

Supplementary Information for Protein-peptide binding site detection using 3D convolutional neural networks

Igor Kozlovskii and Petr Popov*

*iMolecule, Center for Computational and Data-Intensive Science and Engineering,
Skolkovo Institute of Science and Technology, Moscow, Russia, 121205*

E-mail: p.popov@skoltech.ru

Table S1: Similarity between proteins in TR1154 (1154 complexes) and TS125 (125 complexes) in terms of the sequence identity (normalized by the alignment length). Column 'Threshold' corresponds to the sequence identity threshold; 'Pairs' – the total number of similar pairs; 'Num TR1154' – number of complexes from TR1154 that have similar proteins in TS125; 'Num TS125' – number of complexes from TS125 that have similar proteins in TR1154.

Threshold	Pairs	Num TR1154	Num TS125
0.30	63	50	28
0.40	33	29	22
0.50	24	21	16
0.60	18	16	12
0.70	10	10	8
0.80	3	3	3

Table S2: Similarity between proteins in $BN_{Pp}956$ (956 complexes) and TS125 (125 complexes) in terms of the sequence identity (normalized by the alignment length). Column 'Threshold' corresponds to the sequence identity threshold; 'Pairs' – the total number of similar pairs; 'Num $BN_{Pp}956$ ' – number of complexes from $BN_{Pp}956$ that have similar proteins in TS125; 'Num TS125' – number of complexes from TS125 that have similar proteins in $BN_{Pp}956$.

Threshold	Pairs	Num $BN_{Pp}956$	Num TS125
0.30	16	15	15
0.40	6	5	6
0.50	3	2	3
0.60	0	0	0
0.70	0	0	0
0.80	0	0	0

Table S3: Sizes of train and validation splits used in the cross-validation. $BiteNet_{BN_{Pp}956}$ and $BiteNet_{TR1154}$ correspond to the filtered sets used to train the protein-small molecule binding site prediction models for the fine-tuning.

Index	$BiteNet_{TR1154}$		TR1154		$BiteNet_{BN_{Pp}956}$		$BN_{Pp}956$	
	Train	Val	Train	Val	Train	Val	Train	Val
0	4818	721	924	230	4892	710	825	131
1	4782	707	923	231	4812	718	818	138
2	4860	709	923	231	4902	723	564	392
3	4907	684	923	231	4875	717	789	167
4	4713	716	923	231	4952	715	828	128
Final	5008	724	1154		5008	724	956	

Table S4: Number of amino acid residues classified as binding and non-binding in TS125 with respect to the distance threshold. The first line indicates the number of binding and non-binding residues from previous studies. Column 'Overlap' shows the fraction of identical labels between the labels defined by a distance threshold and the labels from the previous studies.

Distance (Å)	Binding	Non-binding	Fraction binding	Fraction non-binding	Overlap
	1716	29154	0.056	0.944	
3.0	593	30277	0.019	0.981	0.964
3.1	704	30166	0.023	0.977	0.967
3.2	813	30057	0.026	0.974	0.971
3.3	937	29933	0.030	0.970	0.975
3.4	1072	29798	0.035	0.965	0.979
3.5	1218	29652	0.039	0.961	0.984
3.6	1403	29467	0.045	0.955	0.990
3.7	1576	29294	0.051	0.949	0.994
3.8	1719	29151	0.056	0.944	0.996
3.9	1827	29043	0.059	0.941	0.996
4.0	1921	28949	0.062	0.938	0.993

Table S5: The precision, recall, and AP_{hotspot} metrics on cross-validation. Asterisk corresponds to the models trained without fine tuning.

Dataset	Index	Train				Validation				
		Size	Precision	Recall	AP_{hotspot}	Size	Precision	Recall	AP_{hotspot}	
TR1154*	0	924	0.097	0.997	0.461	230	0.088	0.770	0.337	
	1	923	0.103	0.886	0.437	231	0.097	0.781	0.266	
	2	923	0.104	0.895	0.477	231	0.092	0.765	0.268	
	3	923	0.106	0.891	0.469	231	0.074	0.789	0.288	
	4	923	0.104	0.868	0.417	231	0.088	0.804	0.291	
	Average			0.103	0.907	0.452		0.088	0.782	0.290
	Final	1154	0.096	0.902	0.474					
TR1154	0	924	0.137	0.884	0.518	230	0.122	0.743	0.351	
	1	923	0.114	0.881	0.506	231	0.105	0.769	0.316	
	2	923	0.112	0.909	0.534	231	0.093	0.780	0.303	
	3	923	0.104	0.894	0.500	231	0.078	0.827	0.342	
	4	923	0.108	0.844	0.414	231	0.096	0.789	0.317	
	Average			0.115	0.882	0.494		0.099	0.782	0.326
	Final	1154	0.114	0.903	0.542					
BN_{P_p956} *	0	825	0.158	0.899	0.531	131	0.116	0.797	0.362	
	1	818	0.147	0.909	0.549	138	0.106	0.786	0.298	
	2	564	0.095	0.907	0.463	392	0.098	0.863	0.384	
	3	789	0.124	0.913	0.529	167	0.090	0.816	0.304	
	4	828	0.126	0.912	0.528	128	0.112	0.783	0.339	
	Average			0.130	0.908	0.520		0.104	0.809	0.337
	Final	956	0.137	0.910	0.543					
BN_{P_p956}	0	825	0.158	0.910	0.564	131	0.113	0.844	0.385	
	1	818	0.116	0.916	0.540	138	0.088	0.837	0.319	
	2	564	0.112	0.891	0.483	392	0.107	0.843	0.406	
	3	789	0.155	0.917	0.574	167	0.110	0.814	0.388	
	4	828	0.170	0.899	0.554	128	0.137	0.755	0.374	
	Average			0.142	0.907	0.543		0.111	0.819	0.374
	Final	956	0.147	0.915	0.567					

Table S6: The AP_{residues} , ROC AUC, and MCC metrics of BiteNet $_{Pp}$ and the state-of-the-art methods on the TS125 benchmark. The training set is in brackets, and asterisk corresponds to the curated version of a set.

Method	AP_{residues}	AUC	MCC
BiteNet$_{Pp}$ (TR1154)*	0.53	0.91	0.49
BiteNet$_{Pp}$ (BN $_{Pp}$ 956)*	0.52	0.91	0.48
BiteNet$_{Pp}$ (TR1154)	0.48	0.89	0.46
BiteNet$_{Pp}$ (BN $_{Pp}$ 956)	0.44	0.88	0.44
BiteNet ¹	0.25	0.79	0.24
PepBind ²		0.79	0.37
P2Rank-Pept ³		0.85	0.35
PepNN-Struct ⁴		0.84	0.32
MTDSite ⁵		0.76	0.30
SPRINT-Str ⁶		0.78	0.29
PeptiMap ⁷		0.63	0.27
Multi-VORFFIP ⁸		0.78	0.21
SPRINT-Seq ⁹		0.68	0.20
PepSite ¹⁰		0.61	0.20
Visual ¹¹		0.73	0.17

Table S7: The 'hot spot'-based and residue-based performance metrics for the cross-validation sets and the test benchmarks. The "All" scores were calculated for all binding and non-binding residues in the sets, and the "Mean" scores correspond to the scores calculated for each structure separately and averaged by the total number of structures. Asterisk corresponds to the models trained without fine tuning. BiteNet is the model trained on small molecule ligand binding sites;¹ it was used to fine tune BiteNet_{pp}. cv0-4 state for the sets used in the five-fold cross-validation. Note that the 'hot spot'-based scores for TS125_{curated} are not listed, as they are affected by the added protein chains.

Model	Set	Hotspot			All								Mean							
		PREC	REC	AP	AP	AUC	MCC	ACC	PREC	REC	SPEC	F1	AP	AUC	MCC	ACC	PREC	REC	SPEC	F1
TR1154* cv0	cv0	0.038	0.939	0.360	0.426	0.867	0.412	0.943	0.485	0.401	0.975	0.439	0.465	0.808	0.313	0.907	0.365	0.374	0.951	0.341
TR1154* cv1	cv1	0.044	0.920	0.312	0.395	0.870	0.390	0.940	0.457	0.389	0.972	0.421	0.451	0.820	0.310	0.914	0.361	0.362	0.957	0.336
TR1154* cv2	cv2	0.039	0.944	0.314	0.388	0.858	0.369	0.942	0.504	0.314	0.981	0.387	0.433	0.794	0.263	0.908	0.351	0.284	0.965	0.283
TR1154* cv3	cv3	0.033	0.956	0.322	0.360	0.853	0.362	0.945	0.436	0.350	0.976	0.388	0.397	0.782	0.258	0.912	0.311	0.309	0.957	0.286
TR1154* cv4	cv4	0.035	0.942	0.332	0.379	0.868	0.382	0.944	0.447	0.378	0.975	0.409	0.434	0.809	0.291	0.909	0.343	0.351	0.951	0.317
TR1154*	TS125	0.041	0.958	0.443	0.471	0.882	0.452	0.948	0.544	0.421	0.979	0.475	0.506	0.841	0.362	0.916	0.422	0.405	0.958	0.385
TR1154*	TS125 _{refined}	0.041	0.950	0.408	0.474	0.881	0.450	0.948	0.537	0.424	0.978	0.474	0.503	0.836	0.366	0.915	0.428	0.409	0.957	0.391
TR1154*	TS125 _{curated}				0.493	0.894	0.459	0.952	0.598	0.389	0.985	0.471	0.548	0.869	0.368	0.926	0.464	0.373	0.970	0.378
TR1154 cv0	cv0	0.051	0.936	0.404	0.435	0.871	0.421	0.941	0.470	0.436	0.971	0.452	0.476	0.807	0.343	0.904	0.399	0.415	0.943	0.375
TR1154 cv1	cv1	0.049	0.963	0.404	0.439	0.886	0.419	0.940	0.468	0.434	0.971	0.450	0.478	0.836	0.344	0.915	0.393	0.403	0.952	0.369
TR1154 cv2	cv2	0.042	0.932	0.368	0.417	0.867	0.398	0.940	0.482	0.383	0.974	0.427	0.447	0.794	0.298	0.905	0.350	0.353	0.952	0.330
TR1154 cv3	cv3	0.036	0.979	0.403	0.390	0.880	0.388	0.945	0.446	0.389	0.975	0.415	0.429	0.807	0.280	0.911	0.319	0.346	0.951	0.307
TR1154 cv4	cv4	0.036	0.955	0.357	0.426	0.885	0.422	0.946	0.467	0.435	0.973	0.450	0.473	0.822	0.331	0.911	0.367	0.407	0.949	0.354
TR1154	TS125	0.046	0.962	0.480	0.490	0.893	0.459	0.949	0.548	0.431	0.979	0.482	0.530	0.857	0.371	0.917	0.417	0.411	0.959	0.397
TR1154	TS125 _{refined}	0.046	0.958	0.458	0.482	0.891	0.457	0.948	0.540	0.433	0.978	0.481	0.523	0.853	0.378	0.918	0.432	0.417	0.960	0.403
TR1154	TS125 _{curated}				0.525	0.909	0.487	0.953	0.604	0.431	0.984	0.503	0.563	0.882	0.401	0.929	0.473	0.413	0.971	0.418
BN _{pp} 956* cv0	cv0	0.043	0.958	0.425	0.475	0.892	0.448	0.952	0.478	0.469	0.975	0.473	0.484	0.882	0.371	0.936	0.419	0.418	0.969	0.386
BN _{pp} 956* cv1	cv1	0.040	0.971	0.326	0.409	0.891	0.403	0.940	0.416	0.455	0.966	0.435	0.499	0.870	0.391	0.912	0.430	0.471	0.947	0.423
BN _{pp} 956* cv2	cv2	0.037	0.981	0.434	0.474	0.910	0.455	0.951	0.550	0.419	0.981	0.476	0.558	0.912	0.435	0.941	0.522	0.446	0.975	0.444
BN _{pp} 956* cv3	cv3	0.036	0.967	0.373	0.429	0.877	0.427	0.949	0.477	0.432	0.976	0.453	0.468	0.851	0.353	0.934	0.398	0.400	0.968	0.366
BN _{pp} 956* cv4	cv4	0.048	0.981	0.385	0.441	0.864	0.398	0.929	0.445	0.427	0.963	0.436	0.469	0.844	0.325	0.908	0.379	0.394	0.953	0.352
BN _{pp} 956*	TS125	0.043	0.954	0.408	0.418	0.863	0.402	0.938	0.437	0.434	0.967	0.435	0.500	0.837	0.355	0.909	0.403	0.406	0.957	0.381
BN _{pp} 956*	TS125 _{refined}	0.043	0.946	0.395	0.423	0.863	0.419	0.939	0.452	0.451	0.968	0.451	0.497	0.834	0.367	0.911	0.408	0.422	0.958	0.393
BN _{pp} 956*	TS125 _{curated}				0.491	0.895	0.461	0.945	0.502	0.478	0.973	0.489	0.545	0.868	0.400	0.922	0.449	0.449	0.963	0.423
BN _{pp} 956 cv0	cv0	0.042	0.966	0.480	0.519	0.921	0.481	0.949	0.456	0.564	0.967	0.504	0.525	0.905	0.426	0.933	0.431	0.518	0.959	0.443
BN _{pp} 956 cv1	cv1	0.036	0.967	0.384	0.444	0.903	0.427	0.939	0.419	0.503	0.962	0.457	0.526	0.887	0.417	0.912	0.453	0.501	0.947	0.447
BN _{pp} 956 cv2	cv2	0.044	0.986	0.489	0.504	0.915	0.474	0.952	0.549	0.454	0.979	0.497	0.587	0.918	0.469	0.942	0.554	0.487	0.973	0.477
BN _{pp} 956 cv3	cv3	0.043	0.962	0.436	0.489	0.907	0.462	0.950	0.495	0.482	0.975	0.488	0.531	0.888	0.405	0.937	0.440	0.462	0.965	0.419
BN _{pp} 956 cv4	cv4	0.063	0.964	0.417	0.458	0.873	0.414	0.933	0.476	0.425	0.968	0.449	0.491	0.860	0.342	0.910	0.402	0.403	0.955	0.368
BN _{pp} 956	TS125	0.044	0.949	0.425	0.448	0.882	0.435	0.941	0.468	0.466	0.969	0.467	0.500	0.843	0.370	0.912	0.417	0.428	0.957	0.395
BN _{pp} 956	TS125 _{refined}	0.046	0.942	0.411	0.444	0.881	0.440	0.941	0.470	0.472	0.969	0.471	0.505	0.842	0.374	0.913	0.407	0.434	0.957	0.400
BN _{pp} 956	TS125 _{curated}				0.518	0.910	0.484	0.948	0.535	0.489	0.975	0.511	0.569	0.880	0.411	0.925	0.445	0.460	0.967	0.432
BiteNet	TS125	0.075	0.730	0.187	0.250	0.789	0.245	0.943	0.461	0.156	0.989	0.233	0.413	0.773	0.196	0.912	0.341	0.154	0.989	0.193
BiteNet	TS125 _{refined}	0.077	0.717	0.189	0.252	0.786	0.236	0.943	0.454	0.148	0.990	0.223	0.414	0.771	0.186	0.912	0.324	0.144	0.989	0.183
BiteNet	TS125 _{curated}				0.269	0.821	0.220	0.943	0.451	0.129	0.991	0.201	0.445	0.809	0.157	0.913	0.289	0.120	0.990	0.154

Table S8: The AP_{residues}, ROC AUC, and MCC performance metrics calculated for the raw, refined, and curated structures in TS125 (PDB IDs : 1QRJ, 3LYV, and 4HFX did not pass the refinement procedure).

Name	TS125			TS125 _{refined}			TS125 _{curated}		
	AP _{residues}	AUC	MCC	AP _{residues}	AUC	MCC	AP _{residues}	AUC	MCC
1avpA	0.291	0.858	-0.042	0.234	0.801	-0.094	0.234	0.801	-0.094
1b05A	0.899	0.996	0.812	0.906	0.996	0.792	0.906	0.996	0.792
1bbrH	0.052	0.566	-0.041	0.051	0.566	0.024	0.863	0.993	0.709
1dpuA	0.106	0.199	-0.122	0.121	0.317	0.000	0.232	0.499	0.000
1e6iA	0.821	0.972	0.629	0.834	0.976	0.758	0.858	0.983	0.723
1emuA	0.736	0.947	0.443	0.747	0.950	0.519	0.747	0.950	0.519
1fchA	0.694	0.877	0.714	0.688	0.877	0.714	0.818	0.981	0.673
1h8bA	0.686	0.918	0.566	0.686	0.918	0.566	0.929	0.979	0.805
1htmD	0.903	0.987	0.617	0.903	0.987	0.617	0.890	0.987	0.691
1hv2A	0.229	0.725	0.231	0.230	0.726	0.231	0.814	0.924	0.437
1ik9A	0.055	0.536	-0.039	0.043	0.422	-0.045	0.287	0.866	0.305
1j19A	0.232	0.921	0.364	0.232	0.919	0.354	0.232	0.919	0.354
1j4xA	0.688	0.905	0.608	0.726	0.918	0.655	0.726	0.918	0.655
1jmqA	0.338	0.780	-0.007	0.347	0.789	0.125	0.752	0.951	0.000
1kugA	0.752	0.977	0.686	0.565	0.961	0.686	0.565	0.961	0.686
1m1dA	0.924	0.981	0.797	0.918	0.979	0.784	0.907	0.969	0.799
1m1jB	0.418	0.975	0.638	0.476	0.979	0.625	0.830	0.994	0.754
1m72C	0.739	0.979	0.687	0.699	0.978	0.742	0.739	0.974	0.677
1mw4A	0.905	0.987	0.658	0.884	0.986	0.600	0.873	0.979	0.767
1n12A	0.960	0.982	0.883	0.961	0.980	0.883	0.959	0.971	0.883
1ntvA	0.787	0.932	0.685	0.814	0.942	0.710	0.814	0.942	0.710
1o9kA	0.281	0.915	0.093	0.289	0.924	0.281	0.582	0.895	0.438

1o9kB	0.090	0.705	-0.054	0.060	0.624	-0.051	0.200	0.791	0.000
1om2A	0.697	0.959	0.514	0.739	0.960	0.514	0.775	0.951	0.000
1ozsA	0.660	0.913	0.634	0.688	0.906	0.620	0.693	0.815	0.573
1p4uA	0.767	0.941	0.649	0.767	0.941	0.753	0.767	0.941	0.753
1qrjB	0.091	0.658	0.000						
1r5iD	0.048	0.638	-0.034	0.045	0.616	-0.036	0.746	0.982	0.457
1uj0A	0.774	0.939	0.547	0.718	0.914	0.572	0.718	0.914	0.572
1wlpB	0.597	0.919	0.516	0.605	0.921	0.542	0.605	0.921	0.542
1x3zA	0.445	0.969	0.374	0.228	0.958	0.000	0.383	0.966	0.000
1yjmA	0.263	0.829	-0.079	0.407	0.845	0.237	0.481	0.908	0.119
1ywhA	0.595	0.844	0.510	0.550	0.861	0.510	0.705	0.956	0.404
1zv8A	0.257	0.396	0.000	0.344	0.402	0.000	0.350	0.571	0.000
2a06D	0.100	0.488	-0.028	0.101	0.485	-0.019	0.068	0.375	-0.085
2akaB	0.804	0.993	0.000	0.712	0.987	0.000	0.034	0.670	0.000
2astC	0.267	0.743	-0.077	0.267	0.743	-0.077	0.227	0.730	0.005
2bbuA	0.416	0.913	0.181	0.385	0.903	0.181	0.917	0.981	0.000
2bugA	0.527	0.929	0.452	0.524	0.925	0.452	0.725	0.934	0.606
2f4oA	0.563	0.965	0.441	0.346	0.929	0.000	0.521	0.960	0.302
2fciA	0.780	0.938	0.623	0.780	0.937	0.623	0.830	0.915	0.628
2ghqA	0.681	0.970	0.592	0.677	0.968	0.592	0.633	0.969	0.560
2h1cA	0.727	0.926	0.605	0.734	0.926	0.578	0.742	0.926	0.578
2ht9A	0.052	0.525	-0.065	0.053	0.540	-0.065	0.044	0.465	-0.058
2k3wA	0.648	0.866	0.188	0.648	0.866	0.188	0.622	0.805	0.361
2ks9A	0.032	0.633	0.000	0.021	0.522	0.000	0.036	0.778	-0.014
2lajA	0.821	0.935	0.741	0.776	0.921	0.692	0.829	0.928	0.572
2m32A	0.396	0.915	0.000	0.441	0.918	0.000	0.538	0.950	0.000

2np0A	0.036	0.798	-0.013	0.030	0.783	-0.012	0.030	0.783	-0.012
2oddA	0.611	0.748	0.521	0.646	0.754	0.513	0.779	0.873	0.488
2oeiA	0.818	0.910	0.584	0.818	0.910	0.584	0.818	0.910	0.584
2pg1E	0.633	0.849	0.302	0.529	0.853	0.399	0.829	0.930	0.653
2pkuA	0.927	0.984	0.767	0.927	0.984	0.767	0.965	0.994	0.851
2pvcA	0.667	0.983	0.492	0.688	0.985	0.384	0.661	0.985	0.589
2rr4A	0.786	0.930	0.645	0.777	0.929	0.621	0.898	0.962	0.749
2v5pA	0.152	0.834	0.000	0.250	0.809	0.000	0.238	0.805	0.000
2vdaA	0.130	0.846	0.073	0.119	0.837	0.064	0.096	0.817	0.021
2vn5C	0.608	0.936	0.309	0.629	0.940	0.385	0.725	0.937	0.000
2voyB	0.317	0.430	0.000	0.318	0.417	0.000	0.380	0.437	0.000
2xe4A	0.283	0.917	0.449	0.295	0.911	0.388	0.295	0.911	0.388
2y65C	0.629	0.908	0.582	0.649	0.916	0.604	0.765	0.967	0.589
2ykaA	0.683	0.905	0.511	0.581	0.872	0.449	0.832	0.968	0.626
2yleA	0.641	0.874	0.498	0.632	0.844	0.541	0.632	0.844	0.541
2zkrp	0.019	0.038	-0.022	0.019	0.045	-0.022	0.013	0.028	0.000
3a79A	0.065	0.863	-0.009	0.129	0.880	0.117	0.836	0.899	0.539
3bxmA	0.466	0.962	0.507	0.586	0.971	0.539	0.586	0.971	0.539
3chxC	0.109	0.504	-0.027	0.119	0.530	0.000	0.215	0.721	0.000
3eabA	0.822	0.949	0.727	0.790	0.932	0.695	0.750	0.911	0.682
3h6zA	0.279	0.890	0.415	0.436	0.887	0.431	0.400	0.852	0.370
3hqiA	0.723	0.973	0.634	0.700	0.973	0.548	0.719	0.981	0.567
3hufA	0.768	0.991	0.738	0.785	0.994	0.738	0.727	0.994	0.738
3j31Q	0.052	0.331	-0.078	0.082	0.483	-0.062	0.110	0.567	-0.026
3juaG	0.083	0.632	0.000	0.168	0.787	0.000	0.082	0.606	-0.043
3jzrA	0.823	0.967	0.737	0.820	0.967	0.737	0.820	0.967	0.737

3lcnB	0.217	0.733	0.182	0.217	0.734	0.093	0.179	0.737	-0.121
3lyvE	0.461	0.558	0.000						
3mp1A	0.065	0.842	-0.011	0.121	0.877	-0.011	0.121	0.877	-0.011
3nfkA	0.892	0.959	0.750	0.902	0.967	0.750	0.899	0.977	0.835
3nk3A	0.770	0.952	0.546	0.777	0.961	0.576	0.804	0.967	0.637
3ntwA	0.686	0.933	0.000	0.726	0.936	0.000	0.751	0.942	0.000
3o0eA	0.352	0.855	0.311	0.346	0.850	0.311	0.261	0.775	0.254
3qhrA	0.149	0.880	-0.034	0.148	0.880	-0.029	0.199	0.896	0.000
3qksB	0.020	0.160	-0.085	0.019	0.138	-0.087	0.309	0.953	0.000
3rbqA	0.815	0.955	0.640	0.794	0.950	0.619	0.805	0.950	0.674
3rceA	0.395	0.956	0.500	0.387	0.960	0.536	0.395	0.960	0.571
3sovA	0.829	0.986	0.695	0.819	0.985	0.695	0.819	0.985	0.695
3srjA	0.785	0.951	0.667	0.781	0.954	0.643	0.755	0.943	0.619
3szmC	0.395	0.935	0.243	0.395	0.935	0.243	0.230	0.871	-0.036
3twtC	0.904	0.976	0.744	0.898	0.976	0.744	0.884	0.977	0.724
3tzdA	0.822	0.901	0.587	0.868	0.927	0.587	0.868	0.927	0.587
3ufmA	0.162	0.893	0.151	0.159	0.890	0.140	0.159	0.890	0.140
3upvA	0.797	0.975	0.637	0.797	0.976	0.701	0.797	0.976	0.701
3v2xA	0.852	0.969	0.671	0.866	0.969	0.649	0.866	0.969	0.649
3v7dB	0.871	0.994	0.781	0.861	0.993	0.733	0.852	0.993	0.683
3wdcA	0.250	0.841	0.277	0.250	0.840	0.277	0.250	0.840	0.277
3wmgA	0.019	0.658	-0.020	0.022	0.691	-0.020	0.022	0.691	-0.020
3zdyD	0.329	0.975	0.000	0.674	0.989	0.376	0.482	0.981	0.527
3zg5A	0.226	0.885	0.000	0.170	0.888	0.000	0.111	0.884	-0.012
3zilA	0.771	0.976	0.630	0.788	0.978	0.630	0.788	0.978	0.630
4aktA	0.576	0.942	0.505	0.612	0.940	0.505	0.511	0.939	0.438

4arfA	0.626	0.989	0.771	0.624	0.990	0.809	0.624	0.990	0.809
4b6hA	0.814	0.968	0.783	0.821	0.967	0.707	0.882	0.970	0.634
4btaA	0.849	0.933	0.650	0.870	0.933	0.760	0.860	0.936	0.728
4buoB	0.593	0.973	0.499	0.643	0.973	0.411	0.726	0.982	0.713
4cc9A	0.022	0.639	-0.026	0.022	0.638	-0.030	0.593	0.991	0.560
4cguA	0.691	0.960	0.639	0.711	0.957	0.701	0.713	0.958	0.701
4dcbA	0.846	0.989	0.648	0.848	0.989	0.680	0.848	0.989	0.680
4extC	0.866	0.974	0.663	0.863	0.973	0.724	0.869	0.973	0.724
4gvcA	0.855	0.969	0.634	0.855	0.969	0.634	0.855	0.969	0.634
4gwqG	0.018	0.084	0.000	0.018	0.069	-0.028	0.019	0.379	-0.035
4hanA	0.170	0.846	0.206	0.147	0.821	0.249	0.124	0.816	0.122
4hdqA	0.235	0.887	0.424	0.220	0.884	0.424	0.637	0.954	0.626
4hfxB	0.519	0.775	0.209						
4i6mA	0.040	0.630	-0.036	0.046	0.675	-0.029	0.029	0.447	-0.021
4igqA	0.681	0.982	0.676	0.624	0.981	0.659	0.624	0.981	0.659
4je8A	0.702	0.973	0.597	0.770	0.975	0.522	0.828	0.988	0.723
4ktyB	0.772	0.982	0.634	0.760	0.973	0.642	0.672	0.967	0.605
4l3oA	0.825	0.973	0.635	0.864	0.979	0.607	0.848	0.969	0.607
4lajA	0.160	0.794	0.000	0.128	0.738	0.000	0.161	0.809	-0.039
4m38A	0.268	0.941	0.457	0.768	0.979	0.553	0.602	0.975	0.594
4m91A	0.258	0.888	0.174	0.230	0.870	-0.059	0.251	0.877	-0.059
4m9sB	0.017	0.517	0.000	0.018	0.541	0.000	0.023	0.636	0.000
4mi8B	0.717	0.948	0.580	0.722	0.941	0.552	0.717	0.951	0.580
4nufA	0.125	0.868	0.022	0.126	0.869	0.170	0.126	0.869	0.170
4po6A	0.666	0.905	0.668	0.704	0.916	0.685	0.704	0.916	0.685

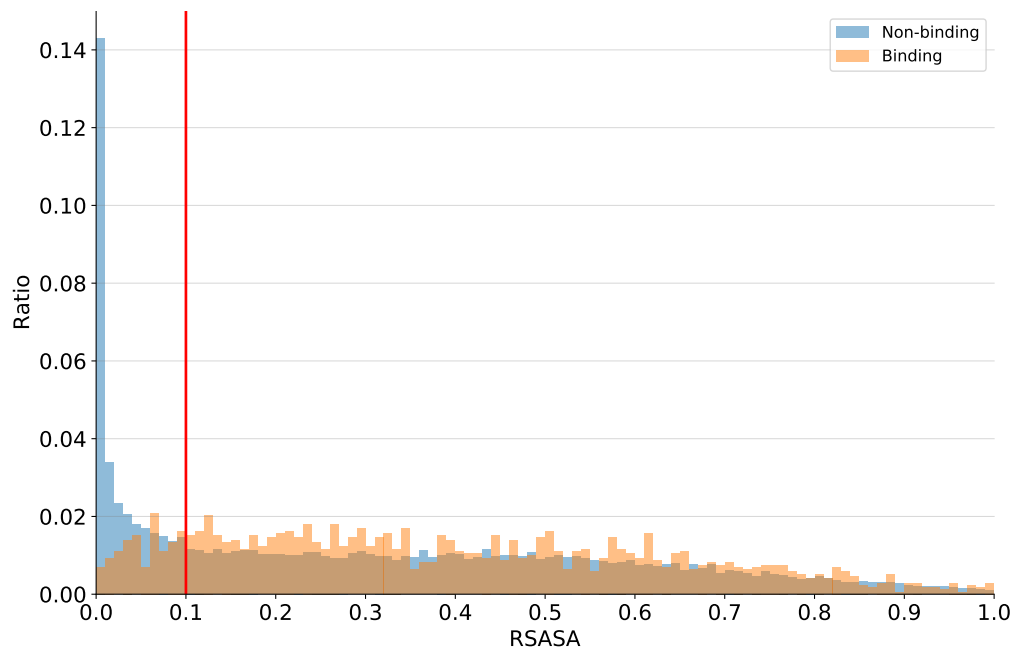


Figure S1: The distribution of the relative solvent accessible surface area for the non-binding (blue) and binding (orange) amino acid residues in TS125.

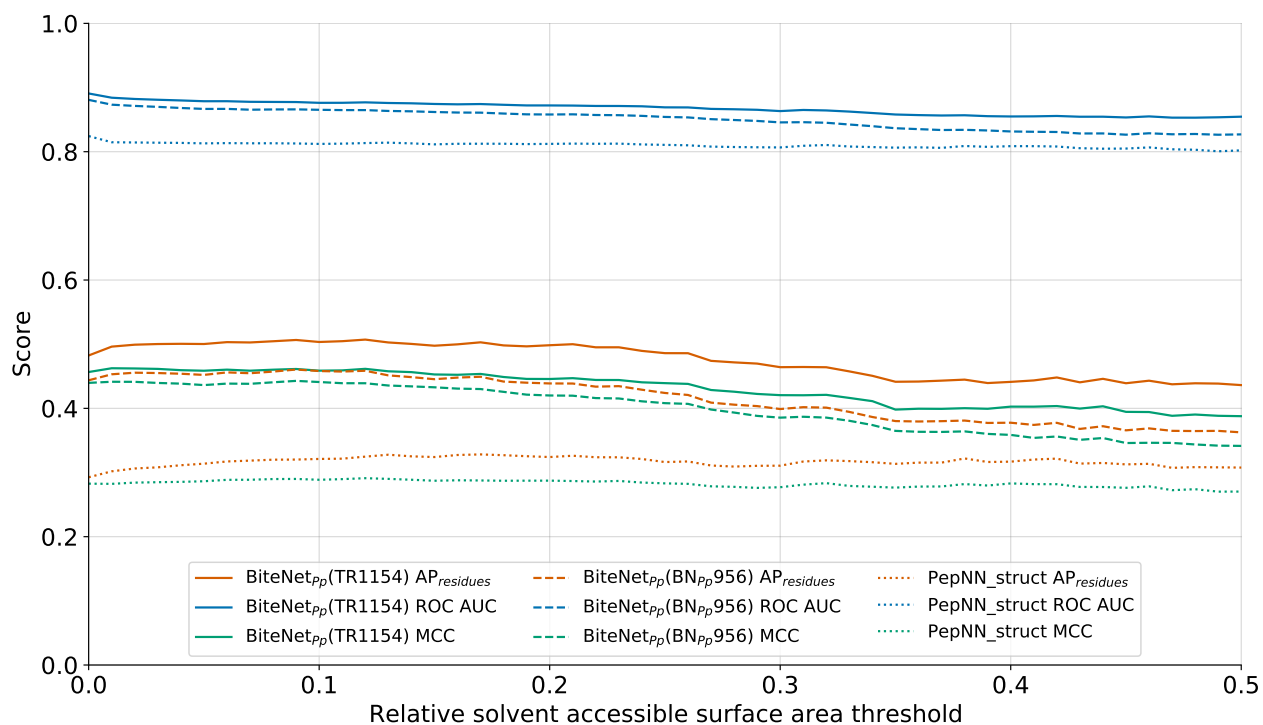


Figure S2: The residue-based AP (orange), ROC AUC (blue) and MCC (green) calculated for the amino acid residues with the relative solvent accessible surface area higher than the threshold. The performance metrics for BiteNet_{pp}(TR1154), BiteNet_{pp}(BN_{pp}956), and PepNN_{struct} are shown with solid, dashed and dotted lines, respectively.

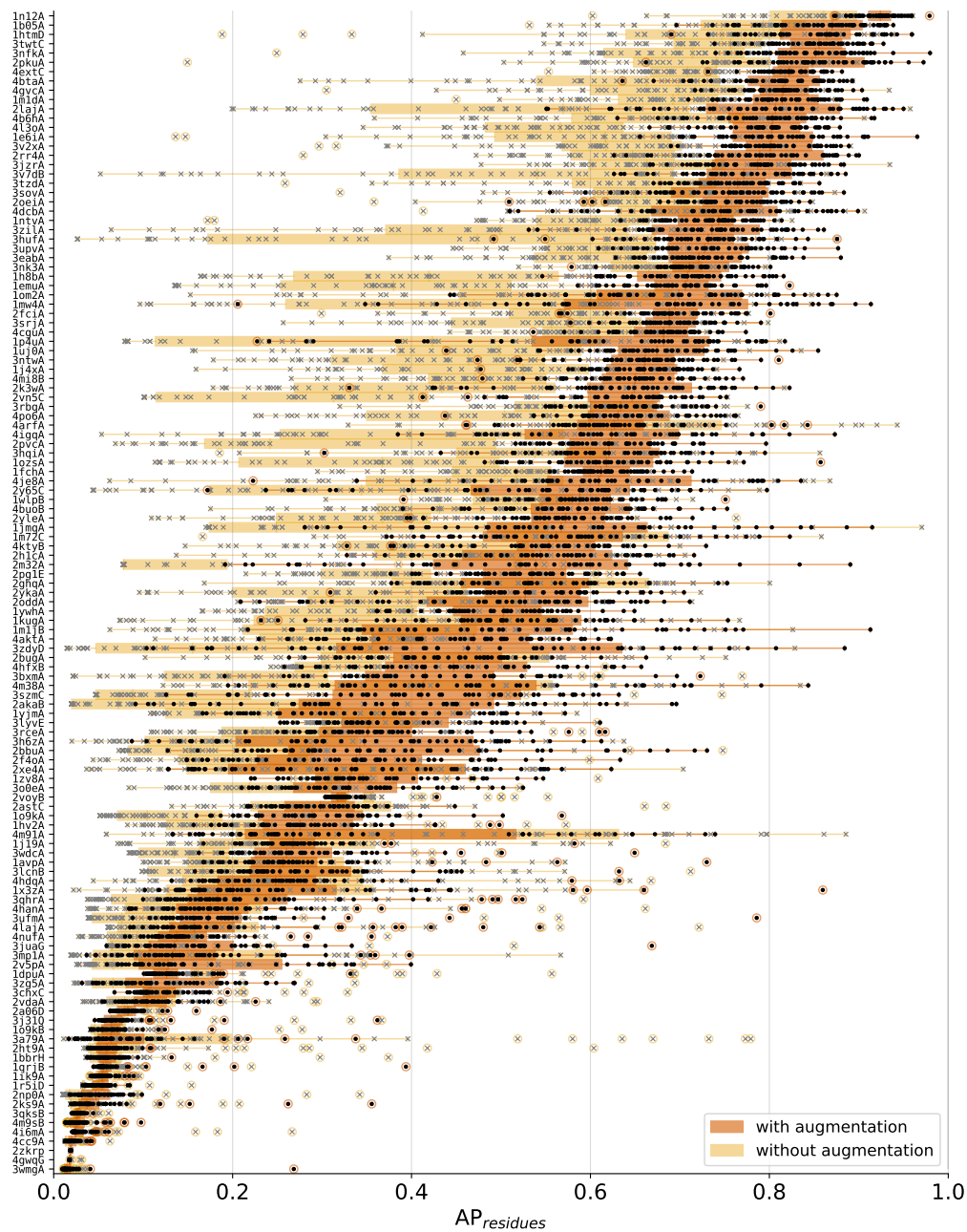


Figure S3: The distribution of the residue-based AP for 51 different orientations (1 original + 50 generated orientations) of each complex in TS125 for BiteNet_{P_p} models trained on BN_{P_p}956 with data augmentation (dark orange bars, black dots) and without it (light orange bars, gray crosses).

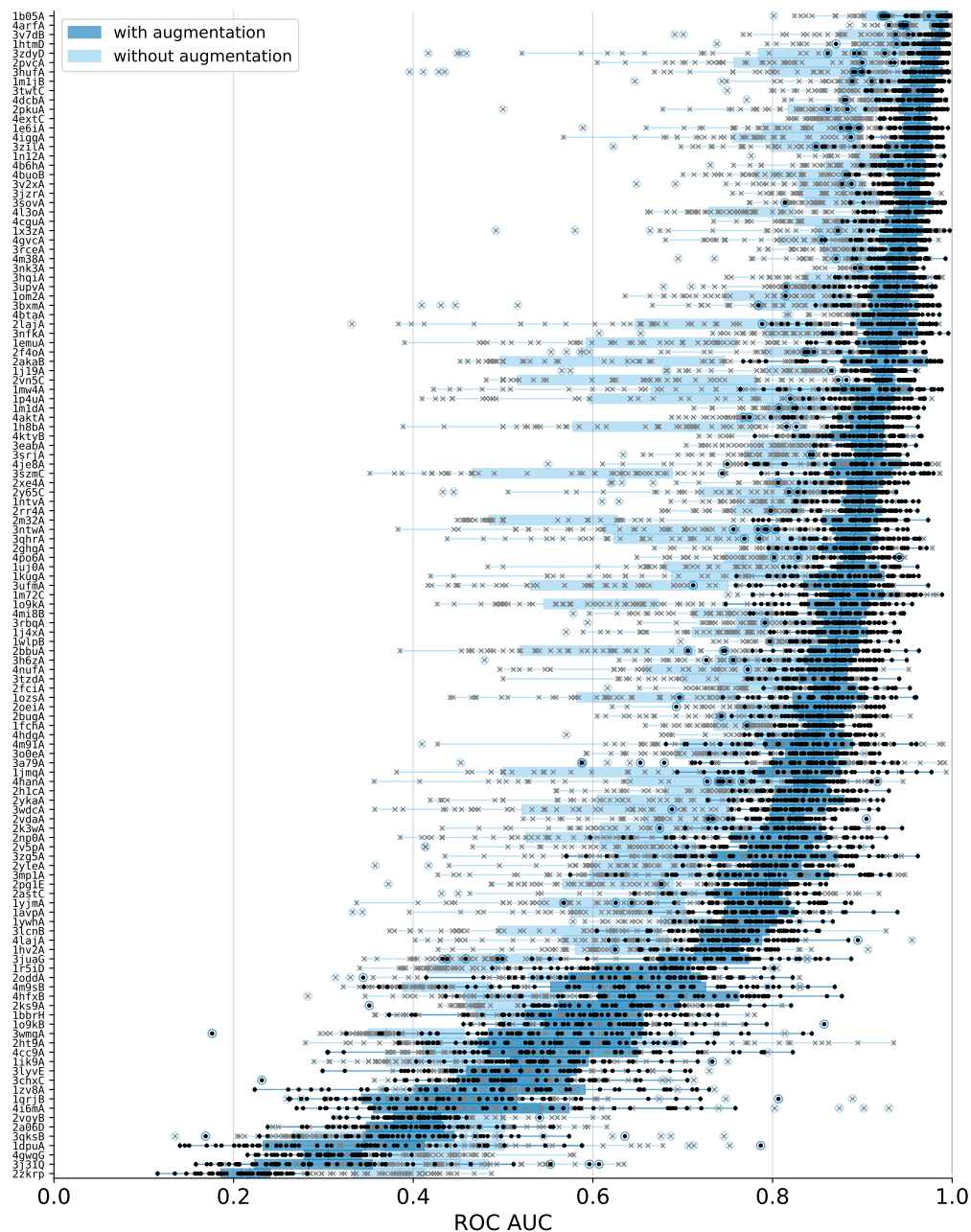


Figure S4: The distribution of the residue-based ROC AUC for 51 different orientations (1 original + 50 generated orientations) of each complex in TS125 for BiteNet_{P_p} models trained on BN_{P_p}956 with data augmentation (dark blue bars, black dots) and without it (light blue bars, gray crosses).

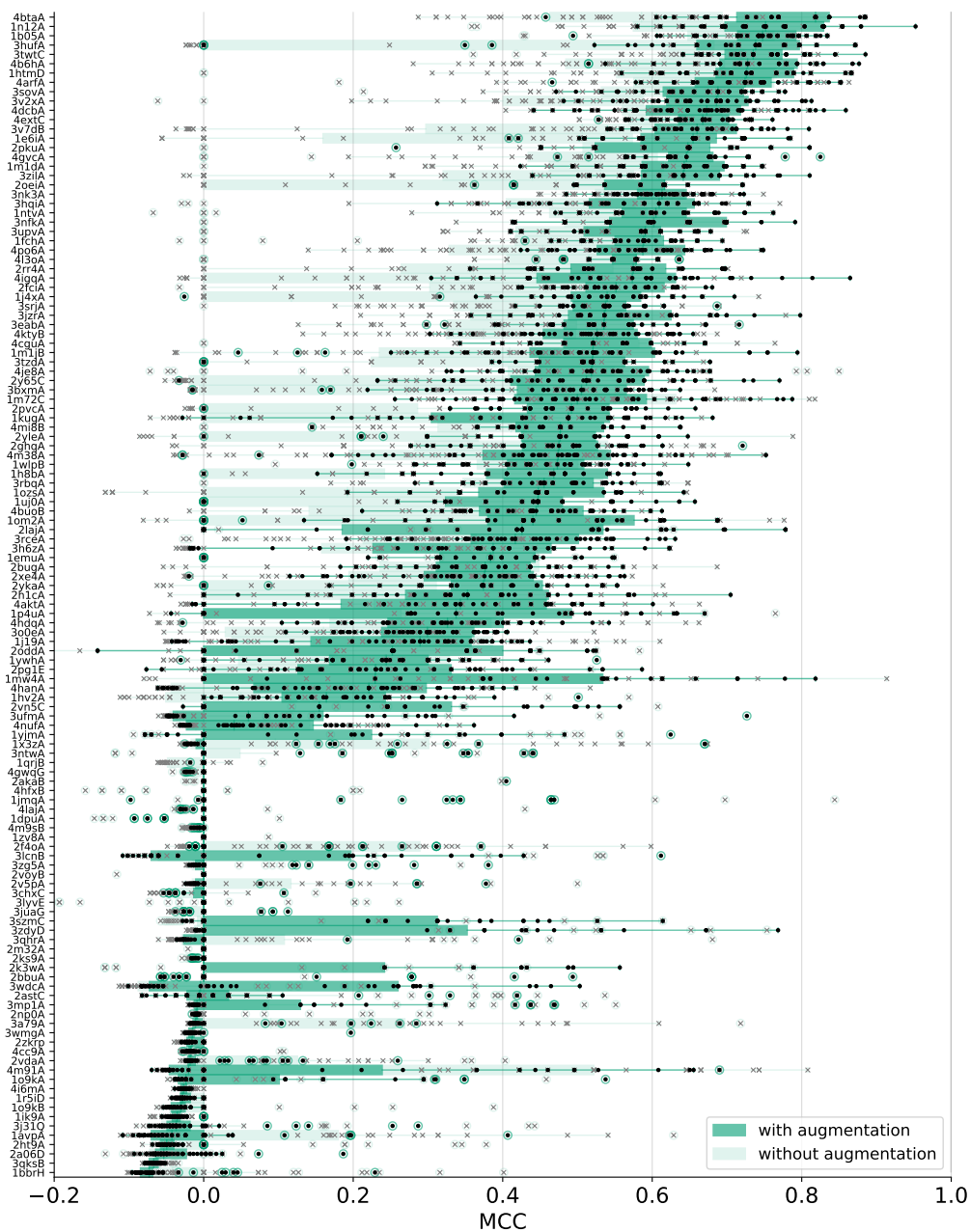


Figure S5: The distribution of the residue-based MCC for 51 different orientations (1 original + 50 generated orientations) of each complex in TS125 for BiteNet_{Pp} models trained on BN_{Pp}956 with data augmentation (dark green bars, black dots) and without it (light green bars, gray crosses).

References

- (1) Kozlovskii, I.; Popov, P. Spatiotemporal Identification of Druggable Binding Sites Using Deep Learning. *Commun. Biol.* **2020**, *3*, 1–12.
- (2) Zhao, Z.; Peng, Z.; Yang, J. Improving Sequence-based Prediction of Protein–peptide Binding Residues by Introducing Intrinsic Disorder and a Consensus Method. *J. Chem. Inf. Model.* **2018**, *58*, 1459–1468.
- (3) Krivák, R.; Jendele, L.; Hoksza, D. Peptide-binding Site Prediction from Protein Structure via Points on the Solvent Accessible Surface. Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2018; pp 645–650.
- (4) Abdin, O.; Wen, H.; Kim, P. M. Sequence and Structure Based Deep Learning Models for the Identification of Peptide Binding Sites. *Machine Learning for Structural Biology Workshop at Conference on Neural Information Processing Systems (NeurIPS)* **2020**,
- (5) Sun, Z.; Zheng, S.; Zhao, H.; Niu, Z.; Lu, Y.; Pan, Y.; Yang, Y. To Improve the Predictions of Binding Residues with Dna, Rna, Carbohydrate, and Peptide via Multiple-task Deep Neural Networks. *bioRxiv* **2020**,
- (6) Taherzadeh, G.; Zhou, Y.; Liew, A. W.-C.; Yang, Y. Structure-based Prediction of Protein–peptide Binding Regions Using Random Forest. *Bioinformatics* **2018**, *34*, 477–484.
- (7) Lavi, A.; Ngan, C. H.; Movshovitz-Attias, D.; Bohnuud, T.; Yueh, C.; Beglov, D.; Schueler-Furman, O.; Kozakov, D. Detection of Peptide-binding Sites on Protein Surfaces: The First Step Toward the Modeling and Targeting of Peptide-mediated Interactions. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 2096–2105.

- (8) Segura, J.; Jones, P. F.; Fernandez-Fuentes, N. A Holistic in Silico Approach to Predict Functional Sites in Protein Structures. *Bioinformatics* **2012**, *28*, 1845–1850.
- (9) Taherzadeh, G.; Yang, Y.; Zhang, T.; Liew, A. W.-C.; Zhou, Y. Sequence-based Prediction of Protein–peptide Binding Sites Using Support Vector Machine. *J. Comput. Chem.* **2016**, *37*, 1223–1229.
- (10) Trabuco, L. G.; Lise, S.; Petsalaki, E.; Russell, R. B. Pepsite: Prediction of Peptide-binding Sites from Protein Surfaces. *Nucleic Acids Res.* **2012**, *40*, W423–W427.
- (11) Wardah, W.; Dehzangi, A.; Taherzadeh, G.; Rashid, M. A.; Khan, M. G.; Tsunoda, T.; Sharma, A. Predicting Protein-peptide Binding Sites with a Deep Convolutional Neural Network. *J. Theor. Biol.* **2020**, 110278.