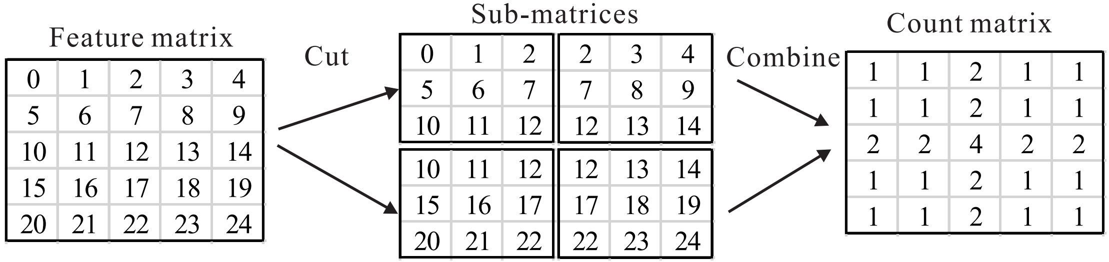
Supplementary Materials



0	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19
20	21	22	23	24

0	1	2
5	6	7
10	11	12

2	3	4
7	8	9
12	13	14

10	11	12
15	16	17
20	21	22

12	13	14
17	18	19
22	23	24

1	1	2	1	1
1	1	2	1	1
2	2	4	2	2
1	1	2	1	1
1	1	2	1	1

Figure S1. Demonstration for the cut-and-combine procedure in the min-batch training of the DeepMeta method. The numbers in the feature matrix and the sub-matrices indicate the identities of elements. The numbers in the count matrix indicate the appearing frequencies of the corresponding elements in the sub-matrices.

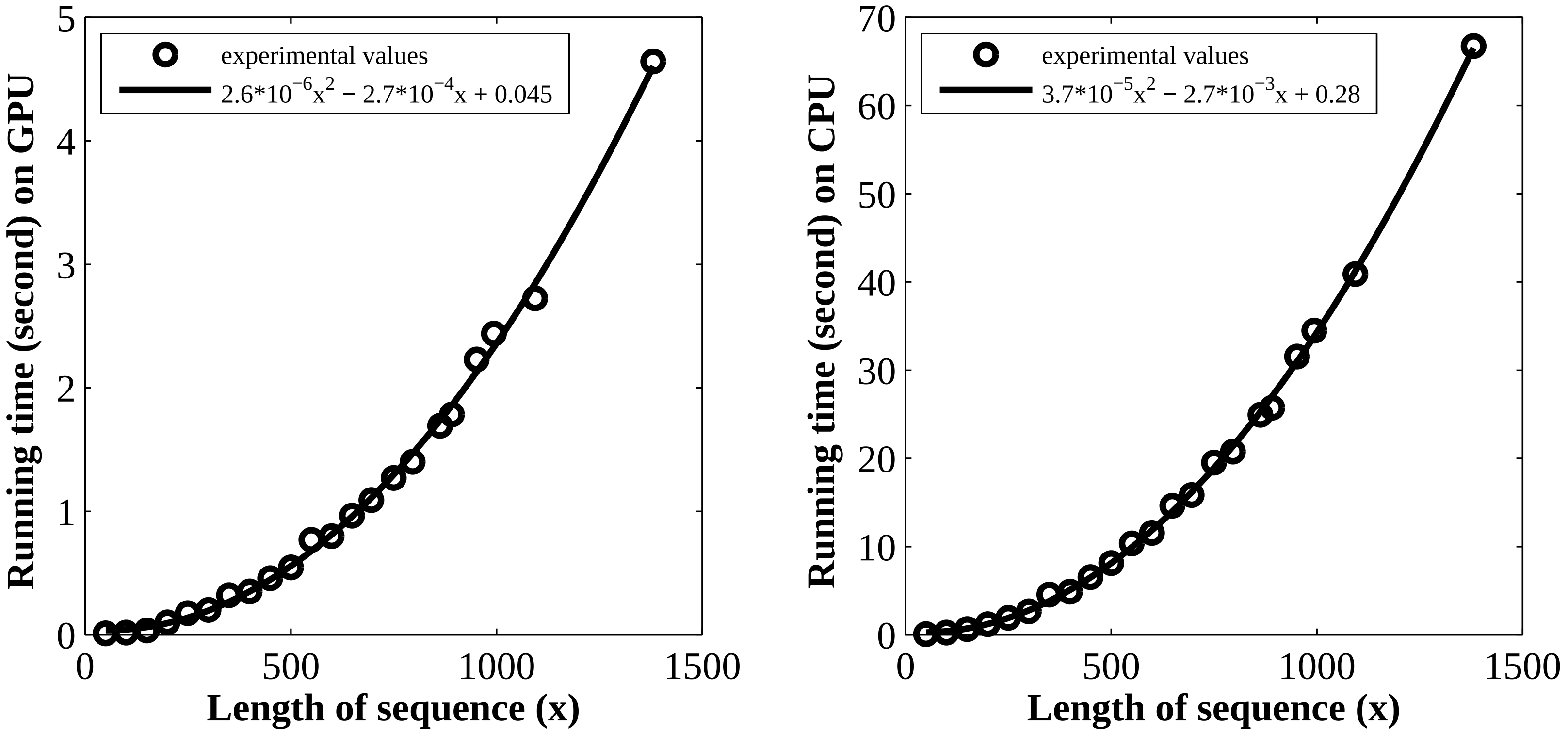


Figure S2. Running time of contact predictions for sequences with different lengths.

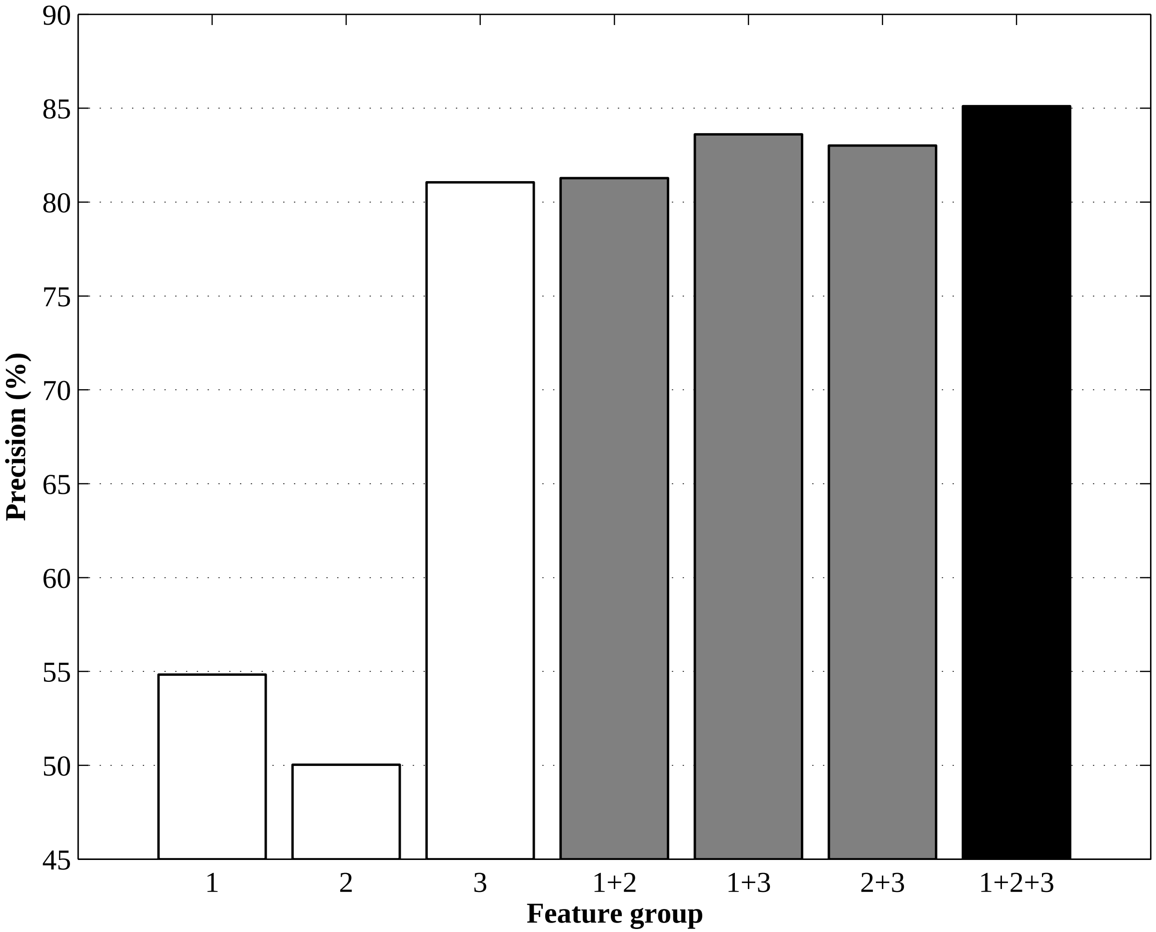


Figure S3. The precision of the predicted contact maps for the validation set with different feature groups. (1) CCMpred-based feature; (2) Sequence profile features, including SS, RSA and the Frobenius norm; and (3) DeepMSA-based feature.

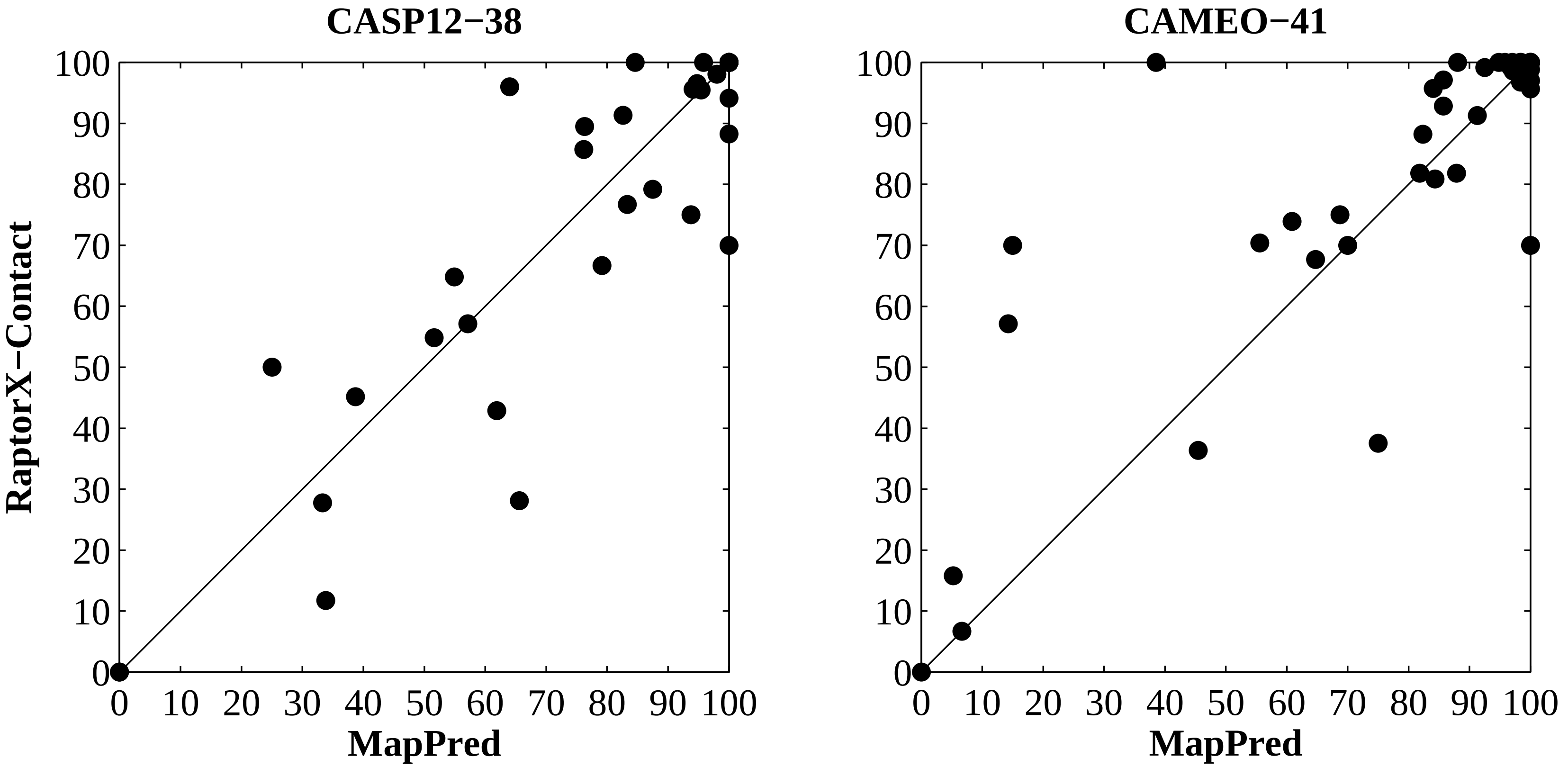
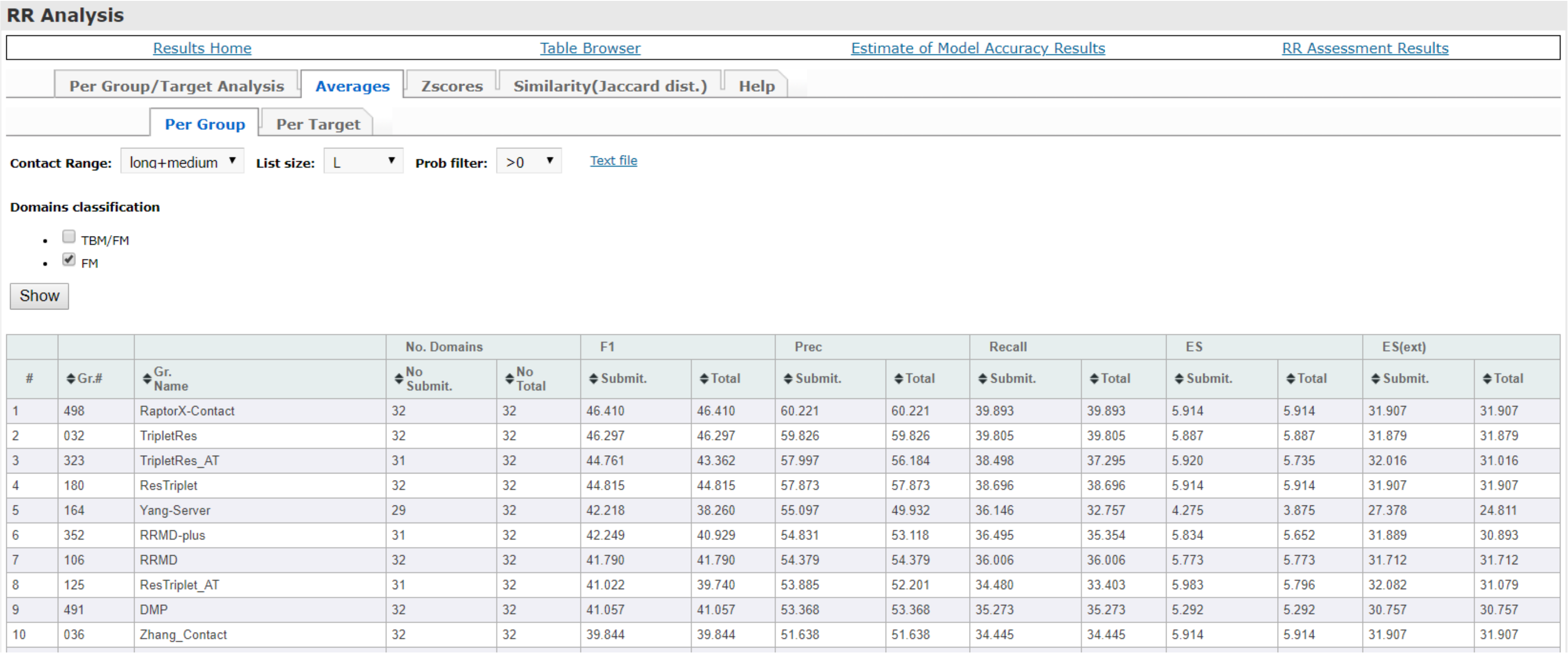


Figure S4. Head-to-head comparison between MapPred and RaptorX-Contact's top L/5 long-range precisions on the datasets CASP12-38 and CAMEO-41.



Contact Range: long+medium **List size:** L **Prob filter:** >0

Domains classification

- ☐ TBM/FM
- ☑ FM

			No. Domains		F1		Prec		Recall		ES		ES(ext)	
#	Gr.#	Gr. Name	No Submit.	No Total	Submit.	Total	Submit.	Total	Submit.	Total	Submit.	Total	Submit.	Total
1	498	RaptorX-Contact	32	32	46.410	46.410	60.221	60.221	39.893	39.893	5.914	5.914	31.907	31.907
2	032	TripletRes	32	32	46.297	46.297	59.826	59.826	39.805	39.805	5.887	5.887	31.879	31.879
3	323	TripletRes_AT	31	32	44.761	43.362	57.997	56.184	38.498	37.295	5.920	5.735	32.016	31.016
4	180	ResTriplet	32	32	44.815	44.815	57.873	57.873	38.696	38.696	5.914	5.914	31.907	31.907
5	164	Yang-Server	29	32	42.218	38.260	55.097	49.932	36.146	32.757	4.275	3.875	27.378	24.811
6	352	RRMD-plus	31	32	42.249	40.929	54.831	53.118	36.495	35.354	5.834	5.652	31.889	30.893
7	106	RRMD	32	32	41.790	41.790	54.379	54.379	36.006	36.006	5.773	5.773	31.712	31.712
8	125	ResTriplet_AT	31	32	41.022	39.740	53.885	52.201	34.480	33.403	5.983	5.796	32.082	31.079
9	491	DMP	32	32	41.057	41.057	53.368	53.368	35.273	35.273	5.292	5.292	30.757	30.757
10	036	Zhang_Contact	32	32	39.844	39.844	51.638	51.638	34.445	34.445	5.914	5.914	31.907	31.907

Figure S5. The screenshot of the CASP13 website: http://predictioncenter.org/casp13/rrc_avrg_results.cgi, which is for the ranking of the **top *L* long+medium-range** predictions.

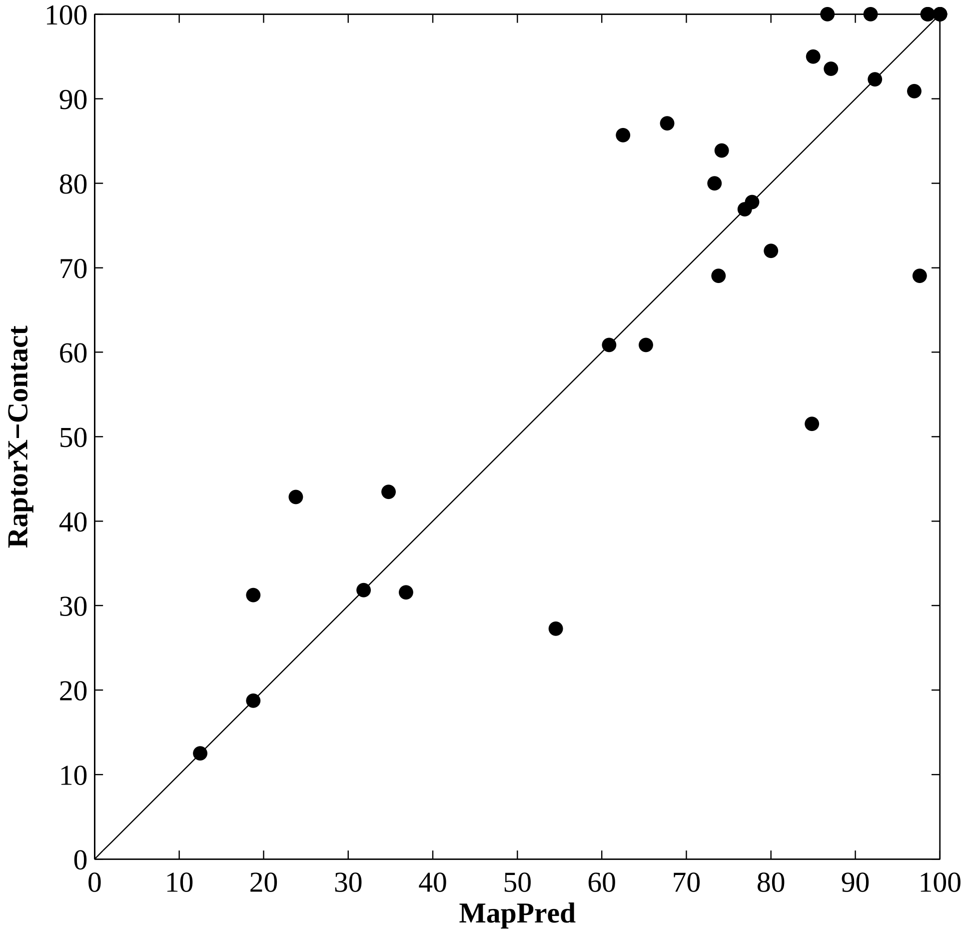


Figure S6. Head-to-head comparison between MapPred and RaptorX-Contact's top L/5 long-range precisions on the 31 FM domains from CASP13.

Table S1. The long-range precisions on the validation set for several neural networks in DeepMSA. Note that a single model was used to save computing time. The respective network architectures (A-C) are shown in the Figure below.

Network	$L/10$	$L/5$	$L/2$	L
CNN (A)	0.795	0.731	0.5864	0.439
Common ResNet (B)	0.825	0.769	0.635	0.483
Dilated ResNet (C)	0.831	0.773	0.636	0.484

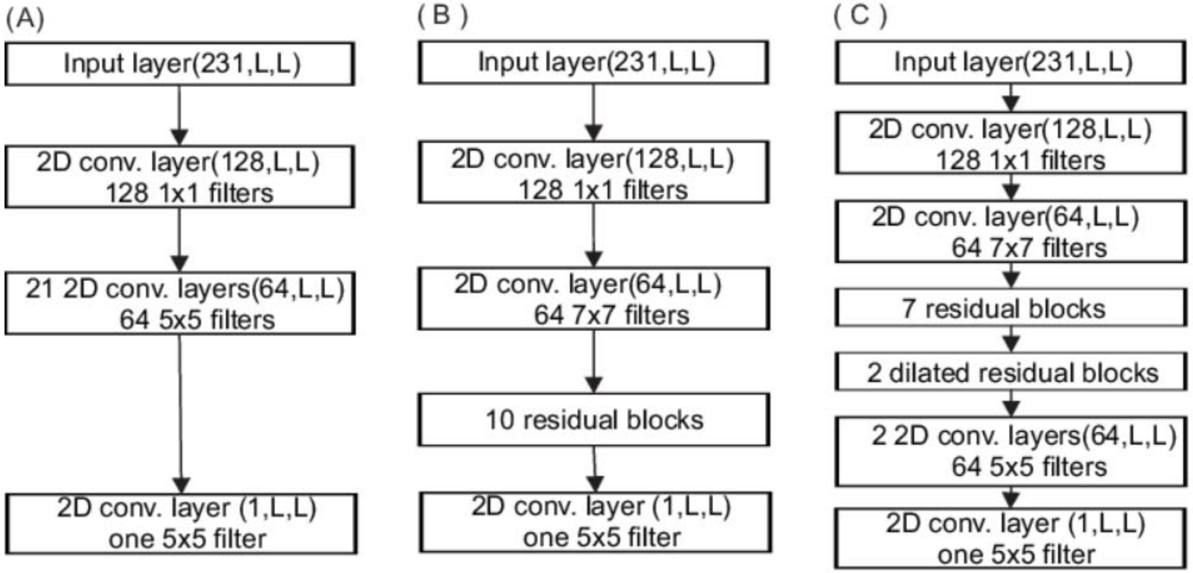


Table S2. The long-range precisions of DeepMSA on the validation set for several ways of using the original 441 channels. Note that a single model was used and each model was trained on 2000 sequences randomly selected from training set, to save training time and computer memory.

No	L/10	L/5	L/2	L	#Channels	Remark	Symmetric
1	0.767	0.696	0.549	0.404	441	Original	No
2	0.761	0.693	0.545	0.403	231	Eq. (2), used in this work	Yes
3	0.748	0.677	0.525	0.388	441	Eq. (S1)	Yes
4	0.752	0.682	0.536	0.397	441	Eq. (S2)	Yes
5	0.749	0.679	0.530	0.390	231	Select a subset of channels with $x \le y$ by the alphabeta order of amino acids	No
6	0.728	0.658	0.511	0.372	231	Select a subset of channels with $x \ge y$ by the alphabeta order of amino acids	No

$$\mathrm{Rcov}(i_x, j_y) = \begin{cases} \mathrm{cov}(i_x, j_y), & \text{if } i \ge j, \\ \mathrm{cov}(i_y, j_x), & \text{if } i < j, \end{cases} \qquad \text{(S1)}$$

$$\mathrm{Rcov}(i_x, j_y) = \begin{cases} \mathrm{cov}(i_x, j_y), & \text{if } i \le j, \\ \mathrm{cov}(i_y, j_x), & \text{if } i > j, \end{cases} \qquad \text{(S2)}$$

Table S3. The precisions for CCMpred and DeepMSA on the three benchmark datasets with MSAs generated from the Uniclast30 and the MetaDB databases. The data in the last row are the P-values are for the statistical tests between the data listed in the third and the fourth rows.

Database	CCMpred			DeepMSA		
	SPOT-228	CAMEO-41	CASP12-38	SPOT-228	CAMEO-41	CASP12-38
Uniclust30	0.3792	0.3956	0.2437	0.6809	0.6374	0.5617
MetaDB	0.469	0.4585	0.3822	0.7546	0.7323	0.7133
P-value	2.83E-77	4.04E-17	3.52E-39	2.89E-63	4.20E-31	4.18E-47