 3D representations of protein AFRs predicted by different models on the D24 dataset. Each row, from left to right, displays the experimentally determined AFRs, followed

5

by predictions from AlloFusion, AllositePro, PASSer Rank, PASSer Ensemble, PASSer Automl and DeepAllo. Green indicates the actual allosteric sites composed of AFRs; red, cyan, orange, light yellow, light blue and light green represent the predicted AFRs by AlloFusion, AllositePro, PASSer Rank, PASSer Ensemble, PASSer Automl and DeepAllo, respectively. Gray represents the cartoon structure of the protein. "N/A" denotes cases where no allosteric site is predicted by the corresponding method.
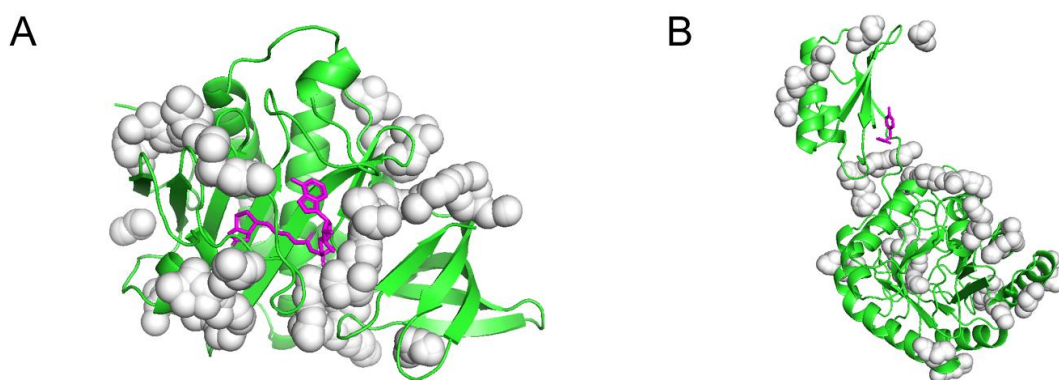
Figure S2. Pocket predictions by Fpocket on proteins (A) PDB ID: 5J94 and (B) PDB ID: 3PG9. Allosteric modulators (purple sticks) are not located within the predicted pockets (white spheres), indicating that these modulators may bind to non-pocket regions.

Table S2. Key parameters set in LoRAConfig for fine-tuning the ProtT5 model.

Parameters	Settings	Rationale for Selection
lora_rank	4	Reduces the number of trainable parameters to ensure a balance between performance and computational efficiency.
lora_init_scale	0.01	Small-scale initialization prevents gradient explosion and maintains training stability.
lora_modules	.*SelfAttention|.*EncDecAttention	Focuses fine-tuning on attention layers to better capture long-range dependencies relevant to allosteric site recognition.
lora_layers	q|k|v|o	Covers all components of the attention mechanism, enhancing the model's ability to learn both local and global information.
trainable_param_names	.*layer_norm.*|.*lora_[ab].*	Restricts fine-tuning to LoRA and LayerNorm layers, reducing computational overhead.
lora_scaling_rank	1	Maintains stable parameter updates and accommodates small-scale LoRA adjustments.

Table S3. Distances between the predicted allosteric site centroids and the actual binding site on D24 using different methods.

PDB_Code	Chain	AlloFusion	AllositePro	PASSer_rank	PASSer Ensemble	PASSer_Automl	DeepAllo
1MC0	A	1.78	63.37	2.31	2.31	2.31	4.78
1WQW	A	2.6	15.9	24.18	16.53	24.18	12.69
2C2B	P	15.23	27.56	0.43	26.75	26.75	0.43
2CDQ	A	9.97	4.52	3.37	3.37	3.37	3.37
2XEM	B	8.25	2.31	2.83	2.83	2.83	2.83
3PG9	B	2.59	N/A	22.74	45.23	22.74	14.39
3UVV	B	1.48	2.58	2.43	2.43	2.43	2.43
4DQW	A	4.25	9.32	9.21	3.13	9.21	9.21
4FXY	P	6.72	31.97	1.01	1.01	1.01	15.97
4NHV	A	1.71	14.3	13.37	13.37	13.37	23.69
4OYA	A	6.96	2.81	24.79	21.81	24.79	2.43
4PCU	A	10.27	54.18	38.69	3.41	3.41	3.41
4PPV	A	13.03	22.31	12.87	23.95	12.87	12.87
4QPL	A	2.98	8.78	19.16	13.96	19.16	8.78
4TPT	A	3.11	1.72	4.58	4.58	4.58	4.68
4YW8	A	13.73	8.04	54.39	54.39	54.39	6.99
4ZLO	A	5.78	2.69	2.28	2.28	2.28	2.28
5BTR	A	12.59	6.23	13.22	3.19	3.19	13.22
5C4T	A	8.15	0.68	0.68	0.68	0.68	14.57
5CGC	A	4.53	4.5	16.31	4.77	4.77	4.77
5DE1	A	13.40	2.35	16.26	2.83	2.83	2.83
5DED	A	5.69	4.16	6.2	3.61	3.61	3.61
5J94	A	10.96	26.64	27.27	23.12	23.12	23.12
6E3U	B	6.15	N/A	2.77	2.77	2.77	2.77