Supporting Information

Unlocking the application potential of AlphaFold3-like approaches in virtual screening

Chao Shen[1,2,3,*], Xujun Zhang[2,4], Shukai Gu[2,4], Odin Zhang[2,4], Qinghan Wang[2,4], Gang Du[2,4], Yihao Zhao[2,4], Linlong Jiang[2,4], Peichen Pan[2,4], Yu Kang[2,4], Qingwei Zhao[1,3], Chang-Yu Hsieh[1,2,4], Tingjun Hou[1,2,4*]

[1]Department of Clinical Pharmacy, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, Zhejiang, China
[2]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China
[3]Zhejiang Provincial Key Laboratory for Drug Evaluation and Clinical Research, Hangzhou 310003, Zhejiang, China
[4]Zhejiang Provincial Key Laboratory for Intelligent Drug Discovery and Development, Jinhua 321016, Zhejiang, China

Corresponding authors

Tingjun Hou
E-mail: tingjunhou@zju.edu.cn
Chao Shen
E-mail: shenchao513@zju.edu.cn

Table S1 Basic information of *GPCR$_{recent}$* dataset. MW, LogP, RB, HBA, HBR and HALX represent the average molecular weight, Wildman-Crippen partition coefficient, number of rotatable bonds, number of hydrogen-bond acceptors, number of hydrogen-bond donors and number of halogens for active and decoy compounds, respectively. To evaluate scaffold diversity for actives, the Tanimoto similarity of each compound to all others was computed based on ECFP4 fingerprints and Murcko scaffolds. The average of these pairwise similarities for all active compounds within a target was then recorded as the scaffold diversity score for that target.

Uniprot ID	Target ID	PDB entry	Release time of first crystal structure	Target Information	Number		MW		LogP		RB		HBA		HBR		HAL		Scaffold diversity for actives
					Act	Dec	Act	Dec	Act	Dec	Act	Dec	Act	Dec	Act	Dec	Act	Dec	
O15552	FFAR2	8j22	2024/01/24	Free fatty acid receptor 2	205	10250	418.10	419.15	3.76	3.76	5.17	5.22	5.46	5.49	1.00	1.01	0.67	0.67	0.245
P0DMS8	ADORA3	8x16	2024/4/24	Adenosine receptor A3AR	300	15000	415.01	416.37	2.62	2.63	5.14	5.20	7.57	7.53	2.24	2.22	0.34	0.33	0.150
P13945	ADRB3	9ije	2024/8/21	Beta-3 adrenergic receptor	300	15000	375.82	376.87	3.70	3.70	5.45	5.48	3.91	4.00	0.90	0.91	0.40	0.39	0.131
P20309	CHRM3	8e9w	2022/11/30	Muscarinic acetylcholine receptor M3	300	15000	375.82	376.87	3.70	3.70	5.45	5.48	3.91	4.01	0.90	0.91	0.40	0.39	0.131
P25021	HRH2	8yn3	2022/6/29	Histamine H2 receptor	300	15000	383.12	384.24	3.24	3.23	7.28	7.31	4.44	4.55	2.31	2.27	0.35	0.34	0.156
P29275	ADORA2B	8hdo	2023/1/18	Adenosine A2b receptor	300	15000	393.06	394.15	2.62	2.62	4.96	5.01	7.14	7.10	1.64	1.63	0.49	0.47	0.138
P30874	SSTR2	7wig	2022/3/9	Somatostatin receptor type 2	268	13400	515.08	517.18	4.52	4.54	7.43	7.50	5.51	5.63	2.28	2.26	0.97	0.95	0.194
P35348	ADRA1A	8thk	2023/07/05	Alpha-1A adrenergic receptor	300	15000	376.73	377.69	3.42	3.43	5.25	5.30	4.47	4.54	1.02	1.02	0.51	0.50	0.145
P35372	OPRM1	8efo	2022/11/9	Mu-type opioid receptor	300	15000	414.41	415.75	3.65	3.67	5.32	5.37	4.72	4.80	1.42	1.42	0.36	0.36	0.134
P46098	HTR3A	8bla	2024/3/6	Human serotonin 5-HT3A	300	15000	341.32	341.89	2.87	2.86	3.17	3.25	4.47	4.53	1.01	1.01	0.43	0.43	0.156
P49682	CXCR3	8hnm	2023/11/29	C-X-C chemokine receptor type 3	300	15000	510.71	512.43	4.43	4.47	6.45	6.54	6.01	6.07	0.74	0.76	1.79	1.76	0.172
P50406	HTR6	7ys6	2022/7/27	5-hydroxytryptamine receptor 6	300	15000	388.58	389.44	3.71	3.69	4.60	4.65	4.52	4.59	1.00	1.00	0.58	0.57	0.141
Q5NUL3	FFAR4	8h4k	2023/3/8	Free fatty acid receptor 4	300	14996	386.75	387.32	4.95	4.90	7.01	7.04	3.72	3.76	1.04	1.03	1.38	1.36	0.196
Q8TDV5	GPR119	7wcm	2022/08/24	Glucose-dependent insulinotropic receptor	300	14914	464.99	467.14	3.87	3.88	5.64	5.62	7.19	7.21	0.43	0.44	0.74	0.73	0.169
Q96RJ0	TAAR1	9jkq	2023/11/15	Trace amine-associated receptor 1	300	15000	330.85	331.19	2.96	2.91	4.44	4.48	4.18	4.24	1.48	1.46	0.61	0.60	0.144
Q9Y5N1	HRH3	8yuu	2022/10/26	Histamine H3 receptor	300	15000	367.06	367.93	3.51	3.51	5.70	5.72	4.13	4.21	0.89	0.88	0.27	0.27	0.147

Table S2 Basic information of the subset of LIT-PCBA dataset used in this study.

Target	Target name	Actives	Inactives	PDB entry
ESR_ago	Estrogen receptor α	13	4378	2P15
ESR_antago	Estrogen receptor α	88	3820	2IOK
TP53	Cellular tumor antigen p53	24	4071	3ZME
VDR	Vitamin D receptor	64	3345	3A2I
MAPK1	Mitogen-activated protein kinase 1	308	61567	4ZZN

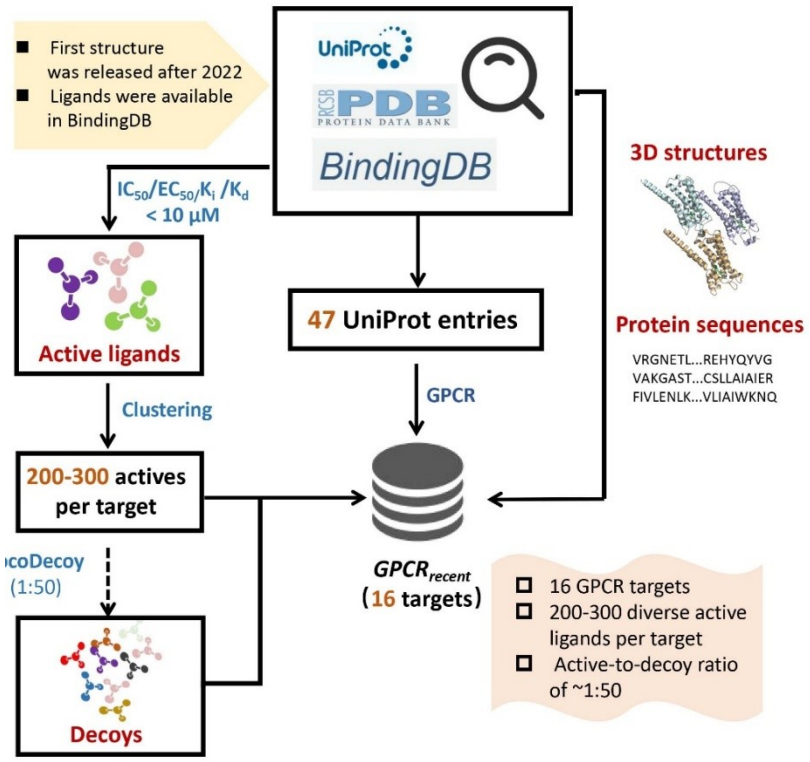

Figure S1. Workflow to construct the *GPCR*$_{recent}$ dataset.

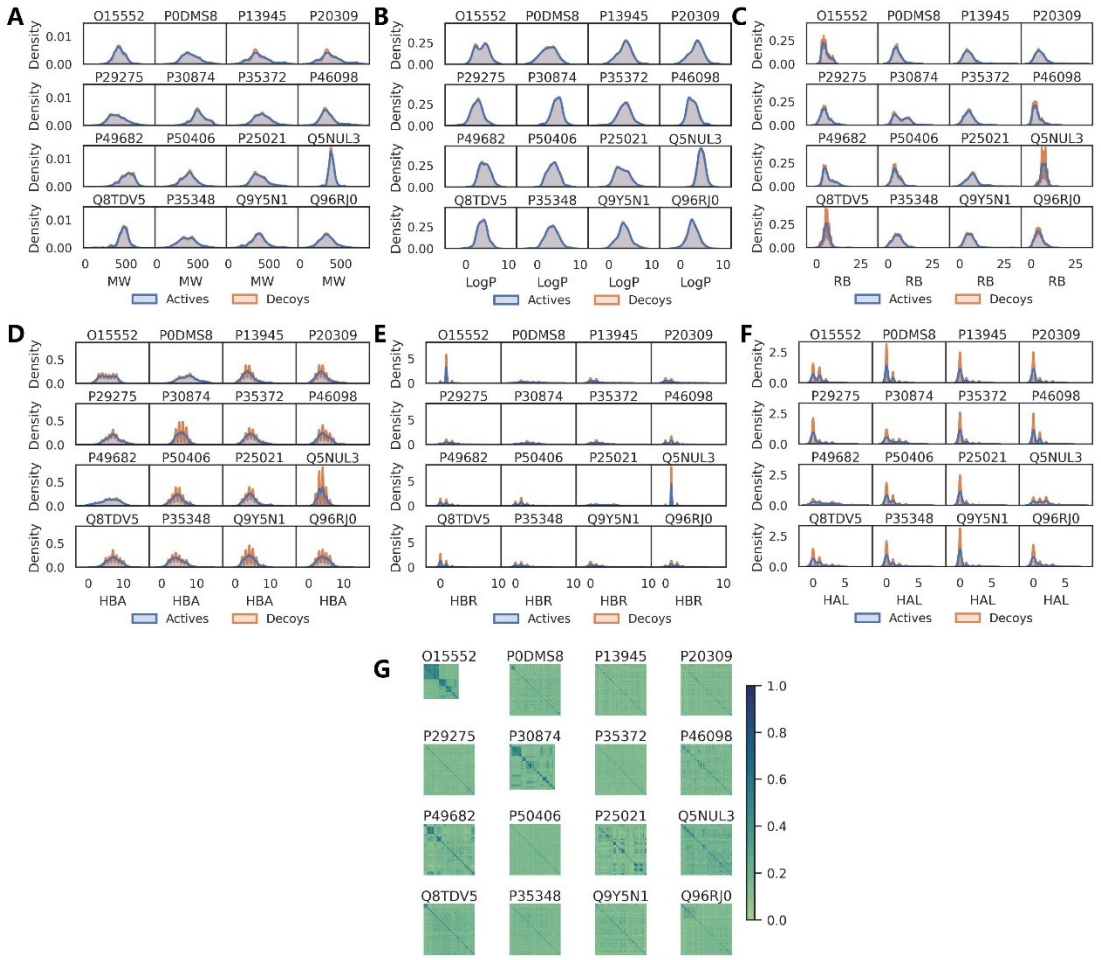

Figure S2. Molecular property distributions and scaffold diversity of active compounds in the *GPCR*$_{recent}$ dataset. (A–F) show the distributions of key physicochemical properties: molecular weight (MW), Wildman–Crippen partition coefficient (LogP), number of rotatable bonds (RB), number of hydrogen-bond acceptors (HBA), number of hydrogen-bond donors (HBD), and number of halogen atoms (HAL). (G) Scaffold diversity was calculated as the pairwise Tanimoto similarity based on ECFP4 fingerprints and Murcko scaffolds across different targets.

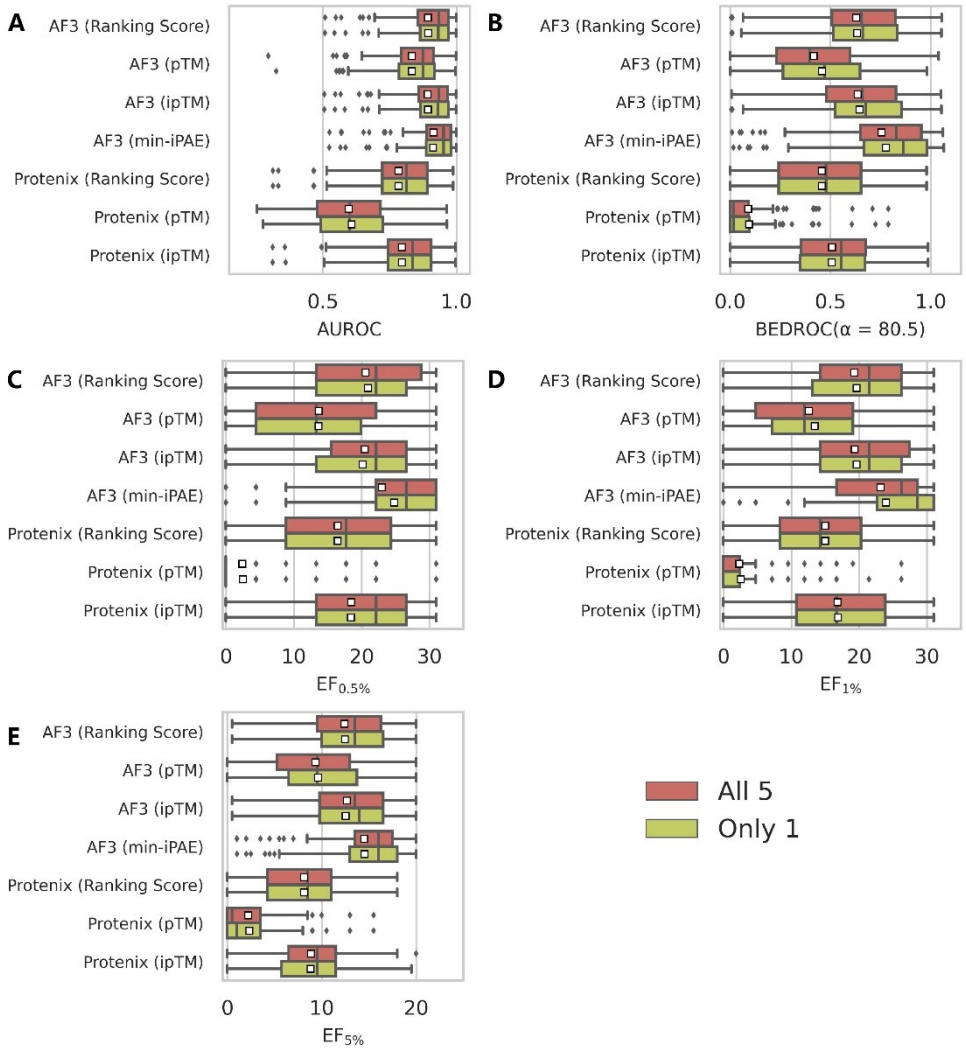

Figure S3. Impact of the number of diffusion samples on performance of AF3 and Protenix on the DEKOIS2.0 benchmark set (N=79). Evaluation metrics include (A) AUROC, (B) BEDROC (a = 80.5), (C) $EF_{0.5\%}$, (D) $EF_{1\%}$, and (E) $EF_{5\%}$. "All 5" denotes the strategy where the top-ranked prediction (by the model's internal Ranking Score) among five samples was selected as the final pose, while "Only 1" refers to using the first generated prediction directly, mimicking the scenario when only one sample was generated using AF3/Protenix. White squares within the box plots represent the mean value for each metric.

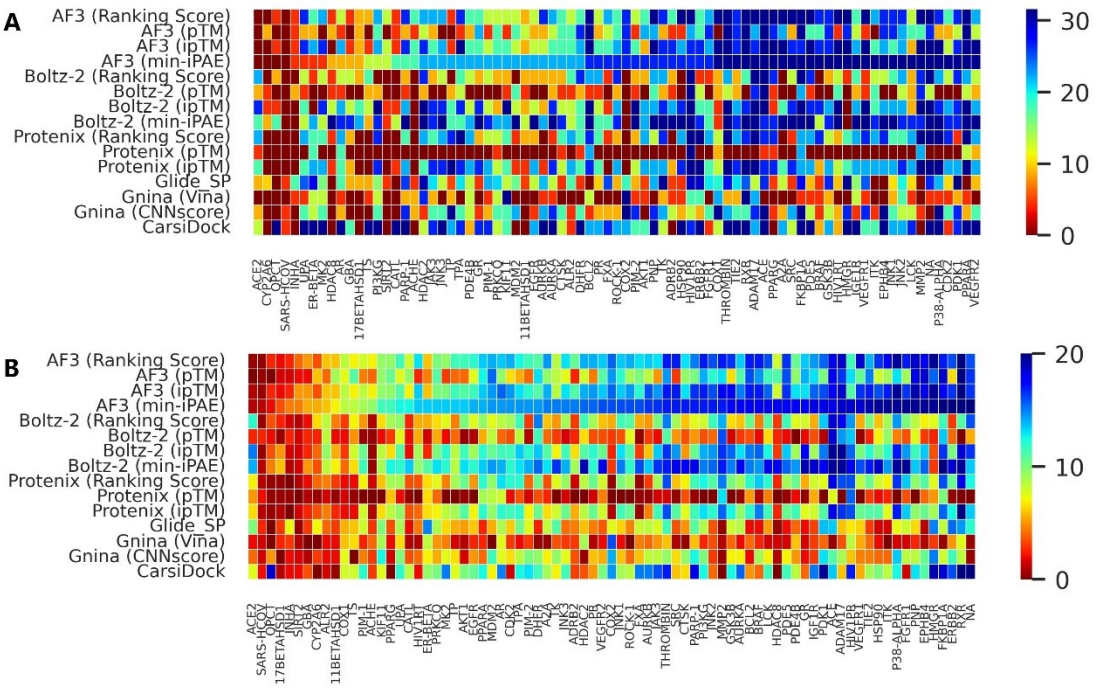

Figure S4. Performance distribution of several representative screening approaches across all targets in DEKOIS2.0 benchmark set (N=79), evaluated by (A) $EF_{0.5\%}$, and (B) $EF_{5\%}$. Targets are sorted in ascending order based on the performance of AF3 (min-iPAE) for each individual metric.

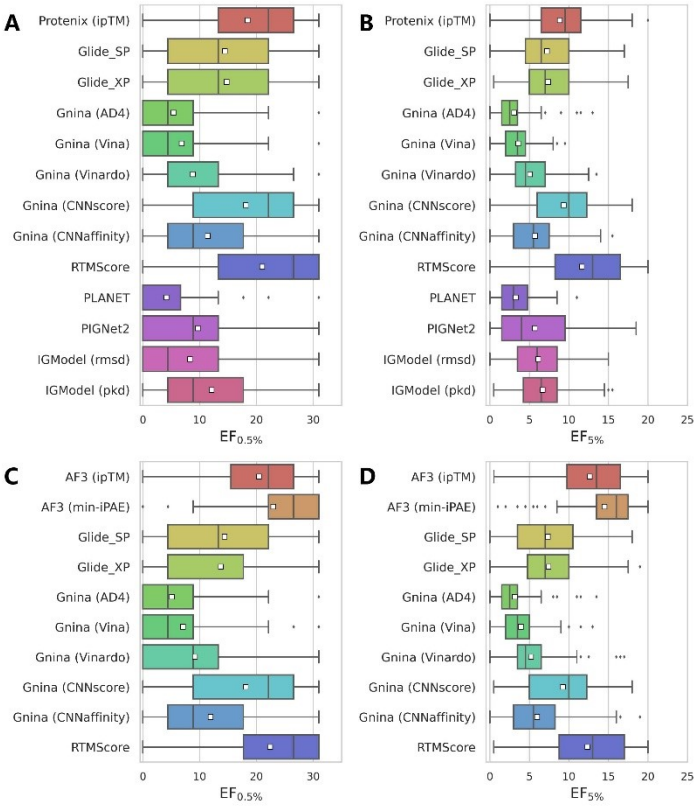

Figure S5. Performance comparison of multiple scoring approaches applied to structures predicted by (A-B) Protenix and (C-D) AF3 on the DEKOIS2.0 benchmark set (N=79). Performance metrics include (A, C) $EF_{0.5\%}$, and (B, D) $EF_{5\%}$. White squares in box plots indicate mean values for each metric.

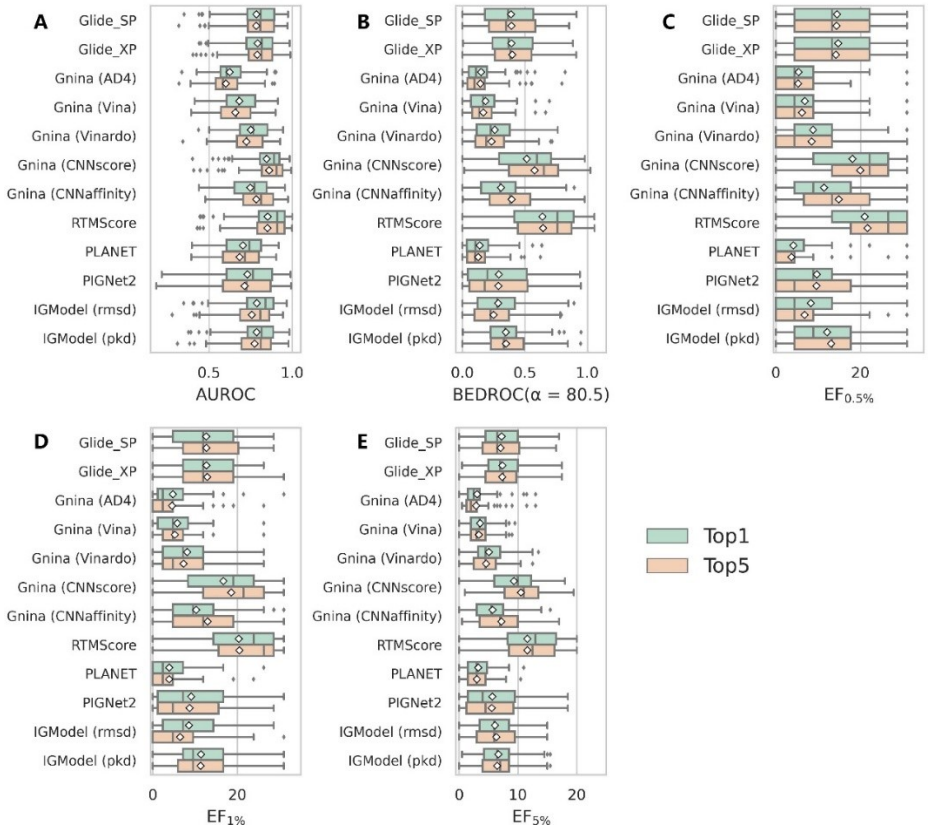

Figure S6. Impact of the number of Protenix-predicted structures on performance of multiple scoring approaches on the DEKOIS2.0 benchmark set (N=79). Evaluation metrics include (A) AUROC, (B) BEDROC (a = 80.5), (C) $EF_{0.5\%}$, (D) $EF_{1\%}$, and (E) $EF_{5\%}$. White squares in box plots indicate mean values for each metric.

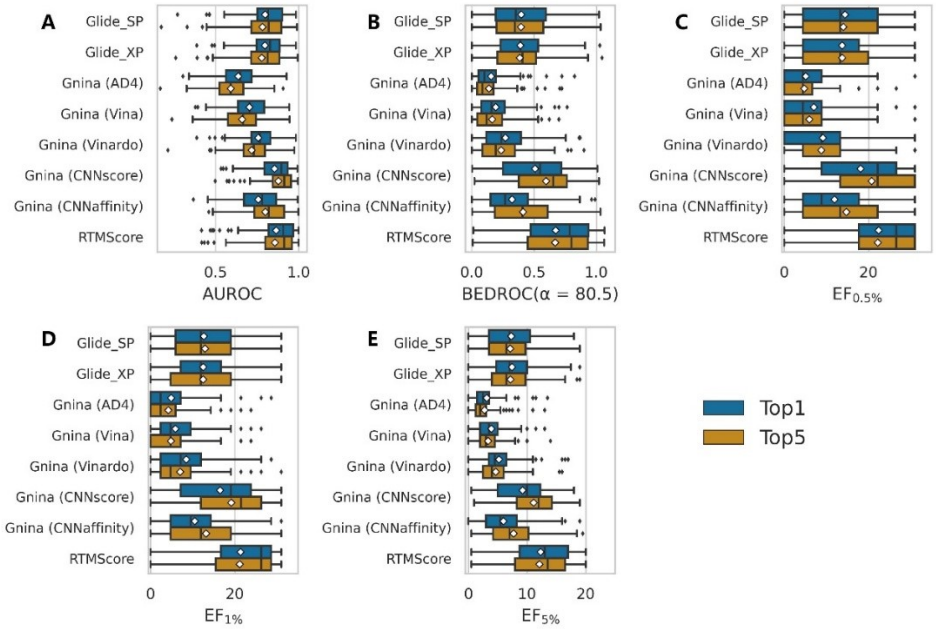

Figure S7. Impact of the number of AF3-predicted structures on performance of multiple scoring approaches on the DEKOIS2.0 benchmark set (N=79). Evaluation metrics include (A) AUROC, (B) BEDROC (a = 80.5), (C) $EF_{0.5\%}$, (D) $EF_{1\%}$, and (E) $EF_{5\%}$. White squares in box plots indicate mean values for each metric.

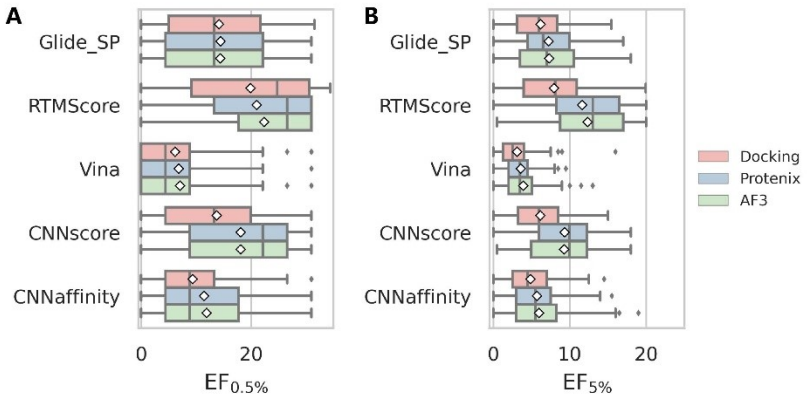

Figure S8. Evaluation of pose generation methods in screening performance, indicated by (A) $EF_{0.5\%}$, and (B) $EF_{5\%}$. The docking engine employed for RTMScore rescoring is Glide SP. White squares in box plots indicate mean values for each metric.

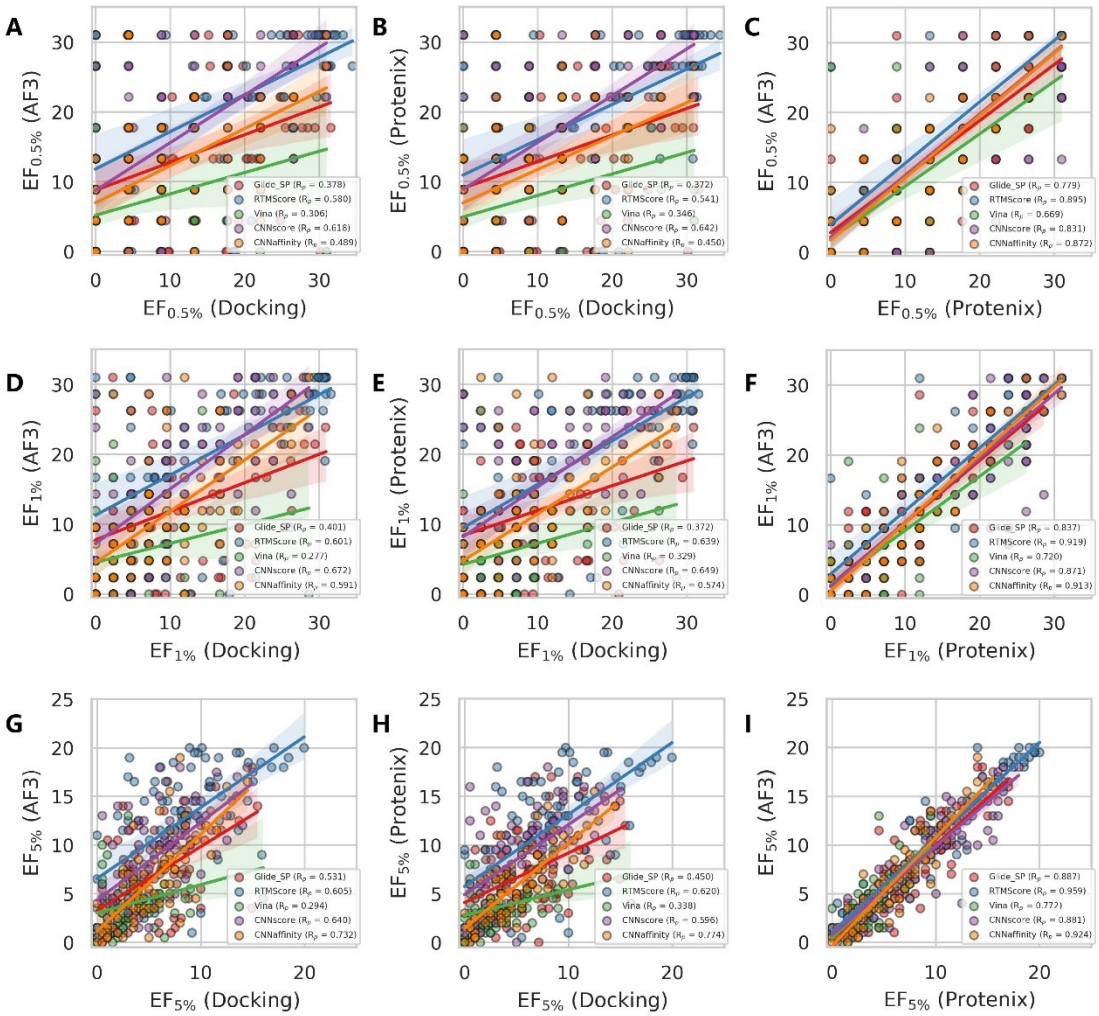

Figure S9. Pairwise correlation analysis of (A-C) $EF_{0.5\%}$, (D-F) $EF_{1\%}$ and (G-I) $EF_{5\%}$ values between different approaches: (A, D, G) Glide/Gnina *vs* AF3, (B, E, H) Glide/Gnina *vs* Protenix, and (C, F, I) AF3 *vs* Protenix. The docking engine employed for RTMScore rescoring is Glide SP.

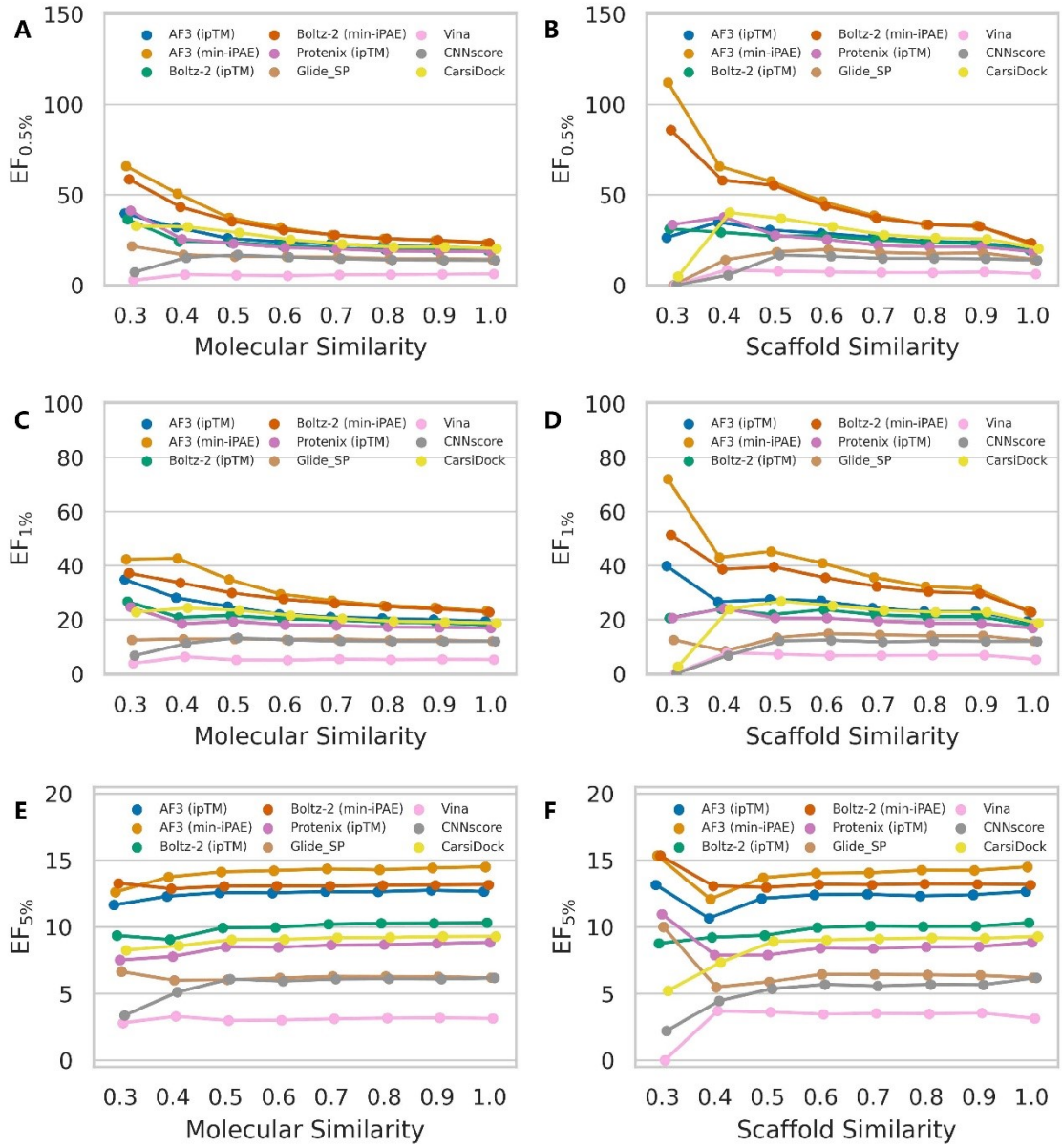

Figure S10. Impact of ligand similarity on the screening performance of multiple docking tools. Performance metrics include (A-B) $EF_{0.5\%}$, (C-D) $EF_{1\%}$ and (E-F) $EF_{5\%}$. For each active compound in the DEKOIS2.0 dataset, ECFP4 fingerprints were computed either for (A, C, E) the entire molecule or (B, D, F) the Murcko scaffold. The minimum Tanimoto similarity to all ligands in the RCSB PDB was then determined. Data point represents the mean value across targets for each metric. Targets for which no active compounds met the specified similarity threshold were excluded from metric calculations.

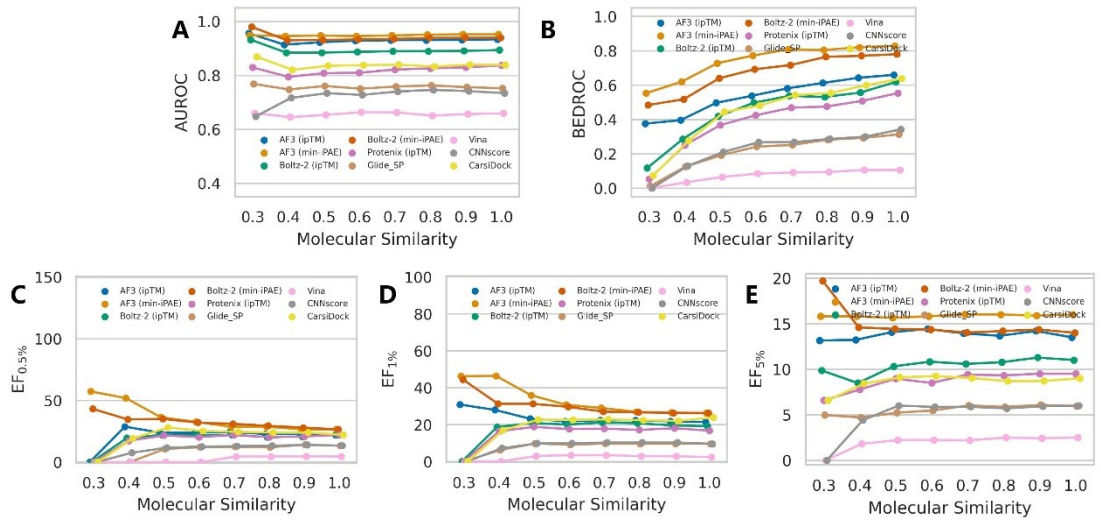

Figure S11. Impact of ligand similarity on the screening performance of multiple docking tools. Performance metrics include (A) AUROC, (B) BEDROC (a = 80.5), (C) $EF_{0.5\%}$, (D) $EF_{1\%}$ and (E) $EF_{5\%}$. For each active compound in the DEKOIS2.0 dataset, ECFP4 fingerprints were computed the entire molecule. The minimum Tanimoto similarity to all ligands in the RCSB PDB was then determined. Data point represents the median value across targets for each metric. Targets for which no active compounds met the specified similarity threshold were excluded from metric calculations.

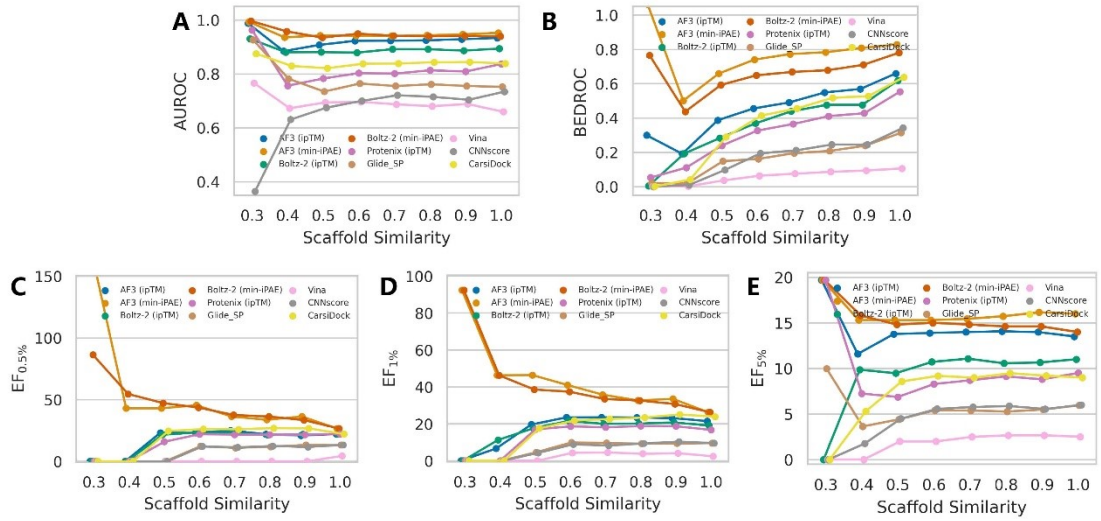

Figure S12. Impact of ligand similarity on the screening performance of multiple docking tools. Performance metrics include (A) AUROC, (B) BEDROC (a = 80.5), (C) $EF_{0.5\%}$, (D) $EF_{1\%}$ and (E) $EF_{5\%}$. For each active compound in the DEKOIS2.0 dataset, ECFP4 fingerprints were computed the Murcko scaffold. The minimum Tanimoto similarity to all ligands in the RCSB PDB was then determined. Data point represents the median value across targets for each metric. Targets for which no active compounds met the specified similarity threshold were excluded from metric calculations.

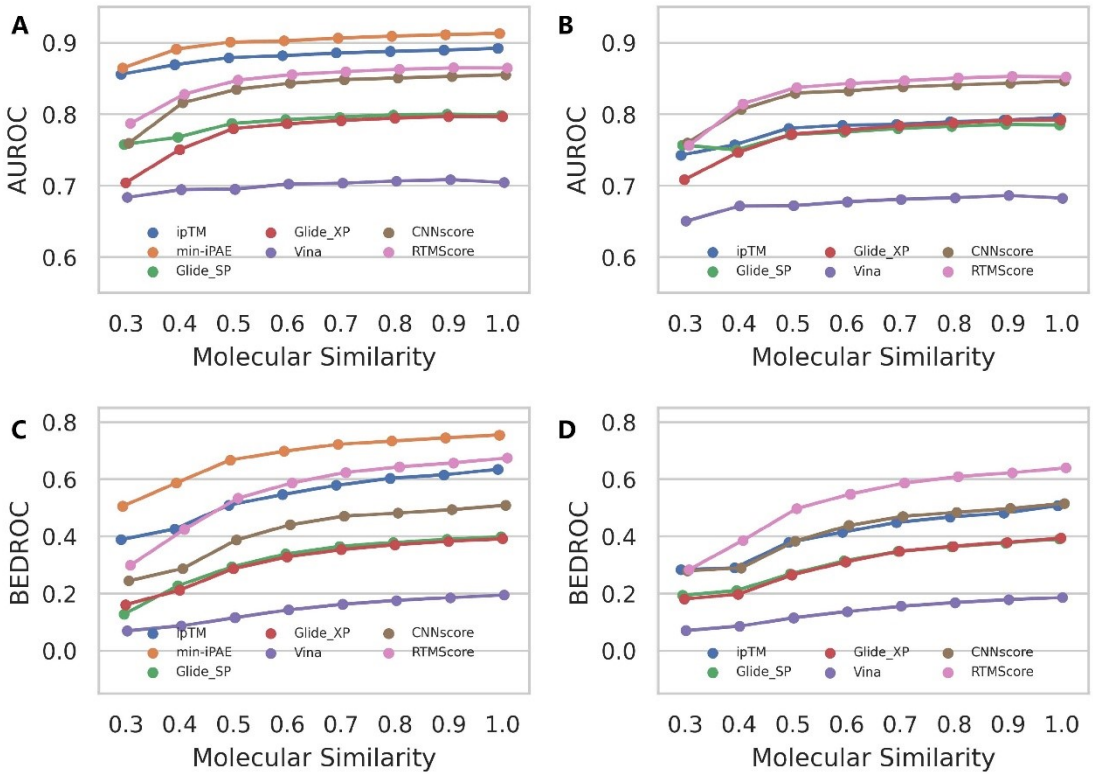

Figure S13. Impact of ligand similarity on the screening performance of multiple scoring schemes based on the structures predicted by (A, C) AF3 and (B, D) Protenix. Performance metrics include (A-B) AUROC and (C-D) BEDROC (a = 80.5). For each active compound in the DEKOIS2.0 dataset, ECFP4 fingerprints were computed for the entire molecule. The minimum Tanimoto similarity to all ligands in the RCSB PDB was then determined. Data point represents the mean value across targets for each metric. Targets for which no active compounds met the specified similarity threshold were excluded from metric calculations.

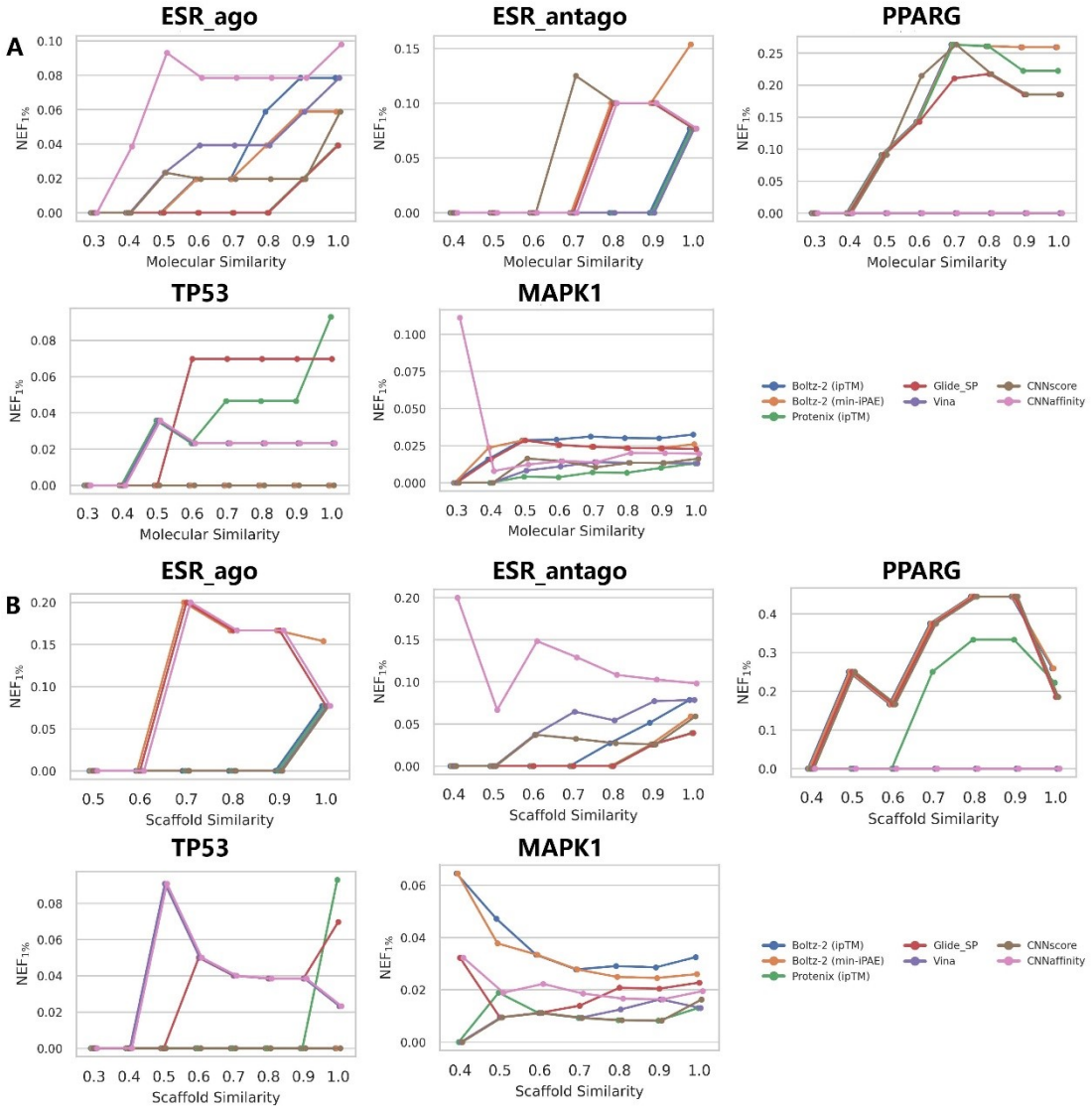

Figure S14. Impact of ligand similarity on the screening performance evaluated by $NEF_{1\%}$ across five targets from the LIT-PCBA dataset. For each active compound, ECFP4 fingerprints were computed for (A) the entire molecule or (B) the Murcko scaffold. The minimum Tanimoto similarity to all ligands in the RCSB PDB was then determined. Targets lacking active compounds above the specified similarity threshold were excluded from metric calculations.

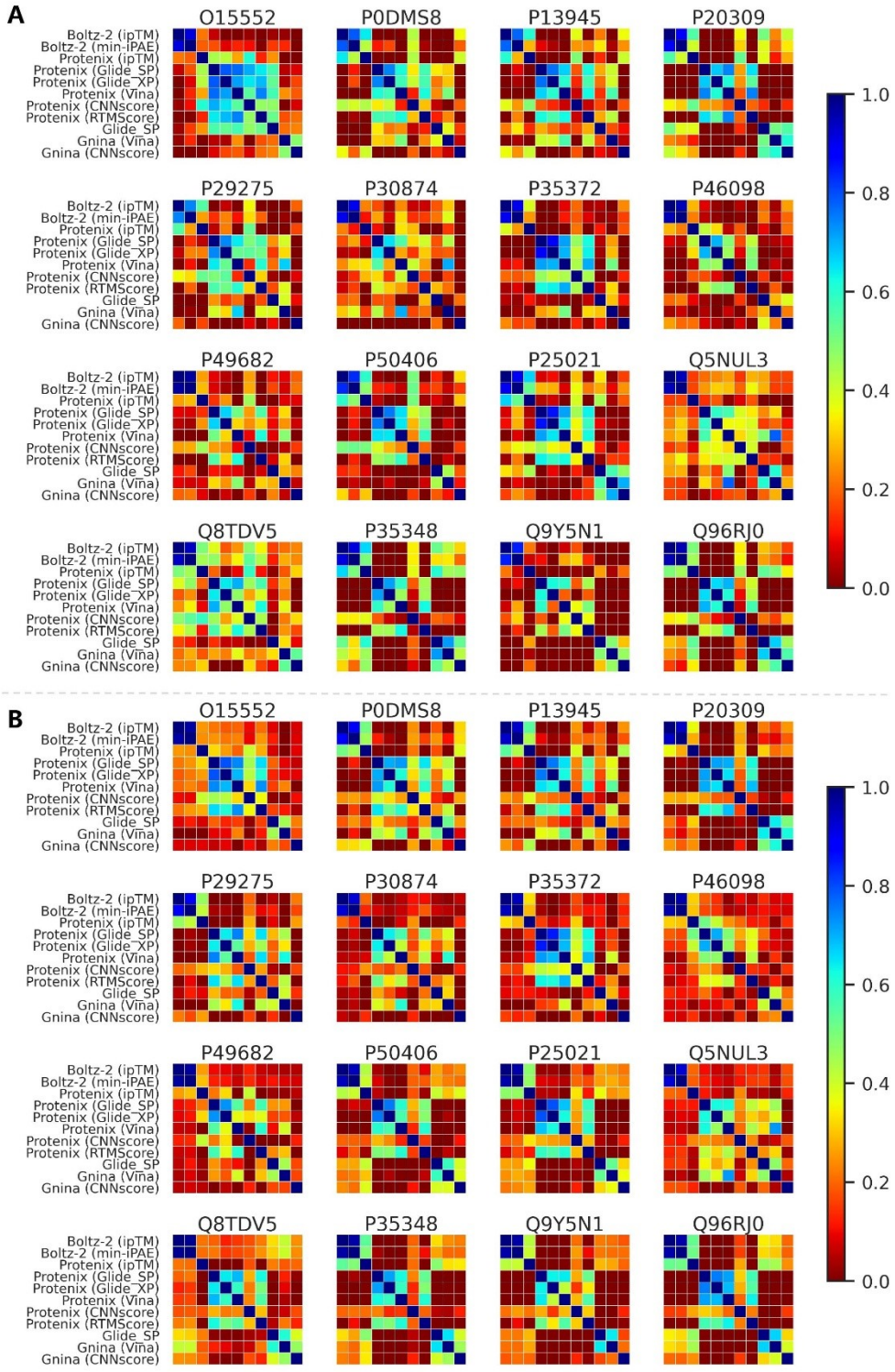

Figure S15. Spearman's rank correlation between different screening tools for (A) active compounds and (B) total compounds across targets in *GPCR*$_{recent}$ dataset.

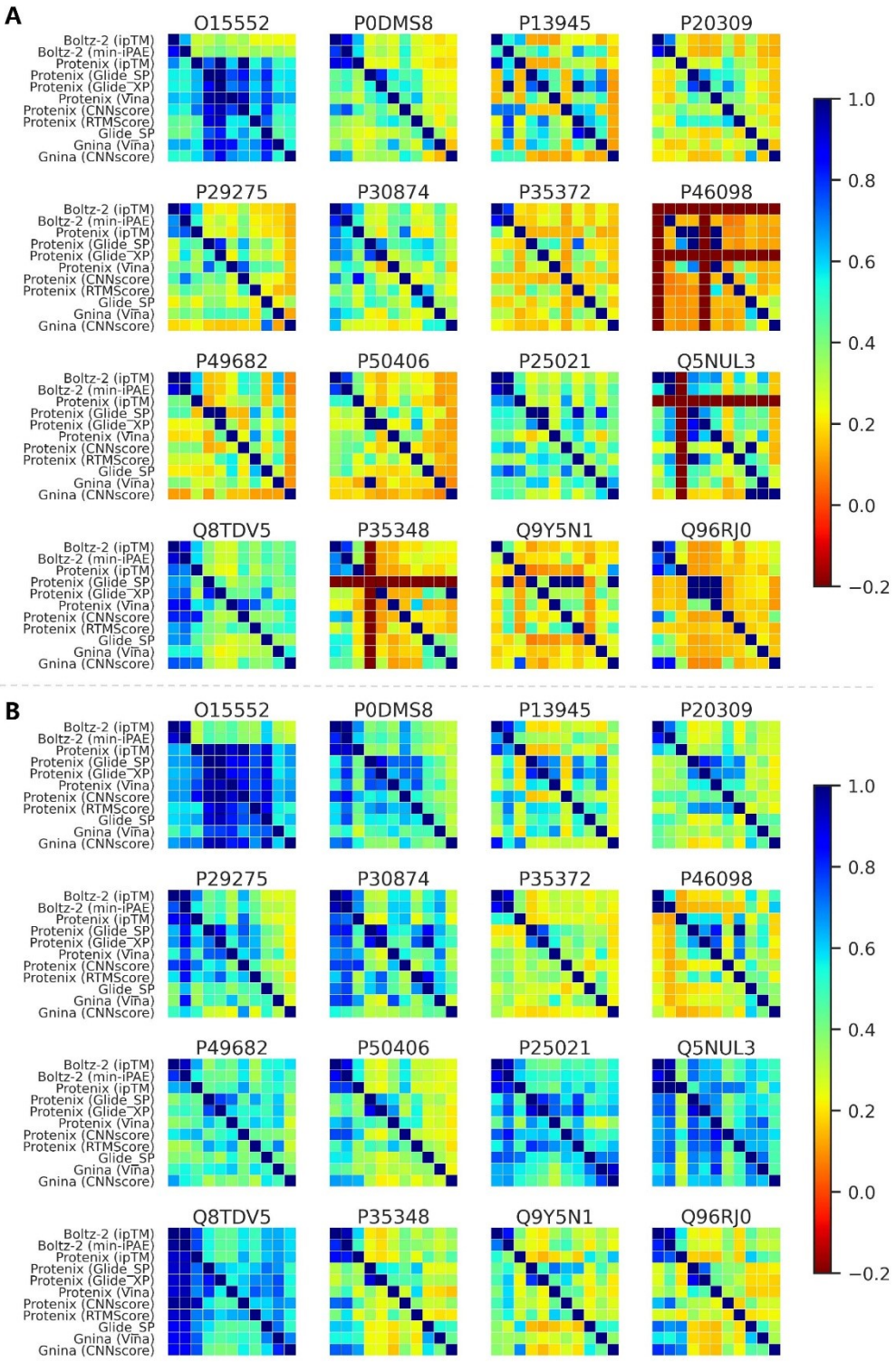

Figure S16. Mean molecular similarities of active compounds to their nearest neighbors among (A) top 100 and (B) top 500 compounds identified by different screening tools across targets in $GPCR_{recent}$ dataset. Similarity is are quantified using the Tanimoto coefficients based on ECFP4 fingerprints. Each row represents the similarities of the actives identified by one method (row label), with values indicating their mean similarity to the nearest neighbors found by each of the other approaches (column label). A value below 0 indicates that no active compound was identified by one of the compared methods at the specified threshold.

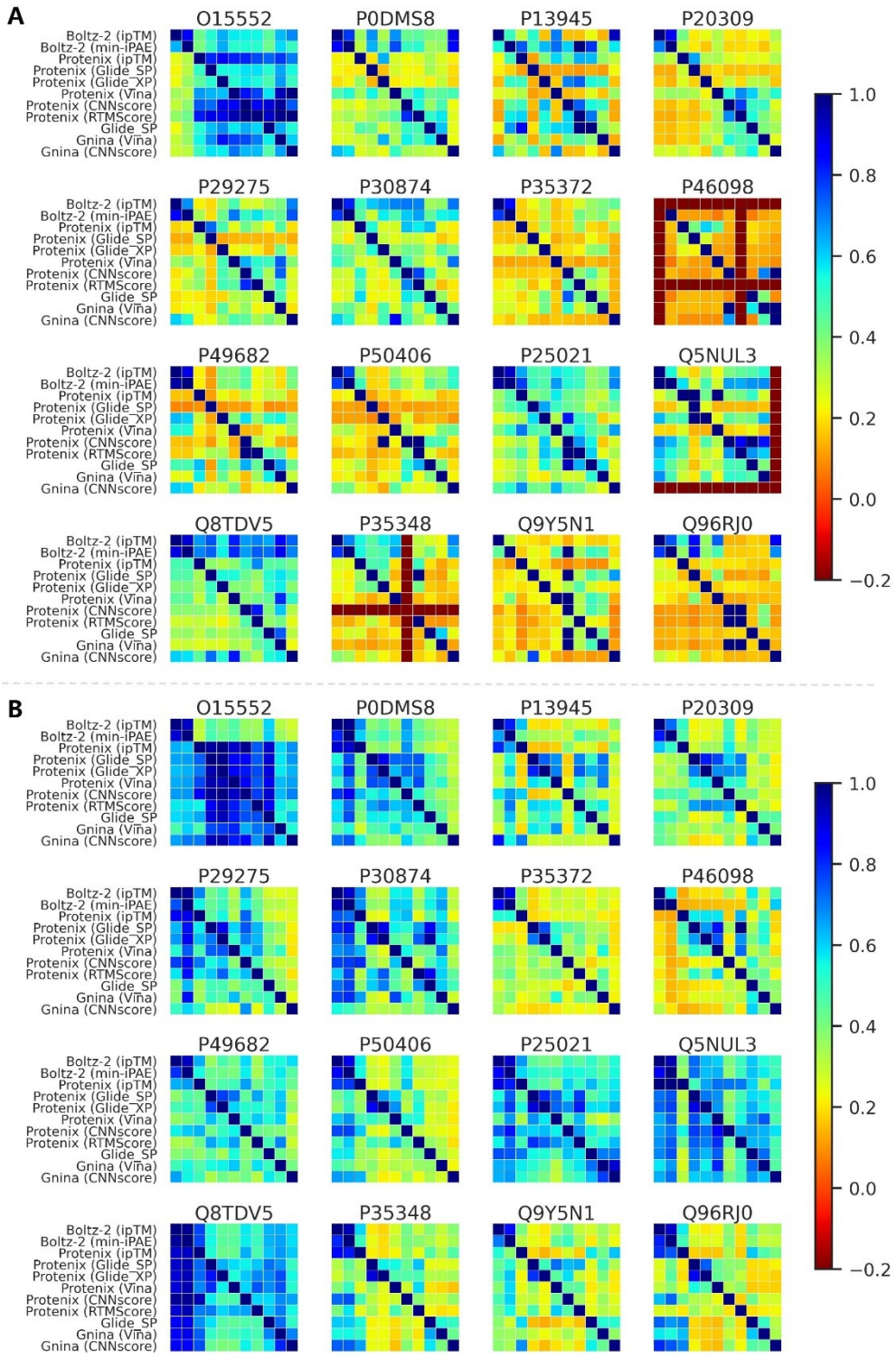

Figure S17. Mean Murcko scaffold similarities of active compounds to their nearest neighbors among (A) top 100 and (B) top 500 compounds identified by different screening tools across targets in *GPCR*$_{recent}$ dataset. Similarity is are quantified using the Tanimoto coefficients based on ECFP4 fingerprints. Each row represents the similarities of the actives identified by one method (row label), with values indicating their mean similarity to the nearest neighbors found by each of the other approaches (column label). A value below 0 indicates that no active compound was identified by one of the compared methods at the specified threshold.