

Supplementary Information

Zhijun Zhang^{1,‡}, Lijun Quan^{1,2,3,‡,*}, Junkai Wang¹, Liangchen Peng¹, Qiufeng Chen¹,
Bei Zhang¹, Lexin Cao¹, Yelu Jiang¹, Geng Li¹, Liangpeng Nie¹, Tingfang Wu^{1,2,3,*},
and Qiang Lyu^{1,2,3,*}

¹School of Computer Science and Technology, Soochow University, Suzhou 215006, China

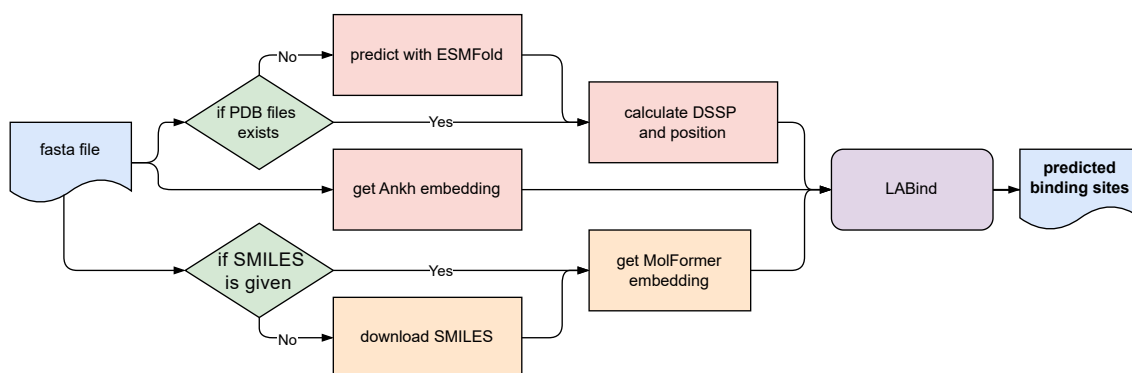
²Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 210000, China

³Jiangsu Key Laboratory of Drug Discovery and Translational Research for Brain Diseases, Soochow University, Suzhou 215006, China

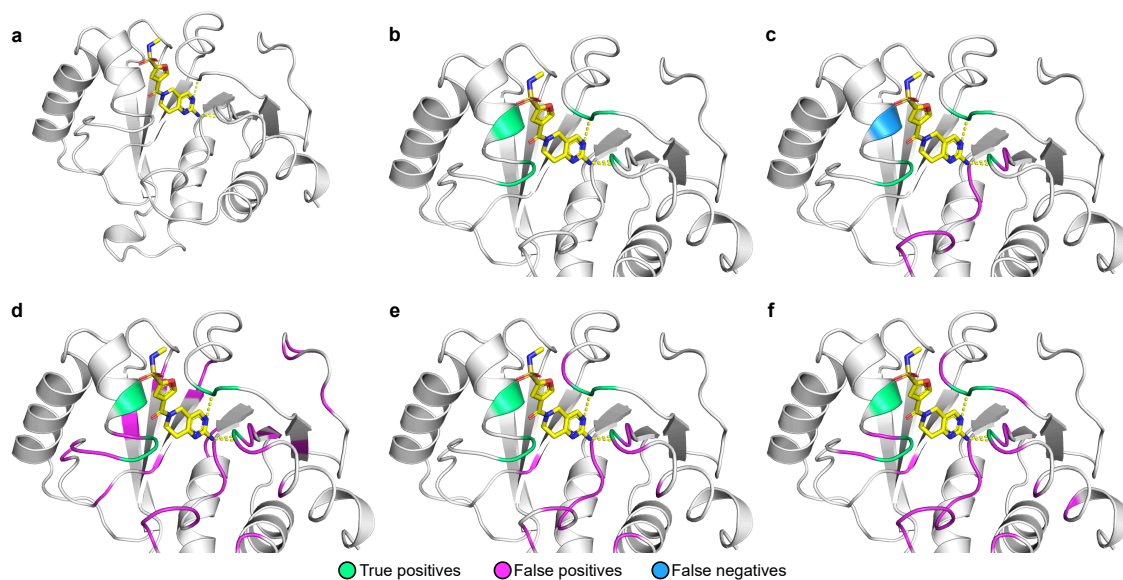
[‡]The authors contribute to the work equally.

*Corresponding Author. E-mail: ljquan@suda.edu.cn, tfwu@suda.edu.cn, and qiang@suda.edu.cn.

1 Supplementary Figures



Supplementary Fig. 1 The program flow for predicting protein-ligand binding sites.



Supplementary Fig. 2 The visualization of all methods on binding sites of the SARS-CoV-2 NSP3 macrodomain. **a** The structure of protein (PDB ID: 5SQQ, chain: A) and ligand (Chemical ID: QU3). **b** The prediction result of LABind. **c** The prediction result of LigBind. **d** The prediction result of P2Rank. **e** The prediction result of DeepSurf. **f** The prediction result of DeepPocket. True positives, false positives, and false negatives represent binding residues correctly predicted, non-binding residues incorrectly predicted as binding, and binding residues incorrectly predicted as non-binding, respectively.

2 Supplementary Tables

Supplementary Table 1 Performance of LABind with other methods for 19 ligands on the DS1 dataset.

Ligand	Method	Rec	Pre	F1	MCC	AUC	AUPR
Ca ²⁺	IonCom	0.257	0.211	0.232	0.220	0.705	0.153
	DELIA	0.151	0.599	0.241	0.296	0.794	0.238
	GraphBind	0.360	0.616	0.455	0.465	0.883	0.424
	LigBind	0.322	0.686	0.439	0.465	0.866	0.420
	LABind-NL	0.488	0.515	0.501	0.494	0.913	0.458
	LABind	0.541	0.825	0.653	0.664	0.947	0.660
Mg ²⁺	IonCom	0.231	0.275	0.251	0.246	0.678	0.142
	DELIA	0.137	0.639	0.225	0.293	0.766	0.200
	GraphBind	0.272	0.385	0.319	0.319	0.799	0.220
	LigBind	0.244	0.525	0.333	0.354	0.845	0.261
	LABind-NL	0.214	0.454	0.291	0.308	0.892	0.230
	LABind	0.350	0.617	0.447	0.461	0.921	0.403
Mn ²⁺	IonCom	0.424	0.217	0.287	0.291	0.818	0.273
	DELIA	0.482	0.657	0.556	0.558	0.893	0.518
	GraphBind	0.464	0.689	0.555	0.561	0.907	0.521
	LigBind	0.580	0.646	0.611	0.608	0.937	0.581
	LABind-NL	0.584	0.760	0.660	0.663	0.963	0.684
	LABind	0.688	0.741	0.713	0.710	0.970	0.743
Zn ²⁺	IonCom	0.854	0.120	0.210	0.299	0.945	0.609
	LigBind	0.670	0.847	0.748	0.750	0.965	0.756
	LABind-NL	0.695	0.811	0.749	0.748	0.982	0.777
	LABind	0.823	0.868	0.845	0.843	0.994	0.901
Fe ³⁺	IonCom	0.629	0.238	0.346	0.377	0.866	0.436
	LigBind	0.670	0.694	0.681	0.678	0.978	0.675
	LABind-NL	0.733	0.753	0.743	0.740	0.992	0.784
	LABind	0.699	0.849	0.766	0.768	0.993	0.827
Fe ²⁺	IonCom	0.341	0.737	0.467	0.498	0.865	0.438
	LigBind	0.610	0.765	0.679	0.680	0.968	0.734
	LABind-NL	0.829	0.703	0.761	0.761	0.995	0.830
	LABind	0.829	0.872	0.850	0.848	0.997	0.885
Cu ²⁺	IonCom	0.469	0.386	0.423	0.416	0.817	0.367
	LigBind	0.498	0.625	0.554	0.552	0.917	0.557
	LABind-NL	0.639	0.688	0.662	0.658	0.976	0.626
	LABind	0.726	0.825	0.773	0.771	0.989	0.822
Na ⁺	IonCom	0.060	0.117	0.079	0.066	0.572	0.057
	LigBind	0.121	0.27	0.167	0.166	0.733	0.115
	LABind-NL	0.302	0.188	0.232	0.212	0.778	0.122
	LABind	0.397	0.409	0.403	0.387	0.833	0.302
K ⁺	IonCom	0.136	0.127	0.132	0.114	0.654	0.055
	LigBind	0.190	0.311	0.236	0.232	0.789	0.181
	LABind-NL	0.497	0.155	0.236	0.253	0.873	0.137

CO ₃ ²⁻	LABind	0.381	0.538	0.446	0.444	0.922	0.427
	IonCom	0.182	0.110	0.137	0.129	0.691	0.133
	LigBind	0.222	0.222	0.222	0.213	0.820	0.100
	LABind-NL	0.414	0.114	0.179	0.201	0.841	0.101
NO ₂ ⁻	LABind	0.303	0.435	0.357	0.357	0.834	0.257
	IonCom	0.14	0.316	0.194	0.196	0.645	0.159
	LigBind	0.326	0.452	0.378	0.369	0.799	0.275
	LABind-NL	0.372	0.160	0.224	0.213	0.699	0.105
SO ₄ ²⁻	LABind	0.140	0.857	0.240	0.340	0.770	0.259
	IonCom	0.261	0.181	0.214	0.197	0.717	0.138
	LigBind	0.352	0.403	0.376	0.364	0.866	0.298
	LABind-NL	0.352	0.285	0.315	0.300	0.857	0.188
PO ₄ ³⁻	LABind	0.478	0.430	0.453	0.441	0.912	0.408
	IonCom	0.241	0.206	0.222	0.207	0.741	0.182
	LigBind	0.454	0.412	0.432	0.422	0.907	0.415
	LABind-NL	0.479	0.385	0.427	0.418	0.915	0.328
ADP	LABind	0.585	0.529	0.556	0.548	0.946	0.516
	LigBind	0.710	0.635	0.670	0.663	0.977	0.700
	LABind-NL	0.656	0.632	0.644	0.635	0.979	0.636
	LABind	0.775	0.667	0.717	0.712	0.987	0.761
AMP	LigBind	0.567	0.465	0.511	0.497	0.945	0.469
	LABind-NL	0.457	0.469	0.463	0.447	0.924	0.379
	LABind	0.662	0.488	0.562	0.553	0.954	0.553
ATP	DELIA	0.496	0.601	0.544	0.535	0.924	0.527
	GraphBind	0.528	0.449	0.485	0.472	0.891	0.445
	LigBind	0.586	0.592	0.589	0.578	0.96	0.608
	LABind-NL	0.631	0.614	0.622	0.612	0.964	0.597
	LABind	0.658	0.694	0.675	0.667	0.976	0.707
GDP	LigBind	0.627	0.652	0.639	0.628	0.967	0.701
	LABind-NL	0.661	0.557	0.605	0.593	0.949	0.604
	LABind	0.661	0.715	0.687	0.678	0.975	0.755
GTP	LigBind	0.520	0.618	0.565	0.554	0.936	0.577
	LABind-NL	0.606	0.606	0.606	0.594	0.952	0.549
	LABind	0.610	0.749	0.672	0.667	0.974	0.721
HEM	DELIA	0.582	0.665	0.62	0.596	0.928	0.648
	GraphBind	0.817	0.537	0.648	0.633	0.966	0.699
	LigBind	0.799	0.699	0.746	0.728	0.979	0.801
	LABind-NL	0.637	0.682	0.659	0.635	0.939	0.676
	LABind	0.788	0.683	0.732	0.713	0.979	0.800
AVG	LigBind	0.477	0.554	0.504	0.500	0.903	0.485
	LABind-NL	0.539	0.502	0.504	0.499	0.915	0.464
	LABind	0.584	0.673	0.608	0.609	0.941	0.616

Note: Boldface indicates the best performance for this ligand test set across all methods. AVG stands for the average value across all ligands.

Supplementary Table 2 AUPR comparison on the DS1 dataset under different training and test settings.

Ligand	Train: Native	Train: ESMFold	Train: Native	Train: ESMFold
	Test: Native	Test: ESMFold	Test: ESMFold	Test: Native
Ca ²⁺	0.660	0.552	0.549	0.573
Mg ²⁺	0.403	0.346	0.344	0.358
Mn ²⁺	0.743	0.704	0.705	0.702
Zn ²⁺	0.901	0.841	0.831	0.815
Fe ³⁺	0.827	0.797	0.811	0.800
Fe ²⁺	0.885	0.853	0.859	0.817
Cu ⁺	0.822	0.791	0.767	0.727
Na ⁺	0.302	0.252	0.250	0.257
K ⁺	0.427	0.302	0.307	0.343
CO ₃ ²⁻	0.257	0.257	0.203	0.267
NO ₂ ⁻	0.259	0.119	0.206	0.129
SO ₄ ²⁻	0.408	0.360	0.320	0.369
PO ₄ ³⁻	0.516	0.503	0.495	0.486
ADP	0.761	0.702	0.71	0.704
AMP	0.553	0.507	0.487	0.505
ATP	0.707	0.673	0.668	0.67
GDP	0.755	0.690	0.685	0.714
GTP	0.721	0.643	0.645	0.624
HEM	0.800	0.773	0.758	0.761

Note: Native indicates the use of native structures, ESMFold indicates the use of structures predicted by ESMFold.

Supplementary Table 3 Performance comparison of LABind with category and LABind.

Ligand	MCC			AUC			AUPR		
	Mol	Ion	All	Mol	Ion	All	Mol	Ion	All
Ca ²⁺	-	0.657	0.664	-	0.947	0.947	-	0.654	0.660
Mg ²⁺	-	0.453	0.461	-	0.919	0.921	-	0.399	0.403
Mn ²⁺	-	0.707	0.710	-	0.967	0.970	-	0.734	0.743
Zn ²⁺	-	0.846	0.843	-	0.994	0.994	-	0.903	0.901
Fe ³⁺	-	0.773	0.768	-	0.992	0.993	-	0.838	0.827
Fe ²⁺	-	0.832	0.848	-	0.997	0.997	-	0.876	0.885
Cu ²⁺	-	0.755	0.771	-	0.989	0.989	-	0.798	0.822
Na ⁺	-	0.365	0.387	-	0.839	0.833	-	0.280	0.302
K ⁺	-	0.445	0.444	-	0.931	0.922	-	0.427	0.427
CO ₃ ²⁻	-	0.298	0.357	-	0.828	0.834	-	0.193	0.257
NO ₂ ⁻	-	0.301	0.340	-	0.762	0.770	-	0.214	0.259
SO ₄ ²⁻	-	0.451	0.441	-	0.914	0.912	-	0.408	0.408
PO ₄ ³⁻	-	0.547	0.548	-	0.943	0.946	-	0.528	0.516
ADP	0.704	-	0.712	0.986	-	0.987	0.750	-	0.761
AMP	0.533	-	0.553	0.949	-	0.954	0.552	-	0.553
ATP	0.680	-	0.667	0.975	-	0.976	0.710	-	0.707
GDP	0.693	-	0.678	0.973	-	0.975	0.761	-	0.755
GTP	0.656	-	0.667	0.966	-	0.974	0.718	-	0.721
HEM	0.715	-	0.713	0.977	-	0.979	0.792	-	0.800

Note: Mol represents training on the small molecule datasets; Ion represents training on the metal and acid radical ions datasets; All represents training on all ligands of the DS1 dataset.

Supplementary Table 4 The Chemical IDs of the unseen ligands.

Chemical ID
02A, 06Q, 07D, 08N, 08T, 0AN, 0NI, 0V9, 118, 1JW, 1O4, 1OI, 3CF, 3JF, 42C, 4J1, 560, 56I, 5IK, 5IV, 5OV, 5X6, 5YG, 5ZK, 5ZY, 64X, 7F1, 7ID, 7OT, 7PZ, 7Q3, 7QB, 7QG, 7R3, 7RI, 7ZX, 80M, 80T, 80Y, 81Q, 84Q, 85I, 85N, 87A, 89K, 89W, 8C2, 8I4, 8KH, 8KZ, 8W7, 94L, 9B0, 9GF, 9SC, 9SL, 9XX, 9Y0, 9YF, 9YX, 9Z9, 9ZN, A1AC4, A1AC5, A1ACI, A1H2K, A1H52, AG7, AGG, AI0, AJP, AQV, AT9, AUF, AX9, AXT, BP4, C0N, C7Z, CI7, CL0, CRT, CZU, D39, DQE, EIQ, EUJ, EZX, F0X, F26, F39, FD0, FHK, FNO, G5F, G7C, G7L, GN9, GV6, GZM, H8G, H9R, HCB, HE9, HEA, HEO, HQG, HUF, HWC, I1F, I1Q, I49, I55, I5F, I70, I7D, I7P, I7R, I7Y, I8A, I8D, I8G, I8M, I9G, I9X, IBW, IEI, IFF, IFJ, IG7, IHT, IIO, IJF, IJR, IO9, IR7, IRM, ISK, ISW, IT4, IUV, IVO, IX7, IXJ, IXU, IY3, IZL, IZR, J40, J6X, J89, JBR, JG0, JHD, JL3, JLQ, JM9, JN9, JQI, JVC, JXL, K6G, K9E, KC2, KDL, KFC, KGD, KHF, KJ0, KJL, KLU, KML, KNR, KP9, KPX, KQ6, KXX, KZB, L2K, L3D, L9I, LFO, LHR, LL0, LMK, LPU, LQ3, LRF, M2L, MF6, MIQ, MJX, MQE, MRC, MV8, MWI, MWO, MWX, N6M, N9U, NB9, ND, NJ9, NJF, NT9, O8I, O8R, OA9, OE9, OI9, OIY, OJC, OLR, OMG, OQZ, OUU, OV3, OYK, OYT, P0M, P0T, P5O, PO8, PSR, PXQ, PXV, Q3R, Q3V, Q3X, Q43, Q4K, Q4L, Q4Z, Q6O, Q7G, Q7L, Q8L, QAK, QBC, QFW, QIY, QLX, QOB, QOD, QOE, QOH, QS2, QSE, QST, R6S, RB8, RBH, RBW, RQ0, RQL, RRX, RVK, RWB, RWR, RXT, SJT, SP2, SQC, SRW, SYQ, T1C, T2W, T2Z, T9K, T9Q, TAH, TBK, TEK, TI0, TK7, TK9, TOO, U35, U4Z, U5L, U5S, U60, U6J, U8O, UB7, UCI, UCJ, UE8, UEB, UFF, UHC, UHL, UI4, UI5, UIS, UIX, UJK, UJQ, UQ7, UQC, US9, USN, UU1, UYY, UZ8, UZZ, V08, V41, V6R, V6Z, V75, V7B, V7N, V8P, V8R, V8U, V99, V9G, V9P, V9X, VBI, VBO, VCW, VEE, VGO, VJI, VJN, VJY, VKA, VKH, VLI, VLS, VND, VNN, VNU, VON, VOT, VOX, VP5, VPE, VPK, VRR, VSZ, VTC, VTL, VTT, VTW, VV0, VWZ, VXF, VXH, VXZ, VYW, VZ6, VZP, VZY, W0A, W0D, W0F, W0G, W0Y, W1D, W1E, W1N, W1P, W28, W2M, W2Q, W3G, W3J, W3M, W3S, W3U, W41, W6H, W6W, W7T, W8K, W9V, W9Y, WA8, WAK, WAS, WBF, WBQ, WBV, WCB, WDC, WEP, WFI, WHL, WKE, WLY, WOY, WQG, WYL, WZY, X0G, X0O, X10, X1J, X2I, X4R, X53, X6M, XB7, XBC, XE0, XFZ, XJ5, XJM, XKS, XP2, XQS, XUH, XXF, XY4, XY6, XYU, XZW, Y33, Y3I, Y6U, Y70, Y72, Y79, Y7L, YDU, YEK, YER, YFR, YG0, YG4, YG5, YG9, YGR, YGW, YK3, YOD, YOJ, YPM, YPS, YRE, YUT, YV2, YV6, YX6, YXB, YXH, YXU, YZK, Z1T, Z2F, Z2Z, Z3H, Z3O, Z41, ZC9, ZCI, ZE0, ZE9, ZG5, ZKM, ZL1, ZNI, ZNS, ZQA, ZQH, ZQV, ZT7, ZTL, ZWY, ZZN, ZXQ

Supplementary Table 5 The performance of LABind on the DS2 test sets using different hyperparameters.

Hidden units	h	ε	N_{thf}	Top_k	Attn Dropout	Dropout	L	lr	ATP AUPR	HEM AUPR	ZN AUPR	CA AUPR	MG AUPR	MN AUPR	AVG AUPR
256	4	0.1	16	30	0.2	0.2	3	$5e-4$	0.696	0.806	0.888	0.675	0.409	0.749	0.704
256	4	0.1	16	30	0.2	0.2	3	$4e-4$	0.695	0.809	0.89	0.674	0.399	0.764	0.705
256	4	0.1	16	30	0.2	0.2	3	$3e-4$	0.695	0.805	0.891	0.67	0.4	0.765	0.704
256	4	0.1	16	30	0.2	0.2	2	$4e-4$	0.659	0.781	0.886	0.651	0.391	0.746	0.686
256	4	0.1	16	30	0.2	0.2	4	$4e-4$	0.704	0.823	0.891	0.682	0.405	0.764	0.712
32	4	0.1	16	30	0.2	0.2	4	$4e-4$	0.654	0.762	0.876	0.643	0.361	0.711	0.668
64	4	0.1	16	30	0.2	0.2	4	$4e-4$	0.673	0.778	0.883	0.65	0.374	0.732	0.682
128	4	0.1	16	30	0.2	0.2	4	$4e-4$	0.68	0.814	0.889	0.668	0.401	0.741	0.699
256	4	0.1	16	10	0.2	0.2	4	$4e-4$	0.622	0.725	0.858	0.629	0.373	0.714	0.653
256	4	0.1	16	20	0.2	0.2	4	$4e-4$	0.697	0.776	0.884	0.673	0.414	0.74	0.697
256	4	0	16	30	0.2	0.2	4	$4e-4$	0.704	0.813	0.896	0.685	0.408	0.75	0.709
256	4	0.2	16	30	0.2	0.2	4	$4e-4$	0.695	0.817	0.893	0.691	0.412	0.763	0.712
256	4	0.3	16	30	0.2	0.2	4	$4e-4$	0.687	0.808	0.893	0.688	0.39	0.74	0.701
256	2	0.2	16	30	0.2	0.2	4	$4e-4$	0.696	0.808	0.892	0.691	0.411	0.754	0.709
256	4	0.1	8	30	0.2	0.2	4	$4e-4$	0.705	0.824	0.891	0.692	0.402	0.771	0.714
256	4	0.1	32	30	0.2	0.2	4	$4e-4$	0.682	0.804	0.888	0.679	0.397	0.759	0.701
256	4	0.1	64	30	0.2	0.2	4	$4e-4$	0.645	0.742	0.878	0.65	0.377	0.735	0.671
256	4	0.1	4	30	0.2	0.2	4	$4e-4$	0.7	0.813	0.892	0.663	0.404	0.768	0.707
256	4	0.1	8	30	0.2	0.1	4	$4e-4$	0.71	0.818	0.892	0.684	0.395	0.761	0.71
256	4	0.1	8	30	0.4	0.3	4	$4e-4$	0.691	0.816	0.891	0.679	0.392	0.759	0.705
256	4	0.1	8	30	0.3	0.3	4	$4e-4$	0.7	0.824	0.892	0.678	0.402	0.767	0.711
256	4	0.1	8	30	0.1	0.3	4	$4e-4$	0.711	0.825	0.897	0.689	0.407	0.765	0.716

Note: Hidden units represent the dimension of the hidden layer, h denotes the number of heads, ε indicates the proportion of noise, and N_{thf} represents the output dimension of the radial basis function. Top_k refers to the number of k -nearest neighbors when constructing the graph. Attn Dropout and Dropout correspond to the attention dropout rate and the dropout rate, respectively. L denotes the number of layers in the graph network, and lr represents the learning rate. The remaining terms refer to the individual and average AUPR on the DS2 dataset.

3 Supplementary Notes

Supplementary Note 1 The calculation methods for all metrics

Consistent with previous studies [1–3], for the evaluation of model performance, we used the following metrics:

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{F1} = 2 \times \frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \quad (3)$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \quad (4)$$

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})d(\text{FPR}) \quad (5)$$

$$\text{AUPR} = \int_0^1 \text{Pre}(\text{Rec})d(\text{Rec}) \quad (6)$$

where true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) represent the number of correctly predicted binding sites, correctly predicted non-binding sites, incorrectly predicted binding sites, and incorrectly predicted non-binding sites, respectively. TPR ($\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$) and FPR ($\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$) represent the true positive rate and false positive rate, respectively. We used AUPR for hyperparameter selection and model performance because AUPR provides higher sensitivity and more information than AUC in imbalanced binary classification tasks.

To holistically evaluate LABind, we employed DCC (distance between the predicted binding site center and the true binding site center) and DCA (distance between the predicted binding site center and the closest ligand atom) to assess its ability in localizing binding sites. Performance was evaluated by calculating the success rate, defined as the proportion of complexes where DCC (DCA) values fell below the predetermined threshold. Consistent with previous studies [4, 5], the threshold is set to 4 Å. Therefore, the success rate is calculated as follows:

$$\text{Success Rate(DCC/DCA)} = \frac{\text{number of predicted binding sites with DCC/DCA} < 4\text{\AA}}{\text{total number of binding sites}} \times 100\%. \quad (7)$$

Supplementary Note 2 Brief introductions to the competitive methods

We compared the multi-ligand-oriented methods P2Rank [4], DeepSurf [5], DeepPocket [6], LMetalSite [2] and GPSite [3], and the single-ligand-oriented methods LigBind [1], IonCom [7], DELIA [8], Graph-Bind [9]. P2Rank, DeepSurf, DeepPocket and LigBind can predict unseen ligands.

P2Rank. P2Rank is based on the random forest algorithm. It utilizes a rapid numerical algorithm to generate a set of regularly spaced points located on the protein’s solvent-accessible surface (SAS). Structural feature descriptors for the SAS points are calculated based on their local chemical neighbors, and a random forest classifier is employed to predict the ligandability scores of the SAS points. It does not rely on homology modeling or ligand information but predicts potential ligand-binding sites based on the atomic structure and physicochemical properties of the protein. To ensure a fair comparison, we attempted to

retrain P2Rank on the DS3 training set. However, we found that P2Rank encountered convergence issues on the DS3 training set. This issue also occurred in [10]. Therefore, we used its standalone program from <https://github.com/rdk/p2rank> for inference and evaluation.

DeepSurf. DeepSurf is an end-to-end deep learning approach that employs surface-based representations, placing numerous 3D voxelized grids on the protein surface to depict the protein structure. It utilizes a convolutional neural network (CNN) model to analyze the protein’s three-dimensional grid, extracting both global and local features, and predicts potential binding sites. Due to the lack of retraining code in the official DeepSurf repository, we could not retrain DeepSurf. We used its standalone program with pre-trained model weights from <https://github.com/stemylonas/DeepSurf> for inference and evaluation.

DeepPocket. DeepPocket executes Fpocket [11] on protein structures to identify potential binding pockets and calculate the centroids of each predicted pocket. Subsequently, it employs a 3D CNN to re-evaluate the predicted pockets. The centroids of the top-ranked pockets, as determined by the CNN scoring, are then subjected to a CNN segmentation model to derive the final pocket structures. We retrained DeepPocket using the default settings from <https://github.com/devalab/DeepPocket>. Since DeepPocket requires the 3D structures of protein-ligand complexes, we downloaded the required PDB files from the RCSB [12] database to complete the process.

LigBind. LigBind is a structure-based predictor of protein-binding sites that enhances the prediction for 1,159 ligands through relation-aware graph networks. By incorporating ligand information, it achieves the functionality of predicting ligands that were not seen during training. It applies domain adaptive neural networks to automatically leverage the diversity and similarity of various ligand-binding patterns. Similar to DeepSurf, LigBind also does not provide the code for training. We used its webserver (<http://www.csbio.sjtu.edu.cn/bioinf/LigBind/index.html>) for evaluation on the DS3 dataset and standalone program with pre-trained model weights from <http://www.csbio.sjtu.edu.cn/bioinf/LigBind/sourcecode.html> for evaluation on the DS2 dataset.

IonCom. IonCom is a structure-based predictor designed to identify binding sites for metal ions and radical anions, integrating IonSeq with COFACTOR [13], TM-SITE, S-SITE, and COACH [14]. IonSeq is trained using sequence-derived features, including Position-Specific Scoring Matrix (PSSM), ligand-specific binding propensities, predicted secondary structure (SS), and predicted relative solvent accessibility (RSA).

DELIA. DELIA is a structure-based predictor of ligand-binding sites, capable of predicting sites for Ca^{2+} , Mn^{2+} , Mg^{2+} , ATP and HEM. It employs the bidirectional long short-term memory network to extract sequence features, including binding propensities from S-SITE, PSSM, and HMM, predicted SS, and predicted RSA. Additionally, it utilizes a CNN to extract feature information from the distance matrices of proteins.

GraphBind. GraphBind is a structure-based predictor for protein-ligand binding sites, representing protein residues as a graph and employing a hierarchical graph neural network to capture the protein residue graph. It utilizes PSSM, HMM, SS and atomic features to represent amino acids and proteins.

LMetalSite. LMetalSite is a sequence-based predictor for metal ion-binding sites, which leverages a pre-trained protein language model to acquire representations of proteins. It employs attention mechanisms and multi-task learning to predict various metal ions and utilizes transformers to extract common characteristics of different ligands.

GPSSite. GPSSite is a sequence-based predictor for multiple ligand-binding sites, which utilizes a pre-trained protein language model to obtain representations of protein sequences. It uses protein structure predictions from the ESMFold [15] structure predictor. By employing multi-task learning, GPSSite predicts various ligands and extracts residue and relational geometric contexts using graph networks.

Supplementary Note 3 The sequence identity between the DS3 test set and other datasets

We used `mmseqs easy-search` [16] to align all the training sets used by P2Rank, DeepSurf, DeepPocket, and LigBind with the DS3 test set. We used the number of protein sequences in the DS3 test set that have a sequence identity exceeding the threshold with sequences in the other training sets to represent the degree of similarity. The sequence identity between the DS3 test set and the training sets of P2Rank, DeepSurf, DeepPocket, and LigBind is shown in [Supplementary Table 6](#). Among all 1,049 proteins in the DS3 test set, there are 481, 44, 163, and 808 protein sequences that share more than 30% sequence identity with the scPDB, CHEN11, JOINED, and LigBind datasets, respectively. Additionally, there are 294, 20, 59, and 483 protein sequences that share more than 80% sequence identity with these datasets, respectively. By removing sequences from the DS3 training and picking sets that share more than 30% sequence identity with those in the DS3 test set, and ensuring that none of the ligands in the DS3 test set appear in the DS3 training set, this approach ensures that the comparison is not biased towards our method LABind.

Supplementary Table 6 The sequence identity between the DS3 test set and other datasets.

Training set	$N_{>30\%}$	$N_{>80\%}$
scPDB	481	294
CHEN11	44	20
JOINED	163	59
LigBind	808	483

Note: $N_{>30\%}$ and $N_{>80\%}$ denote the number of sequences in the DS3 test set with sequence identity greater than 30% and 80%, respectively.

Supplementary Note 4 Mean-shift-based identification of binding site centers and evaluation protocol

We adopted the mean-shift algorithm [17] as the clustering method because it can automatically determine the number of clusters. However, the mean-shift algorithm still requires the specification of a bandwidth parameter. In this study, the bandwidth was set to 12.0. The cluster centers were defined as the average positions of atoms within each cluster. The binding sites were ranked by cluster size from high to low. For P2Rank, DeepSurf, and DeepPocket, we directly used the predicted binding site centers provided by each method. We selected the top- n binding sites to calculate the evaluation metrics, where n represents the number of identical ligands in the corresponding native protein-ligand complex. For example, in identifying the binding site center for a protein with ligand A, we set n to the number of protein-ligand complexes containing ligand A, rather than the total number of all ligands.

Supplementary Note 5 Construction and evaluation settings of the COACH64 dataset

COACH420 was derived from the COACH test set, and consisted of 420 single chain structures containing a mix of drug targets and naturally occurring ligands. COACH420 is a benchmark dataset commonly used

by P2Rank, DeepSurf, and DeepPocket. To avoid data leakage arising from discrepancies among training sets, we applied a redundancy filtering procedure. Following DeepPocket, we removed complexes from the COACH420 that share more than 50% sequence identity with proteins in the DS3 training set, or have both greater than 0.9 ligand similarity and more than 30% sequence identity. Here, ligand similarity was calculated using the Tanimoto coefficient between molecular fingerprints of the two ligands. After filtering, 59 protein chains were retained, containing a total of 64 binding sites. To distinguish it from COACH420, we named the dataset COACH64. To compare performance across methods, we used the pre-trained models of P2Rank, DeepSurf, and DeepPocket for binding site center prediction on COACH64. We used LABind trained on DS3 to predict binding sites on COACH64.

Supplementary Note 6 The experimental details and procedures of molecular docking

For the 285 protein-ligand complexes in the CASF-2016 core set [18], we first performed a preprocessing step. Initially, all protein chains were separated. LABind predictions were then performed on individual chains. Complexes where no binding site was predicted for any chain were excluded. Ultimately, 16 proteins were excluded, retaining 269 protein-ligand complexes for analysis. To ensure that Smina [19] can correctly perform molecular docking, we use the processing files it provides to process the protein and ligand. The command lines executed are as follows:

```
prepare_receptor4.py -i receptor.pdb -o receptor.pdbqt -A hydrogens -U nphs
prepare_ligand4.py -i ligand.mol2 -o ligand.pdbqt -A hydrogens -U nphs
```

When not utilizing predicted binding sites, docking was performed with the grid box centered on the geometric centroid of the protein and dimensions fully encompassing the entire structure. The binding sites predicted by LABind were used to reduce the size of the search box and determine the center position of the search box. The center of the search box is set at the geometric center of all the atoms of the binding residues, with the size of the box determined by the longest distance between atoms in each dimension plus 4 Å. Docking using the cognate ligand was implemented via `smina autobox_ligand`. To evaluate prediction outcomes, we computed the RMSD between the native pose and docked poses using the Hungarian algorithm (implemented via the `spyrmsd` package [20]).

Supplementary Note 7 The details of the DS1 dataset

We selected 19 ligands from the LigBind ligand-specific dataset for performance comparison, which include Ca^{2+} , Mg^{2+} , Mn^{2+} , Zn^{2+} , Fe^{3+} , Fe^{2+} , Cu^{2+} , Na^+ , K^+ , CO_3^{2-} , NO_2^- , SO_4^{2-} , PO_4^{3-} , ADP, AMP, ATP, GDP, GTP and HEM. [Supplementary Table 7](#) presents statistical information for 19 ligands. For the proteins, there are a total of 10,341 non-redundant proteins and 11,121 protein-ligand binding samples. Regarding the binding sites, there are a total of 72,029 binding sites and 3,734,243 non-binding sites, with the non-binding sites being 51.844 times the number of binding sites. A residue is considered to be a binding site if the shortest atomic distance to the metal ion is less than 0.5 Å plus the combined Van der Waal's radii of the two closest atoms.

Supplementary Table 7 Statistics of the nineteen ligand benchmark datasets in the DS1 dataset.

Ligand	Training dataset				Test dataset			
	N_s	N_{pos}	N_{neg}	N_{neg}/N_{pos}	N_s	N_{pos}	N_{neg}	N_{neg}/N_{pos}
Ca ²⁺	1421	7688	453464	58.98	533	3365	211318	62.80
Mg ²⁺	1701	6029	562714	93.33	665	2490	288725	115.95
Mn ²⁺	513	2341	173359	74.05	167	778	63552	81.69
Zn ²⁺	1770	8389	459705	54.80	615	3150	212711	67.53
Fe ³⁺	237	1124	72650	64.64	88	345	30318	87.88
Fe ²⁺	108	453	34930	77.11	31	123	9665	78.58
Cu ²⁺	129	519	38161	73.53	54	241	16348	67.83
Na ⁺	103	620	32904	53.07	23	199	7418	37.28
K ⁺	64	639	22508	35.22	17	147	7294	49.62
CO ₃ ²⁻	69	330	23328	70.69	21	99	8585	86.72
NO ₂ ⁻	19	83	7053	84.98	7	43	1590	36.98
SO ₄ ²⁻	370	2541	108050	42.52	31	230	10582	46.01
PO ₄ ³⁻	393	2447	122409	50.02	73	482	26242	54.44
ADP	399	5438	159487	29.33	158	1903	78956	41.49
AMP	235	2588	76200	29.44	67	786	25887	32.94
ATP	336	4524	130740	28.90	139	1943	72042	37.08
GDP	127	1558	45987	29.52	40	496	15702	31.66
GTP	83	1052	34250	32.56	34	454	14665	32.30
HEM	203	4733	51862	10.96	78	1659	22882	13.79

Note: N_s represents the number of proteins; N_{pos} represents the number of positive residues (binding site); N_{neg} represents the number of negative residues (non-binding site); N_{neg}/N_{pos} represents the ratio of negative residues to positive residues.

Supplementary Note 8 The details of the DS2 dataset

In the DS2 dataset, the metal ions datasets are derived from LMetalSite. Metal ions datasets are selected with resolutions of ≤ 3.0 Å and lengths of 50-1000. ATP and HEM datasets are collected the corresponding binding proteins with resolutions of ≤ 3.0 Å and lengths of 50-1500 from BioLiP [21] released on 29 March 2023. Redundancy was eliminated from each dataset by excluding proteins with sequence identities $>25\%$ over 30% alignment coverage. Molecule (ATP and HEM) datasets and metal ions (Zn²⁺, Ca²⁺, Mg²⁺, Mn²⁺) datasets are respectively split into training and testing sets based on release date, with the cutoffs being 1 January 2021, for molecule datasets and 1 January 2020, for metal ions datasets. There are a total of 6,609 proteins, 39,571 binding residues, and 2,161,755 non-binding residues.

Supplementary Table 8 Statistics of the six ligand benchmark datasets in the DS2 dataset.

Ligand	Training dataset				Test dataset			
	N_s	N_{pos}	N_{neg}	N_{neg}/N_{pos}	N_s	N_{pos}	N_{neg}	N_{neg}/N_{pos}
Ca ²⁺	1554	8442	495704	58.72	183	1034	65820	63.66
Mg ²⁺	1729	6317	569415	90.14	235	893	87913	98.45
Mn ²⁺	547	2556	179143	70.09	57	225	20194	89.75
Zn ²⁺	1646	7727	467128	60.45	211	1039	54981	52.92
ATP	347	5115	125540	24.54	79	1231	38228	31.05
HEM	176	4022	43041	10.70	48	970	14648	15.10

Note: N_s represents the number of proteins; N_{pos} represents the number of positive residues (binding site); N_{neg} represents the number of negative residues (non-binding site); N_{neg}/N_{pos} represents the ratio of negative residues to positive residues.

Supplementary Note 9 Description of issues with the LigBind web server

Due to the inability to parse the SMILES strings of the ligands, the prediction method LigBind-G without ligand information was utilized for 29 binding samples of 12 ligands, including QIY, CL0, 08T, K6G, OQZ, HEA, USN, V9G, OE9, ISW, F0X and 07D. Some protein structures contained atoms from other chains, which prevented LigBind from parsing them, including 8tfpA and 8tfpH from 84K, 7qxeHHH from G7C, 7t0fH and 7t0fL from K71, 7u62H and 7u62L from MOI, 7u63A and 7u63B from OOX, 6yhjH from OQZ and 8y33A from V8U. Thus, these proteins were excluded from the comparison when using LigBind. As shown in [Supplementary Table 9](#), LigBind removed some residues in certain proteins, and these residues were not considered when calculating the metrics. Given the scarcity of problematic protein-ligand prediction outcomes, we deem that it does not impact the overall predictive results of LigBind on the DS3 dataset.

Supplementary Table 9 The number of residues removed by LigBind on the DS3 dataset.

Ligand	Protein chain	N_d
OV9	7xs6A	1
AJP	7mhyA	2
AQV	8hk7A	1
I7Y	7t3nA	1
JXL	8pf8B	1
USN	7b1sD	1
VGO	8cqiB	1
VWZ	8oqsB	1
VXZ	8oqpB	1
ZQV	7myjC	1

Note: N_d represents the number of deleted residues. In protein chains, the first four letters represent the RCSB ID, followed by the chain identifier.

Supplementary Note 10 The details of the DS3 dataset.

We initially divided the BioLiP2 [22] database into several ligand files based on the ligand names. Only protein chains with resolutions of ≤ 3.0 Å and lengths of < 1500 were kept. The redundancy reduction was performed based on sequence identity using `mmseqs easy-cluster` in `MMseqs2`.

```
mmseqs easy-cluster ligand.fasta clusterRes tmp --min-seq-id 0.3 -c 0.3
```

The aforementioned command was executed with sharing sequence identity <30% over 30% alignment coverage to perform redundancy removal. We utilized the DS3 test set as the independent test set. The Chemical IDs of the 474 ligands in the DS3 dataset are presented in [Supplementary Table 4](#) as follows.

Supplementary Note 11 The model algorithm

Here, we will provide a detailed description of the model algorithm. We first perform dimensionality reduction on the embeddings from the pre-trained model, then combine them with graph representations based on spatial information to form a protein fusion graph, which is input into the attention-based learning module, and finally output residue scores through the MLP classifier. The algorithm corresponding to the architecture is shown in [Supplementary Algorithm 1](#). In the architecture, arguments add noise to the input using a standard normal distribution ($X \sim N(0, 1)$) to achieve the purpose of data augmentation. MLP_l , MLP_e , and MLP_o represent the MLPs for the ligand, edge, and output, respectively. They all use SiLU [23] as the activation function. Additionally, we employed LayerNorm and Dropout to accelerate convergence and prevent overfitting.

Supplementary Algorithm 1 The architecture of the LABind.

Input: *protein, ligand* and *position*;

Output: *binding sites*;

```
1:  $x_p \leftarrow \text{Ankh}(\text{protein})$ 
2:  $x_l \leftarrow \text{MolFormer}(\text{ligand})$ 
3:  $x_p, x_l, x_s \leftarrow \text{arguments}(x_p, x_l, \text{position})$ 
4:  $x_p \leftarrow \text{MLP}_{\text{prot}}(x_p)$ 
5:  $x_l \leftarrow \text{MLP}_{\text{lig}}(x_l)$ 
6:  $z_n, x_e \leftarrow \text{GraphConverterModule}(x_s)$ 
7:  $x_n \leftarrow \text{FuseMLP}(\text{concat}(x_p, z_n))$ 
8: for  $i$  from 1 to 4 do
9:    $x_n \leftarrow x_n + \text{NeighborAttention}(x_n, x_e)$ 
10:   $x_n \leftarrow x_n + \text{CrossAttention}(x_n, x_l)$ 
11:   $x_n \leftarrow x_n + \text{FFN}(x_n)$ 
12:   $x_l \leftarrow x_l + \text{MLP}_l(\text{concat}(\text{Mean}(x_n), \text{Max}(x_n), x_l))$ 
13:   $x_e \leftarrow x_e + \text{MLP}_e(\text{GatherEdges}(x_n, x_e))$ 
14: end for
15:  $\text{binding sites} \leftarrow \text{MLP}_o(x_n)$ 
16: return binding sites
```

Supplementary Note 12 Version of other software

In addition to using Python and PyTorch, we also employ a variety of other tools to assist in our research. We utilize Biopython v1.83 [24] to parse protein structures. MSMS v2.6.1 [25] and DSSP v2.0.4 [26] are respectively used to calculate the protein surface and secondary structure. The tool for calculating TM-score employs US-align (version 20240705) [27], Clustering and redundancy removal are performed using MMseq2 (version d4841a8efad066e9758b6626cc64c5ef5ee53055). Various evaluation metrics are calculated using Scikit-learn v1.3.2 [28]. We use Smina (version 2020.12.10) [19] for the validation of molecular docking. We use PyMOL v3.0.0 [29] to generate the final binding site figures and some visualization images.

Supplementary References

- [1] Ying Xia, Xiaoyong Pan, and Hong-Bin Shen. LigBind: Identifying Binding Residues for Over 1000 Ligands with Relation-Aware Graph Neural Networks. *Journal of Molecular Biology*, 435(13):168091, 2023.
- [2] Qianmu Yuan, Sheng Chen, Yu Wang, Huiying Zhao, and Yuedong Yang. Alignment-free metal ion-binding site prediction from protein sequence through pretrained language model and multi-task learning. *Briefings in Bioinformatics*, 23(6):bbac444, 2022.
- [3] Qianmu Yuan, Chong Tian, and Yuedong Yang. Genome-scale annotation of protein binding sites via language model and geometric deep learning. *eLife*, 13:RP93695, 2024.
- [4] Radoslav Krivák and David Hoksza. P2Rank: Machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):39, 2018.
- [5] Stelios K Mylonas, Apostolos Axenopoulos, and Petros Daras. DeepSurf: A surface-based deep learning approach for the prediction of ligand binding sites on proteins. *Bioinformatics*, 37(12):1681–1690, 2021.
- [6] Rishal Aggarwal, Akash Gupta, Vineeth Chelur, C. V. Jawahar, and U. Deva Priyakumar. DeepPocket: Ligand Binding Site Detection and Segmentation using 3D Convolutional Neural Networks. *Journal of Chemical Information and Modeling*, 62(21):5069–5079, 2022.
- [7] Xiuzhen Hu, Qiwen Dong, Jianyi Yang, and Yang Zhang. Recognizing metal and acid radical ion-binding sites by integrating *ab initio* modeling with template-based transferals. *Bioinformatics*, 32(21):3260–3269, 2016.
- [8] Chun-Qiu Xia, Xiaoyong Pan, and Hong-Bin Shen. Protein–ligand binding residue prediction enhancement through hybrid deep heterogeneous learning of sequence and structure data. *Bioinformatics*, 36(10):3018–3027, 2020.
- [9] Ying Xia, Chun-Qiu Xia, Xiaoyong Pan, and Hong-Bin Shen. GraphBind: Protein structural context embedded rules learned by hierarchical graph neural networks for recognizing nucleic-acid-binding residues. *Nucleic Acids Research*, 49(9):e51–e51, 2021.
- [10] Yang Zhang, Wenbing Huang, Zhewei Wei, Ye Yuan, and Chongxuan Li. Equipocket: an e(3)-equivariant geometric graph neural network for ligand binding site prediction, 2024.
- [11] Vincent Le Guilloux, Peter Schmidtke, and Pierre Tuffery. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10(1):168, 2009.
- [12] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, January 2000.

- [13] Chengxin Zhang, Peter L. Freddolino, and Yang Zhang. COFACTOR: Improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic Acids Research*, 45(W1):W291–W299, 2017.
- [14] Jianyi Yang, Ambrish Roy, and Yang Zhang. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, 29(20):2588–2595, 2013.
- [15] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan Dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.
- [16] Martin Steinegger and Johannes Söding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, 2017.
- [17] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, May 2002.
- [18] Minyi Su, Qifan Yang, Yu Du, Guoqin Feng, Zhihai Liu, Yan Li, and Renxiao Wang. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *Journal of Chemical Information and Modeling*, 59(2):895–913, 2019.
- [19] David Ryan Koes, Matthew P. Baumgartner, and Carlos J. Camacho. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *Journal of Chemical Information and Modeling*, 53(8):1893–1904, August 2013.
- [20] Rocco Meli and Philip C. Biggin. Spyrmsd: Symmetry-corrected RMSD calculations in Python. *J Cheminform*, 12(1):49, December 2020.
- [21] Jianyi Yang, Ambrish Roy, and Yang Zhang. BioLiP: A semi-manually curated database for biologically relevant ligand–protein interactions. *Nucleic Acids Research*, 41(D1):D1096–D1103, 2012.
- [22] Chengxin Zhang, Xi Zhang, Peter L Freddolino, and Yang Zhang. BioLiP2: An updated structure database for biologically relevant ligand–protein interactions. *Nucleic Acids Research*, 52(D1):D404–D412, January 2024.
- [23] Stefan Elfving, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 107:3–11, 2018.
- [24] Peter J. A. Cock, Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J. L. De Hoon. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009.
- [25] Michel F. Sanner, Arthur J. Olson, and Jean-Claude Spehner. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–320, 1996.

- [26] R. P. Joosten, T. A. H. Te Beek, E. Krieger, M. L. Hekkelman, R. W. W. Hooft, R. Schneider, C. Sander, and G. Vriend. A series of PDB related databases for everyday needs. *Nucleic Acids Research*, 39(Database):D411–D419, 2011.
- [27] Chengxin Zhang, Morgan Shine, Anna Marie Pyle, and Yang Zhang. US-align: Universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods*, 19(9):1109–1115, 2022.
- [28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [29] LLC Schrödinger and Warren DeLano. Pymol. <http://www.pymol.org/pymol>.