Supplementary information for improved protein model quality assessment by integrating sequential and pairwise features using deep learning

Xiaoyang Jing, Jinbo Xu

Toyota Technological Institute at Chicago, Chicago, IL 60637, USA

Table S1. The number of protein targets and decoys of different datasets.

Dataset	Data Source	#Targets		#Decoys	
Training	CASP 7-11	495	13916	107060	335468
	CAMEO	1011		38444	
	CATH	12410		189964	
Validation	CASP 7-11	32	729	6688	18437
	CAMEO	53		2010	
	CATH	644		9739	
Test	CASP12	64		9423	
	CASP13	76		11371	

Table S2. Summary of input features

Type	Feature	Type	Shape
Derived from Sequence	One-hot encoding of sequence	Sequential	L*21
	rPosition	Sequential	L*1
	PSSM	Sequential	L*20
	Predicted SS3	Sequential	L*3
	Predicted ACC	Sequential	L*3
	Co-evolution	Pairwise	L*L*4
	Predicted distance potential	Pairwise	L*L*14
Derived from model structure	SS3	Sequential	L*3
	RSA	Sequential	L*1
	distance (CaCa, CbCb and NO)	Pairwise	L*L*3
Total			Sequential: L*52 Pairwise: L*L*21

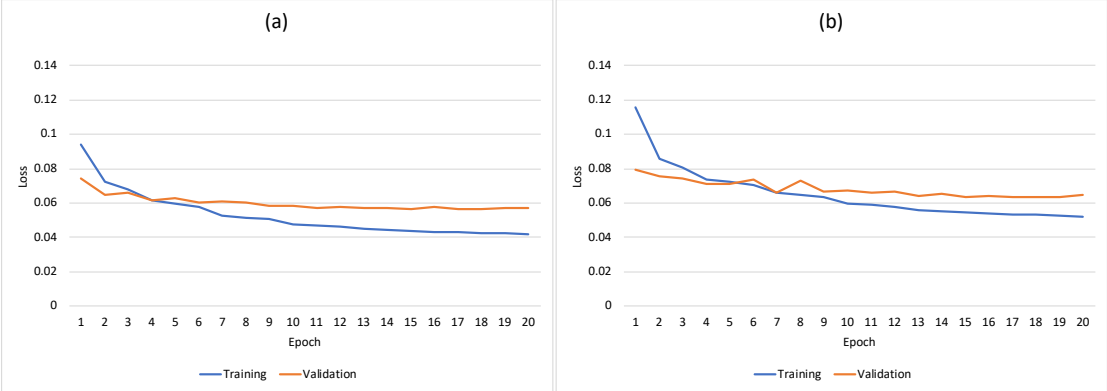

Fig. S1. The training and validation losses of ResNetQA and ResNetQA-R. (a) the training and validation losses of ResNetQA, the mean and standard deviation for the training loss are 0.0527 and 0.0128, and those for the validation loss are 0.0599 and 0.0043, respectively; (b) the training and validation losses of ResNetQA-R, the mean and standard deviation for the training loss are 0.0650 and 0.0151, and those for the validation loss are 0.0681 and 0.0047, respectively. The training loss has a larger standard deviation because we employed minibatch-based training. The gap between the validation and the training loss is also likely incurred by minibatch training. The training loss of one epoch is the averaged loss of all the minibatches in this epoch. The loss of one minibatch is calculated when it is being used to update the model parameters. At each iteration, we minimize the loss on one minibatch and thus, the model parameters are biased towards the proteins in this minibatch. As such, it is not surprising that the loss calculated on this minibatch is usually smaller than the validation loss, which is always calculated on the same set of many more proteins.

Table S3. Z-scores[1] of single-model methods on local (S-score) and global (GDT_TS) QA

Dataset	Method	Local			Global				
		PCC	ASE	AUC	PCC	Spearman	Kendall	Diff	Loss
CASP12 Stage 2	**ResNetQA[2]**	0.9650	**0.8942**	**0.8890**	0.6529	0.5925	0.6343	**0.7620**	0.2905
	ResNetQA-R[3]	**0.9773**	0.8371	0.8687	**0.7619**	**0.6729**	**0.7288**	0.4259	0.4160
	ProQ3	0.2064	0.0729	0.2397	0.0972	0.1410	0.0665	0.2141	**0.4298**
	ProQ2	0.1571	-0.3423	0.1917	-0.0442	0.0175	-0.0586	-0.0752	0.2943
	Wang4	0.0649	0.0387	0.0622	0.0119	-0.0075	-0.0557	-0.0190	-0.1084
	VoroMQA	0.0589	-0.2259	0.0568	-0.0694	-0.0618	-0.1359	-0.5319	0.1066
	SPARKS-X	-0.0534	0.0225	-0.0569	0.1855	0.2426	0.2071	-0.1386	0.2248
CASP13 Stage 2	**ResNetQA**	0.6380	**0.8094**	0.5682	0.6317	0.3979	0.3933	**0.7656**	-0.1158
	ResNetQA-R	**0.7007**	0.7330	**0.6325**	**0.7335**	**0.5695**	**0.5955**	0.4081	0.1211
	ProQ3D	0.3005	0.3126	0.3390	-0.0979	-0.1302	-0.2012	0.4285	**0.1229**
	ProQ3	0.2840	0.2433	0.3737	-0.4067	-0.3973	-0.4546	0.2327	0.0871
	ProQ4	0.2532	-0.3786	0.2671	0.1932	0.1183	0.1561	0.1427	0.0669
	VoroMQA-A	0.2293	-0.0446	0.2749	-0.1434	-0.2366	-0.3040	-0.1397	0.0326
	VoroMQA-B	0.2206	-0.0501	0.2623	-0.2758	-0.3162	-0.3786	-0.1659	0.1198

1. To calculate the Z-score for each target, we collect the quality estimations of all groups participating in CASP12 and CASP13, evaluate their performances by local and global QA metrics, respectively, and then calculate the mean and standard deviation of the scores of all groups for each metric, which are then used to calculate Z-score. The Z-score of our methods (ResNetQA and ResNetQA-R) are calculated using the same mean and standard deviation as other methods. Note that the CASP assessors may set a Z-score to -2 or 0 if it is smaller than -2 or 0, respectively, but here we do not do that.
2. The model trained using local and global MSE loss.
3. The model trained using local and global MSE loss plus global margin ranking loss.

Table S4. Performances of single-model methods on global QA (lDDT)

Dataset	Method	Global				
		PCC↑	Spearman↑	Kendall↑	Diff↓	Loss↓
CASP12 Stage 2	**ResNetQA**	0.8298	0.7586	0.5919	0.1035	0.0515
	ResNetQA-R	**0.8527**	**0.7860**	**0.6186**	**0.0712**	0.0469
	ProQ3	0.7239	0.6760	0.5025	0.1283	**0.0382**
	ProQ2	0.7131	0.6746	0.4999	0.1032	0.0448
	Wang4	0.7130	0.6432	0.4822	0.1672	0.0828
	VoroMQA	0.6986	0.6501	0.4798	0.1957	0.0530
	SPARKS-X	0.7754	0.7163	0.5466	0.1691	0.0471
CASP13 Stage 2	**ResNetQA**	**0.8673**	0.7918	0.6272	0.1127	0.0707
	ResNetQA-R	0.8661	**0.8096**	**0.6487**	**0.0821**	**0.0541**
	ProQ3D	0.7246	0.6849	0.5156	0.1470	0.0675
	ProQ3	0.6711	0.6418	0.4761	0.1511	0.0711
	ProQ4	0.8125	0.7636	0.6120	0.1309	0.0597
	VoroMQA-A	0.6968	0.6560	0.4909	0.1929	0.0838
	VoroMQA-B	0.6676	0.6315	0.4687	0.1945	0.0796

Table S5. Performances of single-model methods on local and global QA (lDDT)

Dataset	Method	Local			Global				
		PCC↑	ASE↑	AUC↑	PCC↑	Spearman↑	Kendall↑	Diff↓	Loss↓
CASP12 Stage 2	**ResNetQA-lDDT[1]**	0.7178	0.9032	0.8709	0.8632	0.7972	0.6402	**0.0645**	0.0484
	ResNetQA-R-lDDT[2]	**0.7235**	**0.9039**	**0.8747**	**0.8935**	**0.8265**	**0.6688**	0.0726	0.0393
	ProQ4	0.5899	0.8844	0.8051	0.8350	0.7822	0.6085	0.0764	**0.0308**
CASP13 Stage 2	**ResNetQA-lDDT**	0.7186	0.8994	0.8859	0.8434	0.8011	0.6471	**0.0698**	0.0641
	ResNetQA-R-lDDT	**0.7318**	**0.9012**	**0.8913**	**0.8719**	**0.8267**	**0.6722**	0.0719	0.0693
	ProQ4	0.6038	0.8799	0.8222	0.8266	0.7902	0.6207	0.0818	**0.0536**

1. The model trained using local and global MSE (mean-squared error) loss based on lDDT.
2. The model trained using local and global MSE loss plus global margin ranking loss based on lDDT.

Table S6. Z-scores of single-model methods on CASP12&13 FM targets (S-score and GDT_TS)

Dataset	Method	Local			Global				
		PCC	ASE	AUC	PCC	Spearman	Kendall	Diff	Loss
CASP12 Stage 2	**ResNetQA**	1.3887	**1.2560**	**1.2853**	0.8417	0.9632	1.0463	**0.7803**	0.4747
	ResNetQA-R	**1.4393**	1.1927	1.2708	**1.0490**	**1.0936**	**1.2041**	0.2849	0.5350
	ProQ3	0.1910	-0.0364	0.1698	0.1536	0.1278	0.0720	0.2546	**0.6030**
	ProQ2	0.2323	-0.5391	0.1909	0.0936	0.0791	0.0366	-0.3403	0.2560
	Wang4	0.1222	0.2101	0.0707	0.2330	0.3010	0.2819	0.5007	0.1269
	VoroMQA	0.1570	-0.2644	0.0920	-0.0413	-0.1599	-0.2139	0.3391	0.0168
	SPARKS-X	-0.0195	0.6476	-0.0758	0.3237	0.4362	0.4200	0.4192	0.2335
CASP13 Stage 2	**ResNetQA**	0.7084	**1.0135**	0.5505	0.9249	0.7465	0.7526	**0.8024**	0.1904
	ResNetQA-R	**0.8612**	0.9589	**0.7229**	**1.1230**	**1.0214**	**1.0852**	0.1785	**0.3093**
	ProQ3D	0.2075	0.3199	0.2987	-0.1303	-0.2248	-0.2529	0.5391	-0.1015
	ProQ3	0.2205	0.2413	0.3568	-0.4350	-0.5446	-0.5481	0.2366	-0.1511
	ProQ4	0.2492	-0.5419	0.2292	0.1495	0.1027	0.1040	0.1390	0.2304
	VoroMQA-A	0.1772	-0.1309	0.2225	-0.3291	-0.3572	-0.4173	0.3736	0.1182
	VoroMQA-B	0.1608	-0.1259	0.2010	-0.4215	-0.4458	-0.4873	0.3508	0.1982

Table S7. Performances of single-model methods on CASP12&13 FM targets (global lDDT)

Dataset	Method	Global				
		PCC↑	Spearman↑	Kendall↑	Diff↓	Loss↓
CASP12 Stage 2	**ResNetQA**	0.7291	0.7390	0.5732	0.1103	0.0434
	ResNetQA-R	**0.7827**	**0.7845**	**0.6143**	**0.0589**	**0.0380**
	ProQ3	0.6405	0.6215	0.4524	0.1149	0.0433
	ProQ2	0.6781	0.6603	0.4867	0.0941	0.0512
	Wang4	0.7016	0.6900	0.5207	0.1242	0.0603
	VoroMQA	0.6399	0.5981	0.4331	0.1082	0.0612
	SPARKS-X	0.7352	0.7225	0.5483	0.1263	0.0478
CASP13 Stage 2	**ResNetQA**	0.8438	0.8175	0.6437	0.1205	0.0625
	ResNetQA-R	**0.8599**	**0.8490**	**0.6874**	**0.0587**	**0.0543**
	ProQ3D	0.6524	0.6410	0.4711	0.1358	0.0994
	ProQ3	0.5871	0.5772	0.4151	0.1466	0.1036
	ProQ4	0.7613	0.7533	0.5964	0.0734	0.0552
	VoroMQA-A	0.5803	0.5728	0.4105	0.1133	0.0908
	VoroMQA-B	0.5637	0.5530	0.3966	0.1151	0.0866

Table S8. Performances of single-model methods on CASP12&13 FM targets (local and global lDDT)

Dataset	Method	Local			Global				
		PCC↑	ASE↑	AUC↑	PCC↑	Spearman↑	Kendall↑	Diff↓	Loss↓
CASP12 Stage 2	**ResNetQA-lDDT**	0.6692	0.9149	0.8333	0.8328	0.8297	0.6711	**0.0597**	0.0412
	ResNetQA-R-lDDT	**0.6802**	**0.9154**	**0.8381**	**0.8642**	**0.8544**	**0.6987**	0.0728	**0.0285**
	ProQ4	0.5020	0.8957	0.7425	0.7708	0.7429	0.5657	0.0566	0.0375
CASP13 Stage 2	**ResNetQA-lDDT**	0.6688	0.9051	0.8469	0.8513	0.8618	0.7055	0.0661	**0.0432**
	ResNetQA-R-lDDT	**0.6929**	**0.9128**	**0.8548**	**0.8666**	**0.8710**	**0.7228**	0.0679	0.0492
	ProQ4	0.5088	0.8872	0.7506	0.7747	0.7641	0.5883	**0.0619**	0.0461

Table S9. Performances of deep ResNet built with different features on all targets[1]

Dataset	Method	Local			Global				
		PCC↑	ASE↑	AUC↑	PCC↑	Spearman↑	Kendall↑	Diff↓	Loss↓
CASP12 Stage 2	**All**[2]	0.5866	**0.8515**	0.8058	**0.8109**	**0.7293**	**0.5606**	7.85	**6.12**
	No Cov[3]	**0.5874**	0.8507	**0.8063**	0.8011	0.7135	0.5459	**7.58**	7.30
	No SS&RSA[4]	0.5643	0.8302	0.7996	0.7788	0.6876	0.5247	8.77	7.33
	No DistPot[5]	0.5319	0.8243	0.7804	0.7192	0.6233	0.4654	9.57	7.76
	1D-DistPot[6]	0.2735	0.6965	0.6412	0.4046	0.3510	0.2444	16.46	11.92
	1D[7]	0.2619	0.6612	0.6347	0.3887	0.3333	0.2322	17.93	14.39
CASP13 Stage 2	**All**	**0.5539**	**0.8373**	**0.7901**	**0.8157**	**0.7396**	**0.5726**	**8.61**	8.44
	No Cov	0.5474	0.8248	0.7889	0.8020	0.7226	0.5551	8.97	**8.29**
	No SS&RSA	0.5416	0.8057	0.7840	0.7771	0.7042	0.5361	9.69	10.68
	No DistPot	0.5000	0.7971	0.7749	0.7450	0.6784	0.5071	10.57	9.99
	1D-DistPot	0.2963	0.6981	0.6618	0.5119	0.4694	0.3336	15.24	16.36
	1D	0.2630	0.6711	0.6441	0.5100	0.4595	0.3261	15.74	14.00

1. The proteins in the training data for ablation studies are filtered out only by BLAST based on sequence identity.
2. All features are used.
3. Excluding coevolution information including mutual information and the information produced by CCMPred.
4. Excluding predicted and model-derived secondary structures and solvent accessibilities.
5. Excluding predicted distance potentials, but coevolution information is used.
6. Using sequential features plus the marginalized distance potentials predicted by RaptorX.
7. Only sequential features are used.

Table S10. Performances of deep ResNet built with different features on FM targets

Dataset	Method	Local			Global				
		PCC↑	ASE↑	AUC↑	PCC↑	Spearman↑	Kendall↑	Diff↓	Loss↓
CASP12 Stage 2	**All**	0.4545	**0.8930**	0.7487	**0.7213**	**0.7270**	**0.5535**	**5.84**	**5.49**
	No Cov	**0.4590**	0.8868	**0.7544**	0.7188	0.7221	0.5491	6.28	6.35
	No SS&RSA	0.4207	0.8647	0.7404	0.6599	0.6857	0.5204	6.87	5.54
	No DistPot	0.3633	0.8505	0.7044	0.5636	0.5599	0.4126	8.71	8.99
	1D-DistPot	0.1983	0.6771	0.6005	0.2919	0.2656	0.1818	18.27	13.51
	1D	0.2025	0.6063	0.6095	0.2978	0.2728	0.1858	21.00	15.73
CASP13 Stage 2	**All**	**0.4045**	**0.8718**	**0.7308**	**0.7978**	**0.7871**	**0.6122**	**7.22**	9.14
	No Cov	0.3997	0.8477	0.7322	0.7766	0.7728	0.5970	8.53	**6.91**
	No SS&RSA	0.3790	0.8237	0.7213	0.7580	0.7482	0.5690	8.60	10.45
	No DistPot	0.3495	0.8199	0.7156	0.6783	0.6611	0.4911	10.13	8.04
	1D-DistPot	0.1673	0.6563	0.5972	0.3253	0.3371	0.2341	18.42	18.87
	1D	0.1527	0.6059	0.5964	0.3359	0.3445	0.2403	20.04	19.06

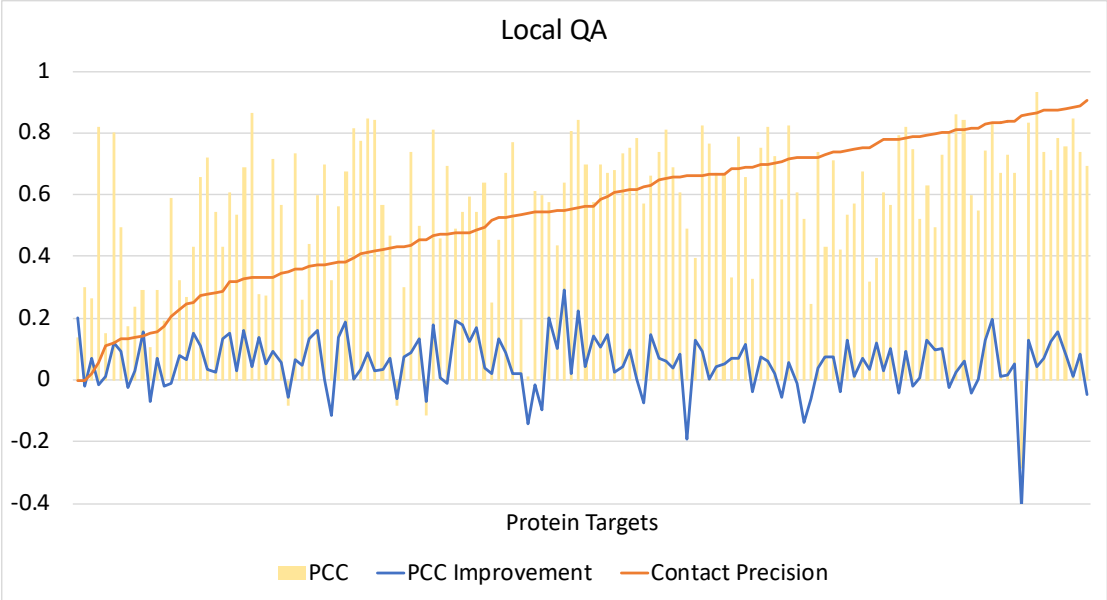

Fig. S2. The correlation between the distance prediction quality and the local QA performance on CASP12 and CASP13 protein targets. Contact Precision: the Top L (L is the sequence length) long-range contact precision of the distance prediction by RaptorX-Contact; PCC: the local QA PCC performance of our method when all features used; PCC Improvement: the improvement on local QA PCC resulting from using predicted distance potential as one feature. The data are sorted from low to high according to contact precision. The correlation between the distance prediction quality and the PCC improvement is -0.1081. That is, the improvement is more pronounced on harder targets (which usually have lower contact precision).

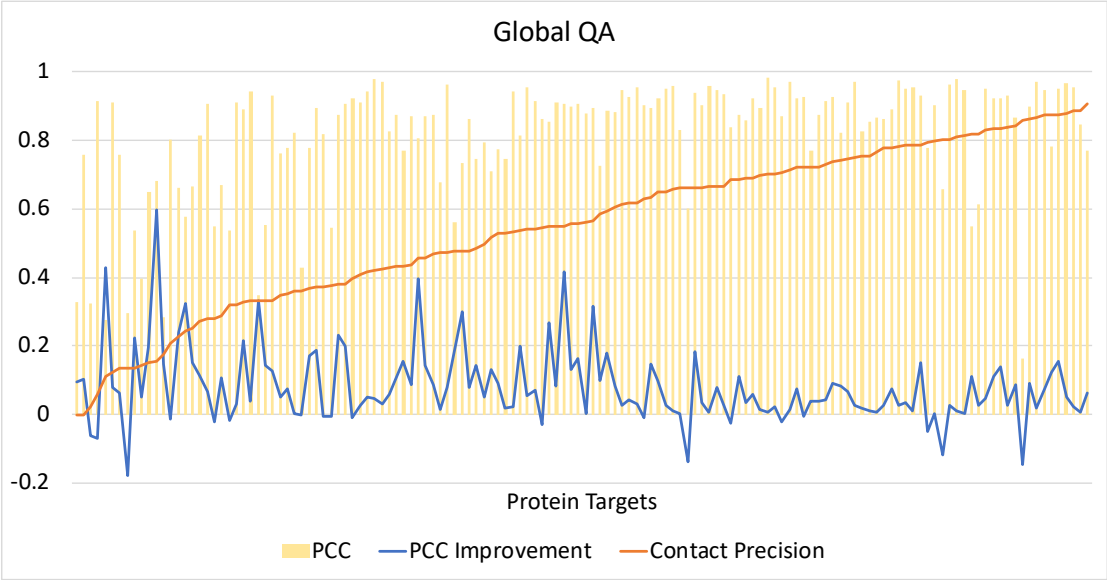

Fig. S3. The correlation between the distance prediction quality and the global QA performance on CASP12 and CASP13 protein targets. Contact Precision: the Top L (L is the sequence length) long-range contact precision of the distance prediction by RaptorX-Contact; PCC: the global QA PCC performance of the model using all features; PCC Improvement: the improvement on global QA PCC resulting from using predicted distance potential as one feature. The data are sorted from low to high according to contact precision. The correlation between the distance prediction quality and the PCC improvement is -0.2783. That is, the improvement is more pronounced on harder targets (which usually have lower contact precision).

Table S11. Performances of deep models trained with and without the CATH dataset

Dataset	Method	Local			Global				
		PCC↑	ASE↑	AUC↑	PCC↑	Spearman↑	Kendall↑	Diff↓	Loss↓
CASP12 Stage 2	ResNetQA	**0.5866**	**0.8515**	**0.8058**	**0.8109**	**0.7293**	**0.5606**	**7.85**	**6.12**
	NoCATH-Pot[1]	0.5450	0.8425	0.7870	0.7520	0.6736	0.5094	9.96	8.93
	NoCATH-PR[2]	0.5375	0.8318	0.7886	0.7563	0.6914	0.5259	10.36	7.02
CASP13 Stage 2	ResNetQA	**0.5539**	**0.8373**	**0.7901**	**0.8157**	**0.7396**	**0.5726**	**8.61**	**8.44**
	NoCATH-Pot	0.5276	0.8134	0.7890	0.7873	0.7217	0.5518	12.11	9.47
	NoCATH-PR	0.5321	0.8037	0.7862	0.7752	0.7005	0.5318	12.14	9.64

1. NoCATH-Pot: predicted distance potential is an input feature. NoCATH means that the CATH data is not used in training.
2. NoCATH-PR: predicted distance probability is an input feature.