NeBcon: Protein contact map prediction using neural network training coupled with naïve Bayes classifiers

Baoji He, S. M. Mortuza, Yanting Wang, Hong-Bin Shen, Yang Zhang

Supporting Information

Table S1. List of the target type, the number of sequences in MSA (N_{MSA}), and the accuracy of top $L/5$ long-range contact predictions by NeBcon and the component predictors for all 98 testing proteins. To save space, the name of each program is represented by the first 4 letters of the full-name, e.g. 'BETA' indicates 'BETACON'.

Protein	Type	N_{MSA}	Accuracy of top $L/5$ long-range contact predictions								
			BETA	SVMc	SVMs	PSIC	CCMp	Free	STRU	Meta	NeBc
1i3cA	Easy	420.5L	0.893	0.321	0.786	0.571	0.464	0.464	0.607	0.929	1.000
1ithA	Easy	21.7L	0.071	0.107	0.214	0.286	0.286	0.143	0.464	0.786	0.750
1c1lA	Easy	6.6L	0.852	0.000	0.192	0.370	0.259	0.111	0.519	1.000	1.000
1czpA	Easy	70.0L	0.947	0.421	0.053	0.526	0.789	0.526	0.579	1.000	1.000
1e29A	Easy	46.8L	0.222	0.111	0.148	0.370	0.556	0.407	0.852	1.000	0.741
1f86A	Easy	6.0L	0.304	0.391	0.696	0.478	0.304	0.261	0.783	0.957	1.000
1gmxA	Easy	125.5L	0.714	0.524	0.857	0.476	0.667	0.619	0.429	0.952	0.905
1h4xA	Easy	60.4L	0.864	0.727	0.909	0.636	0.818	0.773	0.455	0.864	0.955
1h97A	Easy	19.8L	0.345	0.138	0.172	0.276	0.379	0.310	0.207	0.690	0.759
1jbeA	Easy	481.2L	0.960	0.840	0.640	0.600	0.640	0.520	0.640	0.920	0.960
1khyA	Easy	29.5L	0.077	0.000	0.000	0.462	0.577	0.500	0.577	0.885	0.846
1nwwA	Easy	35.0L	0.448	0.207	0.241	0.586	0.552	0.517	0.655	0.828	0.828
1pbjA	Easy	174.4L	0.000	0.000	0.000	0.304	0.478	0.261	0.250	0.609	0.375
1q77A	Easy	76.5L	0.815	0.593	0.654	0.407	0.296	0.370	0.519	0.630	0.889
1q9uA	Easy	5.4L	0.240	0.360	0.440	0.240	0.360	0.160	0.600	0.840	0.960
1r4vA	Easy	0.3L	0.034	0.000	0.069	0.000	0.034	0.034	0.167	0.103	0.400
1rl2A	Easy	16.7L	0.231	0.000	0.154	0.423	0.385	0.385	0.296	0.885	0.852
1rxdA	Easy	43.2L	0.552	0.310	0.345	0.379	0.414	0.483	0.267	0.897	0.767
1s5uA	Easy	77.7L	0.160	0.280	0.280	0.640	0.600	0.440	0.680	1.000	1.000
1sbxA	Easy	1.3L	0.050	0.000	0.389	0.050	0.150	0.050	0.095	0.150	0.381
1seiA	Easy	17.0L	0.346	0.231	0.115	0.462	0.500	0.538	0.462	0.846	0.808
1ss4B	Easy	118.1L	0.000	0.107	0.000	0.500	0.464	0.321	0.310	0.286	0.345
1v05A	Easy	9.5L	0.474	0.579	0.632	0.632	0.421	0.368	0.632	0.895	0.947
1v30A	Easy	11.4L	0.565	0.391	0.696	0.348	0.348	0.217	0.304	0.870	0.957
1w0nA	Easy	1.3L	0.500	0.208	0.125	0.167	0.167	0.125	0.083	0.083	0.875
1yqhA	Easy	8.0L	0.650	0.550	0.600	0.450	0.500	0.400	0.600	0.800	1.000

1yrkA	Easy	28.9L	0.360	0.360	0.600	0.320	0.320	0.200	0.120	0.800	0.920
1zceA	Easy	5.3L	0.259	0.074	0.185	0.407	0.481	0.444	0.414	0.741	0.793
1zd0A	Easy	1.5L	0.464	0.357	0.321	0.143	0.036	0.036	0.107	0.036	0.357
1zmaA	Easy	166.6L	0.783	0.652	0.773	0.609	0.522	0.609	0.565	0.739	0.826
2b0vA	Easy	116.3L	0.207	0.276	0.172	0.621	0.759	0.655	0.690	0.931	1.000
2bwqA	Easy	59.0L	0.542	0.333	0.000	0.583	0.625	0.667	0.417	1.000	0.958
2c6uA	Easy	44.5L	0.750	0.667	0.667	0.478	0.375	0.333	0.542	0.500	1.000
2cc3A	Easy	4.0L	0.464	0.429	0.357	0.107	0.179	0.071	0.536	0.643	0.571
2d9rA	Easy	2.9L	0.938	0.688	0.533	0.438	0.750	0.500	0.765	0.875	0.882
2e56A	Easy	0.5L	0.357	0.071	0.250	0.107	0.036	0.036	0.000	0.036	0.821
2g2cA	Easy	33.8L	0.900	0.400	0.733	0.567	0.633	0.433	0.533	1.000	1.000
2gj3B	Easy	248.5L	0.478	0.391	0.304	0.522	0.652	0.435	0.565	0.522	0.957
2gu3A	Easy	1.6L	0.000	0.000	0.000	0.200	0.160	0.080	0.360	0.120	0.440
2gu9A	Easy	139.2L	0.545	0.091	0.545	0.818	0.864	0.818	0.727	1.000	1.000
2h1cA	Easy	31.6L	0.370	0.444	0.038	0.444	0.556	0.407	0.630	0.889	0.704
2hkvA	Easy	18.9L	0.034	0.172	0.138	0.138	0.310	0.379	0.138	0.793	0.586
2ib0A	Easy	2.0L	0.107	0.107	0.321	0.143	0.214	0.179	0.286	0.607	0.429
2im8A	Easy	0.6L	0.042	0.292	0.167	0.125	0.083	0.000	0.375	0.333	0.500
2jdcA	Easy	221.1L	0.172	0.138	0.241	0.448	0.414	0.448	0.517	0.862	0.862
2o7aA	Easy	7.6L	0.130	0.043	0.045	0.391	0.348	0.391	0.522	0.435	0.522
2o90A	Easy	14.4L	0.261	0.261	0.304	0.522	0.435	0.304	0.565	0.739	0.739
2oggA	Easy	28.6L	0.321	0.000	0.107	0.429	0.464	0.500	0.500	0.536	0.964
2otmA	Easy	35.9L	0.483	0.621	0.621	0.517	0.448	0.448	0.433	0.759	0.867
4nklA	Easy	15.1L	0.040	0.160	0.240	0.600	0.560	0.520	0.600	0.840	0.920
1hxrA	Hard	420.5L	0.286	0.095	0.000	0.095	0.238	0.095	0.190	0.429	0.286
1mwpA	Hard	21.7L	0.000	0.000	0.000	0.000	0.000	0.000	0.263	0.000	0.579
1oo0A	Hard	6.6L	0.071	0.000	0.000	0.000	0.036	0.000	0.071	0.071	0.071
1wnaA	Hard	70.0L	0.458	0.375	0.125	0.000	0.000	0.000	0.120	0.000	0.160
1x0tA	Hard	46.8L	0.095	0.143	0.143	0.095	0.238	0.190	0.333	0.524	0.714
256bA	Hard	6.0L	0.238	0.429	0.143	0.300	0.381	0.190	0.381	0.667	0.714
2djfA	Hard	125.5L	0.043	0.435	0.304	0.261	0.130	0.043	0.870	0.174	0.565
2dqaA	Hard	60.4L	0.042	0.167	0.042	0.000	0.042	0.083	0.250	0.208	0.333
2f6mB	Hard	19.8L	0.190	0.048	0.048	0.095	0.000	0.000	0.190	0.286	0.524
2fzpA	Hard	481.2L	0.192	0.269	0.577	0.000	0.077	0.038	0.423	0.231	0.385
2ga1A	Hard	29.5L	0.300	0.200	0.278	0.316	0.250	0.200	0.050	0.500	0.100
2hd0A	Hard	35.0L	0.120	0.000	0.000	0.000	0.000	0.000	0.040	0.000	0.04
2hl7A	Hard	174.4L	0.000	0.312	0.267	0.438	0.312	0.188	0.312	0.562	0.688
2o27A	Hard	76.5L	0.440	0.200	0.120	0.000	0.040	0.040	0.160	0.240	0.400
2sicI	Hard	5.4L	0.476	0.238	0.095	0.190	0.143	0.143	0.381	0.286	0.429
2v7fA	Hard	0.3L	0.111	0.074	0.154	0.111	0.222	0.148	0.259	0.407	0.481
2vlqB	Hard	16.7L	0.308	0.231	0.708	0.115	0.038	0.038	0.038	0.192	0.385
2w9yA	Hard	43.2L	0.037	0.111	0.038	0.037	0.037	0.000	0.074	0.185	0.519
2x5cA	Hard	77.7L	0.000	0.000	0.111	0.000	0.000	0.000	0.000	0.105	0.211

2x5rA	Hard	1.3L	0.500	0.000	0.043	0.000	0.000	0.000	0.000	0.083	0.042
2xetA	Hard	17.0L	0.353	0.118	0.176	0.529	0.588	0.412	0.471	0.824	0.941
2y6xA	Hard	118.1L	0.000	0.095	0.143	0.048	0.000	0.048	0.095	0.429	0.619
2ygnA	Hard	9.5L	0.107	0.500	0.160	0.074	0.000	0.071	0.143	0.321	0.286
2ynqA	Hard	11.4L	0.138	0.207	0.148	0.034	0.069	0.000	0.069	0.241	0.276
3a4cA	Hard	1.3L	0.048	0.095	0.000	0.000	0.000	0.048	0.048	0.095	0.476
3bcyA	Hard	8.0L	0.034	0.310	0.138	0.000	0.000	0.034	0.172	0.241	0.345
3cbnA	Hard	28.9L	0.321	0.250	0.071	0.357	0.357	0.286	0.286	0.714	0.750
3ci0J	Hard	5.3L	0.000	0.034	0.276	0.241	0.138	0.172	0.233	0.276	0.767
3dt5A	Hard	1.5L	0.000	0.043	0.048	0.000	0.000	0.000	0.087	0.000	0.261
3e9vA	Hard	166.6L	0.261	0.696	0.130	0.217	0.174	0.174	0.250	0.739	0.417
3fdqA	Hard	116.3L	0.036	0.036	0.120	0.000	0.000	0.000	0.107	0.250	0.214
3fkeA	Hard	59.0L	0.333	0.333	0.250	0.000	0.000	0.000	0.375	0.333	0.333
3h0dA	Hard	44.5L	0.276	0.034	0.207	0.103	0.103	0.069	0.097	0.310	0.387
3iezA	Hard	4.0L	0.111	0.167	0.000	0.056	0.111	0.111	0.111	0.056	0.333
3mswA	Hard	2.9L	0.074	0.000	0.000	0.074	0.000	0.000	0.148	0.074	0.259
3pmcA	Hard	0.5L	0.077	0.115	0.080	0.038	0.000	0.000	0.231	0.269	0.462
3rkoG	Hard	33.8L	0.000	0.100	0.053	0.350	0.550	0.300	0.100	0.550	0.800
3u12A	Hard	248.5L	0.000	0.000	0.050	0.000	0.050	0.000	0.100	0.200	0.050
3zn4A	Hard	1.6L	0.552	0.310	0.207	0.000	0.000	0.000	0.621	0.310	0.690
4ajyV	Hard	139.2L	0.586	0.552	0.586	0.207	0.138	0.034	0.483	0.310	0.586
4ipcA	Hard	31.6L	0.500	0.000	0.167	0.125	0.250	0.208	0.333	0.458	0.625
4ipdA	Hard	18.9L	0.458	0.042	0.292	0.167	0.250	0.208	0.292	0.458	0.667
4ipgA	Hard	2.0L	0.458	0.083	0.292	0.250	0.167	0.167	0.250	0.458	0.625
4jm1A	Hard	0.6L	1.000	0.500	0.750	0.000	0.000	0.000	0.688	0.625	0.625
4jw3A	Hard	221.1L	0.227	0.682	0.091	0.227	0.136	0.091	0.864	0.273	0.818
4l1pA	Hard	7.6L	0.000	0.000	0.160	0.200	0.240	0.200	0.28	0.280	0.440
4ljhA	Hard	14.4L	0.167	0.042	0.083	0.000	0.083	0.000	0.292	0.542	0.708
4lvrA	Hard	28.6L	0.000	0.083	0.083	0.000	0.000	0.000	0.000	0.208	0.250
Average	Easy	62.2L	0.406	0.288	0.341	0.406	0.432	0.364	0.459	0.709	0.798
Average	Hard	2.6L	0.198	0.181	0.167	0.134	0.119	0.094	0.242	0.312	0.451
Average	All	33.0L	0.310	0.236	0.255	0.262	0.278	0.232	0.353	0.515	0.628

Table S2. Comparison of accuracy of top L long-range contact predictions between individual methods and pair-wise combinations. The accuracy of individual methods is listed in diagonal cells (highlight in gray), that combined by NBC is in upper triangle cells, and that combined by shearing is in lower triangle cells. The upper part of the table is for Easy targets and the lower part of the table for Hard targets. To save space, the name of each program is represented by the first 4 letters of the full-name, e.g. 'BETA' indicates 'BETACON'. This table is similar to Table 1 but with top L (instead of top $L/5$) long-range contact predictions.

	BETA	SVMc	SVMS	PSIC	CCMp	mfDC
50 Easy targets						
BETA	0.195	0.217	0.228	0.302	0.291	0.291
SVMc	0.214	0.187	0.220	0.287	0.270	0.279
SVMS	0.226	0.211	0.208	0.301	0.289	0.291
PSIC	0.279	0.257	0.274	0.232	0.260	0.260
CCMp	0.198	0.175	0.191	0.206	0.250	0.246
Free	0.271	0.248	0.267	0.245	0.198	0.229
48 Hard targets						
BETA	0.117	0.140	0.129	0.129	0.123	0.106
SVMc	0.137	0.113	0.141	0.136	0.133	0.109
SVMS	0.131	0.135	0.131	0.127	0.118	0.105
PSIC	0.117	0.116	0.110	0.057	0.063	0.069
CCMp	0.097	0.095	0.090	0.058	0.061	0.056
Free	0.109	0.107	0.102	0.066	0.050	0.071

Table S3: Average accuracy of top *L*/5 contact predictions by different methods on 98 non-redundant test proteins. Values in parentheses are p-values in student t-test relative to NeBcon.

Methods	**Short (6-11)**	**Medium (12-24)**	**Long (>24)**
50 Easy targets			
BETACON	0.576 (2×10^{-7})	0.467 (9×10^{-9})	0.406 (2×10^{-12})
SVMSEQ	0.510 (2×10^{-12})	0.409 (5×10^{-11})	0.288 (2×10^{-12})
SVMcon	0.602 (2×10^{-9})	0.500 (6×10^{-8})	0.341 (2×10^{-12})
PSICOV	0.293 (2×10^{-12})	0.369 (2×10^{-12})	0.406 (2×10^{-12})
CCMpred	0.332 (2×10^{-12})	0.400 (2×10^{-11})	0.432 (2×10^{-12})
FreeContact	0.296 (2×10^{-12})	0.356 (2×10^{-12})	0.364 (2×10^{-12})
STRUCTCH	0.633 (6×10^{-5})	0.540 (1×10^{-3})	0.459 (2×10^{-12})
MetaPSICOV	0.647 (7×10^{-3})	0.662 (2×10^{-1})	0.709 (3×10^{-3})
NeBcon	0.707	0.643	0.798
48 Hard targets			
BETACON	0.502 (3×10^{-4})	0.392 (8×10^{-4})	0.198 (2×10^{-7})
SVMSEQ	0.439 (4×10^{-8})	0.376 (3×10^{-5})	0.181 (2×10^{-9})
SVMcon	0.525 (3×10^{-3})	0.409 (2×10^{-3})	0.167 (3×10^{-10})
PSICOV	0.112 (2×10^{-12})	0.117 (2×10^{-12})	0.134 (2×10^{-12})
CCMpred	0.132 (2×10^{-12})	0.152 (2×10^{-12})	0.119 (2×10^{-12})
FreeContact	0.111 (2×10^{-12})	0.115 (2×10^{-12})	0.094 (2×10^{-12})
STRUCTCH	0.577 (2×10^{-1})	0.432 (6×10^{-3})	0.242 (1×10^{-9})
MetaPSICOV	0.502 (9×10^{-5})	0.478 (2×10^{-1})	0.312 (3×10^{-6})
NeBcon	0.593	0.501	0.451

Table S4: Shannon entropy of contact maps generated by different methods using different numbers of top contacts and cell divisions. Values in parentheses are p-values in student t-test of each method relative to the entropy for NeBcon.

Methods	**Short**	**Medium**	**Long**	**All**
Entropy calculated by top-L contacts on **10×10** ***cells***				
BETACON	3.454 ($3.1*10^{-2}$)	2.170 ($8.0*10^{-14}$)	2.573 ($1.4*10^{-11}$)	3.377 ($7.9*10^{-13}$)
SVMSEQ	3.513 ($7.2*10^{-5}$)	3.209 ($2.7*10^{-4}$)	3.493 ($1.6*10^{-2}$)	3.631 ($9.1*10^{-5}$)
SVMcon	3.351 ($3.6*10^{-4}$)	2.839 ($1.4*10^{-13}$)	3.212 ($5.9*10^{-10}$)	3.453 ($4.6*10^{-11}$)
PSICOV	3.058 ($2.2*10^{-2}$)	2.995 ($2.7*10^{-1}$)	3.471 ($2.3*10^{-1}$)	3.705 ($2.8*10^{-1}$)
CCMpred	3.806 ($1.9*10^{-6}$)	3.788 ($4.5*10^{-14}$)	4.388 ($5.9*10^{-14}$)	4.708 ($8.4*10^{-15}$)
FreeContact	3.737 ($5.5*10^{-5}$)	3.751 ($2.4*10^{-13}$)	4.447 ($3.0*10^{-15}$)	4.694 ($2.9*10^{-14}$)
STRUCTCH	3.374 ($9.9*10^{-2}$)	3.110 ($3.2*10^{-1}$)	3.50 ($4.7*10^{-8}$)	3.707 ($9.9*10^{-3}$)
MetaPSICOV	3.557 ($9.2*10^{-4}$)	3.181 ($1.6*10^{-2}$)	3.537 ($6.8*10^{-2}$)	3.757 ($1.1*10^{-1}$)
NeBcon	3.414	3.098	3.615	3.828
Entropy calculated by top-L/5 contacts on **10×10** ***cells***				
BETACON	2.334 ($8.5*10^{-4}$)	1.44 ($1.1*10^{-11}$)	1.680 ($6.5*10^{-9}$)	2.097 ($4.4*10^{-9}$)
SVMSEQ	2.356 ($2.4*10^{-3}$)	1.989 ($8.6*10^{-3}$)	2.301 ($1.9*10^{-2}$)	2.278 ($9.1*10^{-5}$)
SVMcon	2.266 ($2.6*10^{-6}$)	1.860 ($4.5*10^{-6}$)	2.210 ($1.1*10^{-3}$)	2.229 ($1.8*10^{-5}$)
PSICOV	2.502 ($3.7*10^{-1}$)	2.360 ($5.0*10^{-2}$)	2.648 ($1.2*10^{-1}$)	2.808 ($9.7*10^{-2}$)
CCMpred	3.167 ($3.1*10^{-19}$)	3.096 ($2.6*10^{-22}$)	3.357 ($2.1*10^{-16}$)	3.609 ($8.6*10^{-19}$)
FreeContact	3.052 ($1.3*10^{-13}$)	3.035 ($2.4*10^{-20}$)	3.435 ($1.6*10^{-17}$)	3.556 ($2.8*10^{-17}$)
STRUCTCH	2.480 ($3.2*10^{-1}$)	2.290 ($2.4*10^{-3}$)	2.747 ($4.9*10^{-5}$)	2.662 ($2.4*10^{-1}$)
MetaPSICOV	2.503 ($1.5*10^{-1}$)	2.106 ($3.5*10^{-1}$)	2.362 ($5.0*10^{-2}$)	2.444 ($2.1*10^{-2}$)
NeBcon	2.458	2.127	2.468	2.596
Entropy calculated by top-L contacts on **15×15** ***cells***				
BETACON	3.710 ($1.0*10^{-1}$)	2.436 ($2.3*10^{-14}$)	2.880 ($9.0*10^{-12}$)	3.720 ($1.1*10^{-16}$)
SVMSEQ	3.830 ($6.9*10^{-4}$)	3.657 ($9.4*10^{-7}$)	3.946 ($1.0*10^{-2}$)	4.066 ($1.2*10^{-4}$)
SVMcon	3.620 ($2.5*10^{-7}$)	3.177 ($1.4*10^{-15}$)	3.651 ($3.6*10^{-10}$)	3.857 ($5.5*10^{-14}$)
PSICOV	3.335 ($1.6*10^{-2}$)	3.488 ($4.9*10^{-1}$)	3.911 ($2.3*10^{-1}$)	4.107 ($2.5*10^{-1}$)
CCMpred	4.147 ($5.6*10^{-6}$)	4.411 ($7.1*10^{-16}$)	4.952 ($3.3*10^{-12}$)	5.208 ($4.4*10^{-13}$)
FreeContact	4.072 ($1.4*10^{-4}$)	4.371 ($4.6*10^{-15}$)	4.952 ($5.9*10^{-12}$)	5.115 ($3.4*10^{-11}$)
STRUCTCH	3.734 ($4.0*10^{-1}$)	3.556 ($4.3*10^{-2}$)	4.109 ($2.6*10^{-1}$)	4.204 ($8.4*10^{-2}$)
MetaPSICOV	3.866 ($4.3*10^{-3}$)	3.624 ($2.7*10^{-3}$)	4.001 ($1.1*10^{-1}$)	4.193 ($1.1*10^{-1}$)
NeBcon	3.743	3.493	4.079	4.266
Entropy calculated by top-L contacts on **5×5** ***cells***				
BETACON	2.764 ($6.3*10^{-3}$)	1.709 ($7.0*10^{-15}$)	1.931 ($6.2*10^{-10}$)	2.688 ($4.6*10^{-5}$)
SVMSEQ	2.815 ($3.4*10^{-1}$)	2.426 ($3.3*10^{-1}$)	2.524 ($4.7*10^{-2}$)	2.830 ($1.6*10^{-1}$)
SVMcon	2.709 ($1.3*10^{-9}$)	2.271 ($2.7*10^{-9}$)	2.347 ($1.6*10^{-6}$)	2.764 ($7.8*10^{-3}$)
PSICOV	2.351 ($4.1*10^{-4}$)	2.238 ($5.2*10^{-2}$)	2.368 ($4.0*10^{-2}$)	2.728 ($1.6*10^{-1}$)
CCMpred	2.924 ($3.7*10^{-2}$)	2.810 ($5.7*10^{-9}$)	2.997 ($4.6*10^{-9}$)	3.448 ($1.8*10^{-11}$)
FreeContact	2.870 ($1.7*10^{-1}$)	2.800 ($6.6*10^{-9}$)	3.080 ($2.5*10^{-11}$)	3.482 ($3.9*10^{-12}$)
STRUCTCH	2.745 ($3.1*10^{-4}$)	2.434 ($4.2*10^{-1}$)	2.539 ($2.4*10^{-2}$)	2.900 ($3.7*10^{-1}$)
MetaPSICOV	2.851 ($1.0*10^{-1}$)	2.525 ($1.1*10^{-2}$)	2.523 ($1.9*10^{-2}$)	2.965 ($4.3*10^{-2}$)
NeBcon	2.807	2.440	2.610	2.887

Table S5: Shannon entropy of top x contact maps generated by different methods using 10×10 cells, where x is the number of native contacts for each protein. Values in parentheses are p-values in student t-test relative to the entropy for native structures.

Methods	**Short**	**Medium**	**Long**	**All**
BETACON	2.705 ($2.2*10^{-9}$)	1.953 ($5.2*10^{-18}$)	2.656 ($8.4*10^{-16}$)	3.912 ($6.9*10^{-25}$)
SVMSEQ	2.680 ($5.6*10^{-15}$)	2.523 ($2.0*10^{-8}$)	3.540 ($4.9*10^{-7}$)	4.146 ($5.6*10^{-13}$)
SVMcon	2.589 ($5.7*10^{-16}$)	2.402 ($5.2*10^{-13}$)	3.289 ($1.5*10^{-16}$)	3.962 ($1.2*10^{-24}$)
PSICOV	2.676 ($2.6*10^{-2}$)	2.726 ($2.6*10^{-1}$)	3.505 ($6.2*10^{-2}$)	3.959 ($1.23*10^{-2}$)
CCMpred	3.377 ($1.3*10^{-7}$)	3.472 ($8.3*10^{-13}$)	4.415 ($6.9*10^{-9}$)	5.016 ($1.1*10^{-6}$)
FreeContact	3.245 ($2.0*10^{-4}$)	3.426 ($2.1*10^{-11}$)	4.478 ($4.5*10^{-10}$)	4.977 ($5.0*10^{-6}$)
STRUCTCH	2.723 ($1.5*10^{-9}$)	2.647 ($3.0*10^{-4}$)	3.477 ($2.6*10^{-8}$)	4.072 ($7.7*10^{-17}$)
MetaPSICOV	2.958 ($3.8*10^{-1}$)	2.709 ($1.4*10^{-2}$)	3.552 ($4.0*10^{-5}$)	4.217 ($9.7*10^{-6}$)
NeBcon	2.750 ($5.6*10^{-10}$)	2.570 ($7.8*10^{-10}$)	3.665 ($6.5*10^{-5}$)	4.273 ($3.3*10^{-9}$)
Native	2.973	2.823	3.815	4.473

Table S6. Overlap of contact maps by NeBcon with individual predictors.*

Methods	f_{common}	f_{missed}	f_{new}
BETACON	0.110	0.044	0.386
SVMSEQ	0.128	0.048	0.368
SVMcon	0.124	0.047	0.372
PSICOV	0.147	0.050	0.238
CCMpred	0.167	0.070	0.218
FreeContact	0.147	0.061	0.237
STRUCTCH	0.180	0.059	0.316
MetaPSICOV	0.203	0.047	0.180

*$f_{common} = n_{common}/n_{nat}$, $f_{missed} = n_{missed}/n_{nat}$, and $f_{new} = n_{new}/n_{nat}$, where n_{nat} is the total number of native contacts, n_{common} is the number of true contacts that are predicted by both NeBcon and the component program, n_{missed} is the number of true contacts that are predicted by the component program but missed by NeBcon, and n_{new} is the number of true contacts newly predicted by NeBcon but not predicted by the individual component. Only top-L long-range contact predictions are counted for each program.

Table S7. The average accuracy of the top *L*/5 long-range contact predictions by NeBcon on the CASP targets in control with the top ten contact predictors in CASP10 and CASP11. Values in parenthesis are p-value in student's t-test between NeBcon and the control predictors.

Method		**Ave. Acc (FM)**
CASP10 (20 FM domains)	NeBcon	0.4659
	MULTICOM	0.4058 (2.5×10^{-1})
	Distill_roll	0.2804 (7.5×10^{-3})
	Distill	0.2448 (3.5×10^{-3})
	MULTICOM-CONST.	0.2252 (4.0×10^{-4})
	IGBteam	0.2038 (1.3×10^{-4})
	SAM-T08-server	0.1924 (1.8×10^{-4})
	MetaPSICOV	0.1721 (4.5×10^{-5})
	RaptorX-Roll	0.1573 (6.2×10^{-5})
	MULTICOM-NOVEL	0.1514 (7.6×10^{-6})
	ZHOU-SPARKS-X	0.0864 (3.0×10^{-6})
CASP11 (33 FM domains)	NeBcon	0.3763
	MetaPSICOV	0.3632 (3.9×10^{-1})
	Pcons-net	0.2482 (1.0×10^{-3})
	Shen-group	0.2330 (4.5×10^{-3})
	UCI-IGB-CMpro	0.2199 (4.8×10^{-3})
	RBO_Aleph	0.1990 (2.9×10^{-3})
	LEE	0.1988 (1.1×10^{-3})
	MULTICOM-CLUST.	0.1831 (4.1×10^{-4})
	RaptorX-Contact	0.1605 (7.2×10^{-5})
	MULTICOM-CONST.	0.1559 (4.7×10^{-5})
	Distill	0.0782 (2.7×10^{-6})

Table S8. Comparison of NeBcon with two other meta-predictor (PconsC2 and PconsC31) that are not included in NeBcon program. Values in parentheses are p-values in student t-test relative to NeBcon.

Method	CASP12 dataset (FM)	
	Top L/5	Top L/2
NeBcon	0.378	0.319
PconsC2	0.328 (0.135)	0.262 (0.020)
PconsC31	0.378 (0.277)	0.314 (0.332)

Text S1. Illustrative examples of how the pairwise NBC combination improves the accuracy of individual predictors

As an illustration of the observations in Table 1, we present in Figure S1 a combination from two complementary methods, PSICOV and BETACON, on the Easy targets; but different from the data in Table 1 that collects top *L/5* predictions, we display the data from the top-*L* predictions in Figure S1. For 43 out of 50 targets, the accuracy of the combination is improved compared to BETACON, which results in an average accuracy increase from 0.195 by BETACON to 0.305 by PSICOV-BETACON. When we convert the combination from complementary to non-complementary method pairs, i.e., replacing PSICOV by SVMcon, the resultant accuracy is obviously dropped with only 30 out of 50 targets having the accuracy improved. Nevertheless, the average accuracy of all the combinations is higher than the individual predictors.

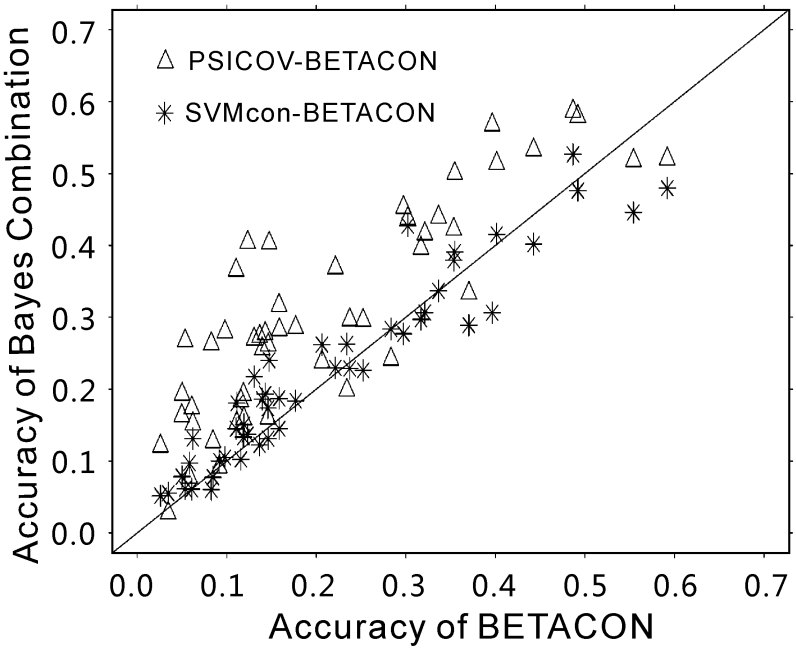

Fig. S1. The accuracies of predicted top *L* contact maps by BETACON versus that by the pair-wise NBC combination from PSICOV-BETACON and SVMcon-BETACON, respectively. Data are shown for the 50 Easy targets.

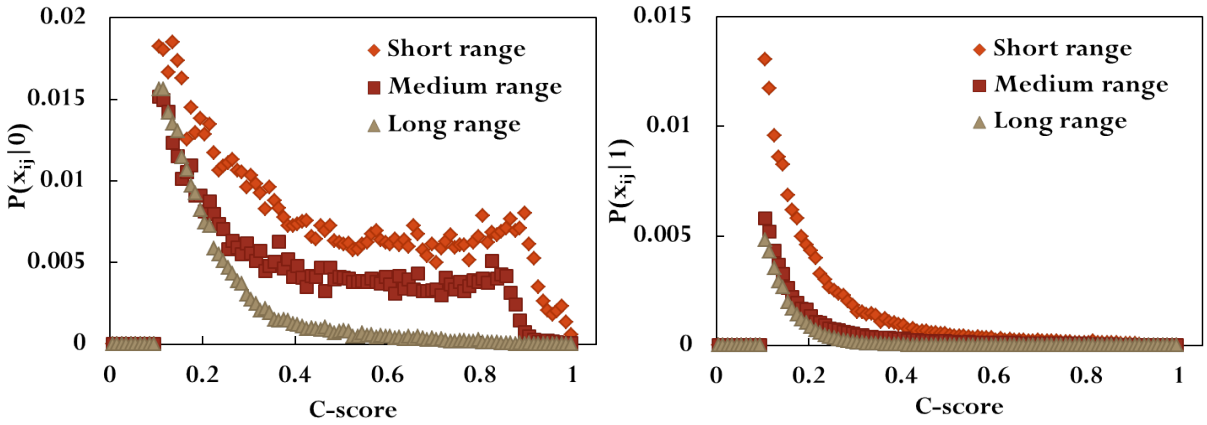

Figure S2: The conditional probability of $P(X_{ij}|0)$ and $P(X_{ij}|1)$ at short-, medium- and long-range contacts at different confidence score (C-score) for BETACON.

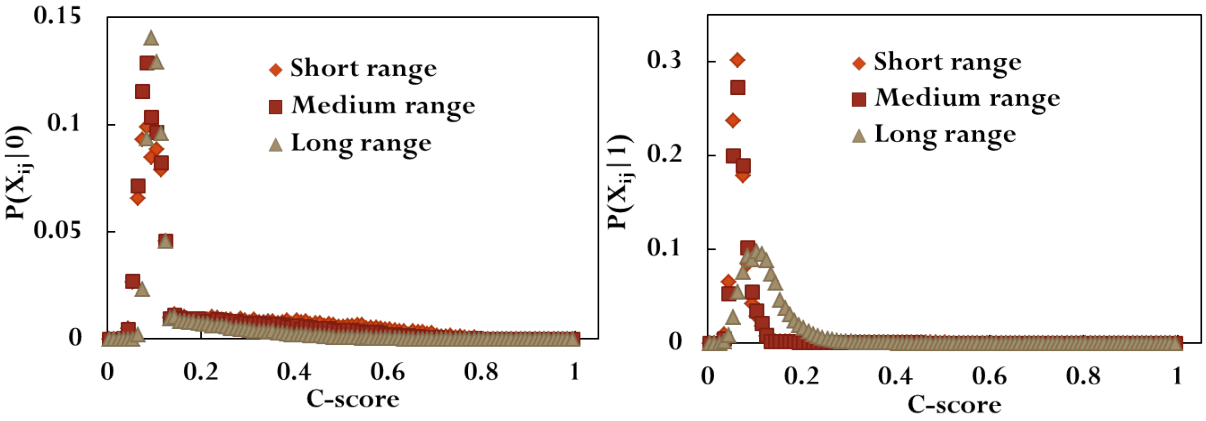

Figure S3: The conditional probability of $P(X_{ij}|0)$ and $P(X_{ij}|1)$ at short-, medium- and long-range contacts at different confidence score (C-score) for SVMSEQ.

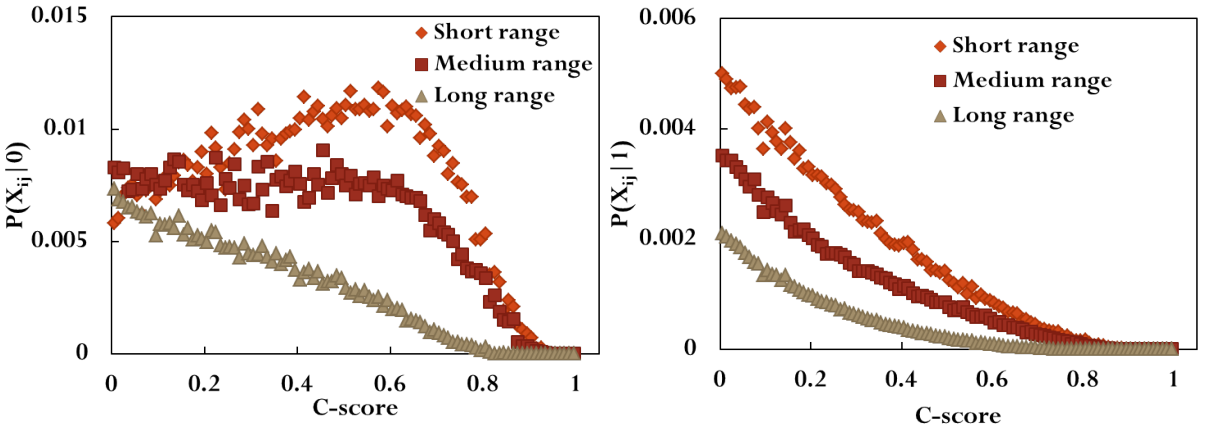

Figure S4: The conditional probability of $P(X_{ij}|0)$ and $P(X_{ij}|1)$ at short-, medium- and long-range contacts at different confidence score (C-score) for SVMcon.

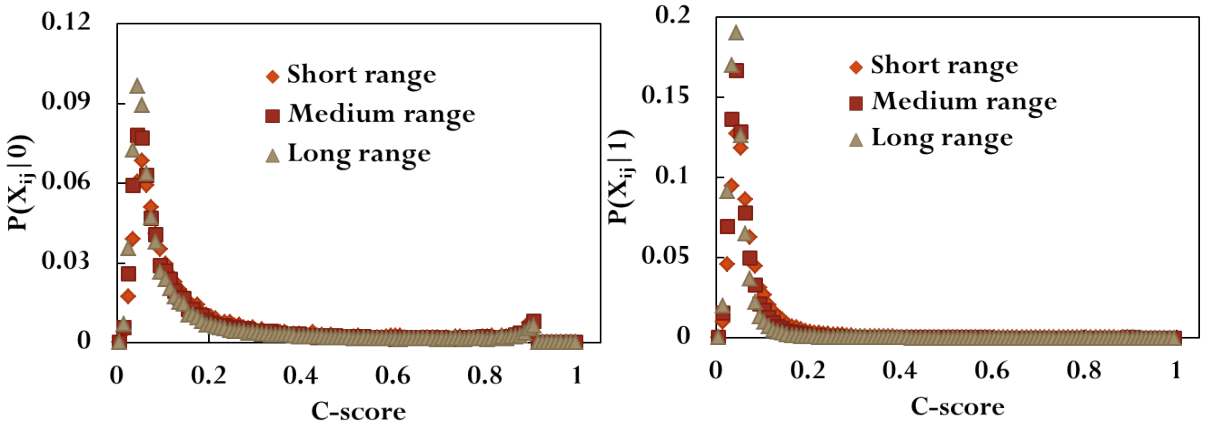

Figure S5: The conditional probability of $P(X_{ij}|0)$ and $P(X_{ij}|1)$ at short-, medium- and long-range contacts at different confidence score (C-score) for PSICOV.

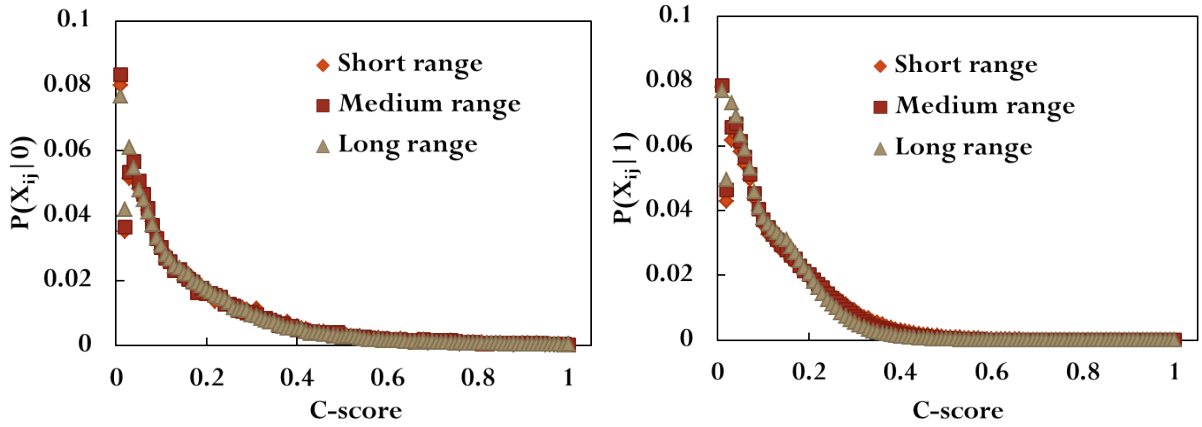

Figure S6: The conditional probability of $P(X_{ij}|0)$ and $P(X_{ij}|1)$ at short-, medium- and long-range contacts at different confidence score (C-score) for CCMpred.

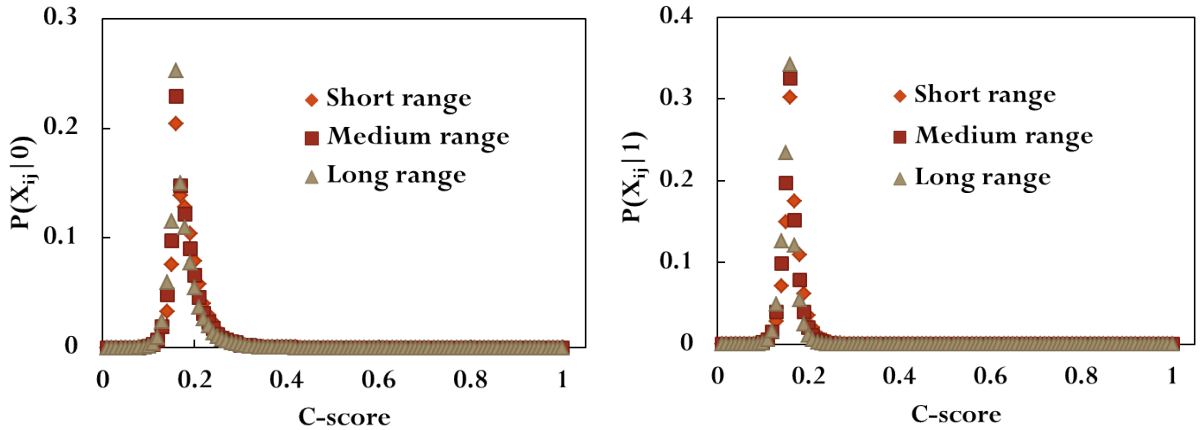

Figure S7: The conditional probability of $P(X_{ij}|0)$ and $P(X_{ij}|1)$ at short-, medium- and long-range contacts at different confidence score (C-score) for FreeContact.

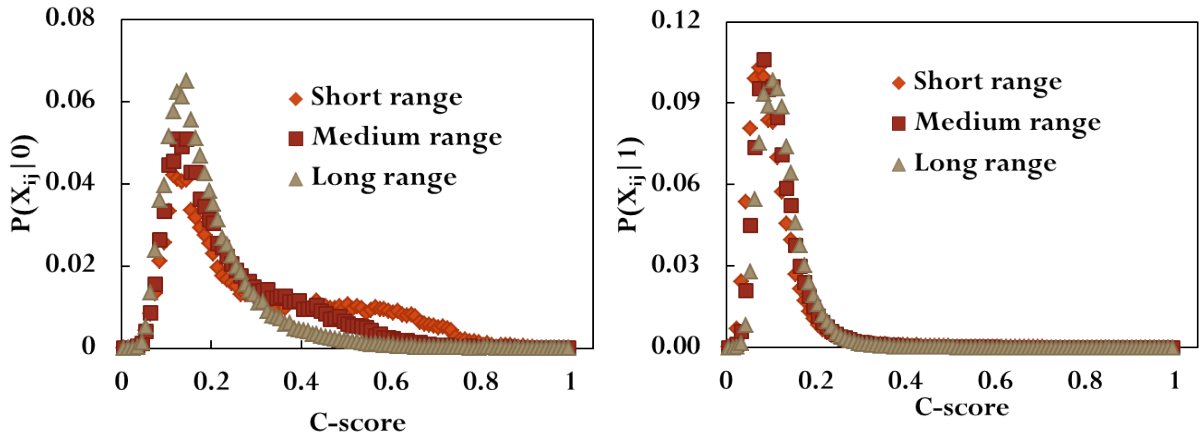

Figure S8: The conditional probability of $P(X_{ij}|0)$ and $P(X_{ij}|1)$ at short-, medium- and long-range contacts at different confidence score (C-score) for STRUCTCH.

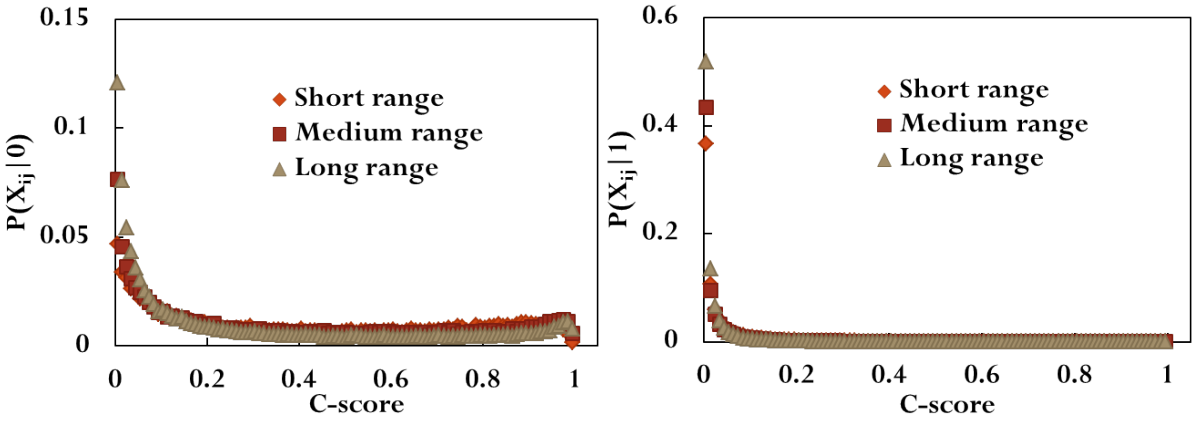

Figure S9: The conditional probability of $P(X_{ij}|0)$ and $P(X_{ij}|1)$ at short-, medium- and long-range contacts at different confidence score (C-score) for MetaPSICOV.

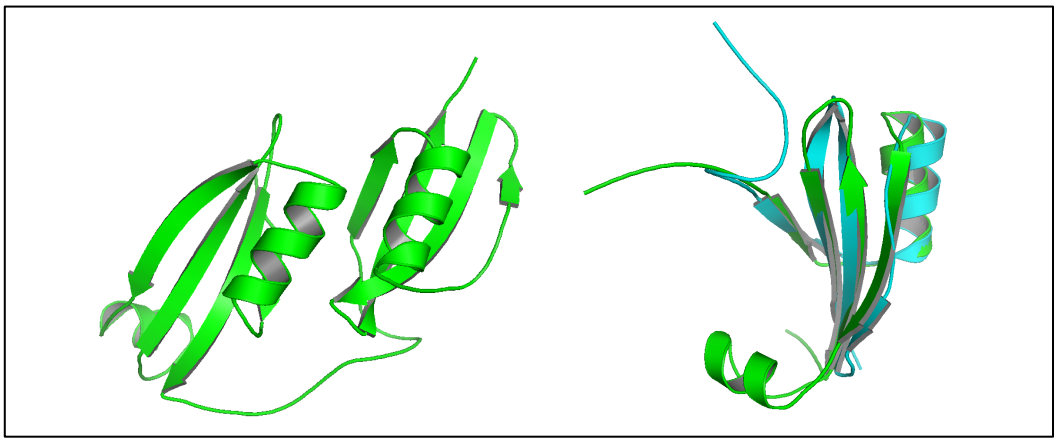

Figure S10: (A) Structure of 1ss4B, which comprises two domains (I and II). (B) The TM-score of superimposed domains is 0.64, indicating structural similarity between the domains.