Received: 29 April 2019 | Revised: 24 July 2019 | Accepted: 27 August 2019

DOI: 10.1002/prot.25810

RESEARCH ARTICLE

PROTEINS WILEY

Analysis of distance-based protein structure prediction by deep learning in CASP13

Jinbo Xu | Sheng Wang

Toyota Technological Institute at Chicago, Chicago, Illinois

Correspondence
Jinbo Xu, Toyota Technological Institute at Chicago, 6045 S Kenwood Avenue, Chicago, IL 60637.
Email: jinbo.xu@gmail.com

Funding information
National Institute of General Medical Sciences, Grant/Award Number: R01GM089753; National Sciences Foundation, Grant/Award Number: DBI-1564955

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1002/prot.25810.

Abstract

This paper reports the CASP13 results of distance-based contact prediction, threading, and folding methods implemented in three RaptorX servers, which are built upon the powerful deep convolutional residual neural network (ResNet) method initiated by us for contact prediction in CASP12. On the 32 CASP13 FM (free-modeling) targets with a median multiple sequence alignment (MSA) depth of 36, RaptorX yielded the best contact prediction among 46 groups and almost the best 3D structure modeling among all server groups without time-consuming conformation sampling. In particular, RaptorX achieved top $L/5$, $L/2$, and L long-range contact precision of 70%, 58%, and 45%, respectively, and predicted correct folds (TMscore > 0.5) for 18 of 32 targets. Further, RaptorX predicted correct folds for all FM targets with >300 residues (T0950-D1, T0969-D1, and T1000-D2) and generated the best 3D models for T0950-D1 and T0969-D1 among all groups. This CASP13 test confirms our previous findings: (a) predicted distance is more useful than contacts for both template-based and free modeling; and (b) structure modeling may be improved by integrating template and coevolutionary information via deep learning. This paper will discuss progress we have made since CASP12, the strength and weakness of our methods, and why deep learning performed much better in CASP13.

KEYWORDS
critical assessment of structure prediction, coevolution analysis, deep convolutional residual neural network, multiple sequence alignment, protein contact and distance prediction, protein folding

1 | INTRODUCTION

Significant progress has been achieved on protein structure prediction due to the development of two major ideas: (a) direct coupling analysis (DCA) for coevolution analysis[1-5]; and (b) very deep and fully convolutional residual neural network (ResNet) for protein contact and distance prediction.[6,7] DCA may recover a small set of long-range native contacts when the protein under study has a large number of sequence homologs. In contrast, deep ResNet not only works very well on proteins without many sequence homologs, but also can directly predict interresidue or interatom distance.[8,9]

In CASP12 and previous CAMEO tests we have demonstrated that deep ResNet can greatly improve contact prediction[6,10-12] and that without extensive conformation sampling, contacts predicted by deep ResNet can be used to build correct folds for (even membrane) proteins without detectable homology in PDB.[13] Afterwards, the power of deep convolutional neural network has been further validated by other research groups who have developed similar deep networks for contact prediction.[14-16] Although contact prediction itself is an important problem that needs further study, our focus switched to distance prediction and accordingly distance-based protein structure modeling. This is because a distance matrix contains finer-grained information than a contact matrix and provides more physical constraints of a protein structure, for example, distance is metric while contact is not. That is, a distance matrix can determine a protein structure much more accurately than a contact matrix. Trained by distance

instead of contact matrices, ResNet may automatically learn more about the intrinsic properties of a protein structure and thus, greatly reduce the conformation space, improve folding accuracy, and shorten running time needed for protein folding.

Although not totally new, there were only few studies on protein distance prediction and their accuracy were not very satisfactory.[17-20] Since 2012, we have employed a probabilistic neural network to predict interresidue distance distribution from sequence profile and mutual information and then from predicted distribution derived protein-specific distance-based potential for decoy ranking,[21] remote homology detection,[22] and folding simulation.[23] In these studies, we have shown that protein-specific distance potential derived from machine learning performs well on decoy ranking and remote homology detection and comparably with other methods in folding simulation.

Prior to CASP13, we have extended our deep ResNet to protein distance prediction and showed that distance-based potential predicted by ResNet may significantly improve protein threading for targets without good templates in PDB.[8] We have developed a simple and efficient distance geometry algorithm that may quickly fold a protein sequence from distance and torsion angles predicted by deep ResNet.[9] Our deep ResNet not only can predict distance matrix from sequence and coevolutionary information, but also from template and alignment information. In this paper we describe our methods for distance prediction and distance-based protein threading and folding, analyze their performance in CASP13 and discuss their strengths and weaknesses. We will also examine a few specific targets and highlight our views on the future development and challenge.

2 | MATERIALS AND METHODS

2.1 | Deep dilated ResNet for protein distance and contact prediction

We use a very similar deep ResNet as described in our previous paper[6] to predict the Euclidean distance distribution of two atoms (of different residues) in a protein to be folded. Our ResNet model consists of one 1D deep ResNet, one 2D deep dilated ResNet, and one Softmax layer (Figure 1). The 1D and 2D ResNets capture long-range sequential and pairwise context, respectively. The 2D ResNet used a dilated instead of the traditional convolutional operation[24] to yield slightly better accuracy with fewer model parameters. The 1D and 2D ResNets use ~7 and ~60 convolutional layers, respectively, and kernel size of 15 and 5 × 5.

We discretize interatom distance into 25 bins: <4.5, 4.5-5, 5-5.5, ..., 15-15.5, 15.5-16, and >16 Å and treat each bin as a label for classification. The ResNet model for distance prediction is trained using the same procedure as before.[6] Contact prediction is fulfilled by summing up the probability of all the C_β-$C_\beta$distance bins falling into interval [0, 8 Å]. Our distance-based contact prediction has 3%-4% better long-range precision than the ResNet model directly trained from contact matrices. Besides C_β-$C_\beta$distance distribution, we also trained individual ResNet models to predict distance distribution for other atom pairs: C_α-C_α, C_α-C_g, C_g-C_g, and N-O. Here C_g represents the first CG atom in an amino acid. When CG does not exist, OG or SG is used. The predicted distance of these five atom pairs is used together to fold a protein, which on average is slightly better than using the predicted C_β-$C_\beta$distance alone.

In addition to distance, we have employed a 1D deep ResNet of 19 convolutional layers to predict 3-state secondary structure and backbone torsion angles ϕ and ψ from position specific scoring matrix (PSSM) generated by HHblits.[25] To predict a distribution function for torsion angles, we train our ResNet model by maximizing the following probability function.

$$P(\phi,\psi\,|\,\bar{\phi},\bar{\psi},\sigma_1,\sigma_2,\rho)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}}\exp\left\{-\frac{1}{1-\rho^2}\left[\frac{1-\cos(\phi-\bar{\phi})}{\sigma_1^2}-\rho\frac{\sin(\phi-\bar{\phi})}{\sigma_1^2}\frac{\sin(\psi-\bar{\psi})}{\sigma_2^2}+\frac{1-\cos(\psi-\bar{\psi})}{\sigma_2^2}\right]\right\} \quad (1)$$

In Equation (1), $\bar{\phi},\bar{\psi}$ are the mean, σ_1, σ_2 are the variance, and ρ is the correlation. That is, our deep ResNet outputs the mean and variance of the torsion angles at each residue.

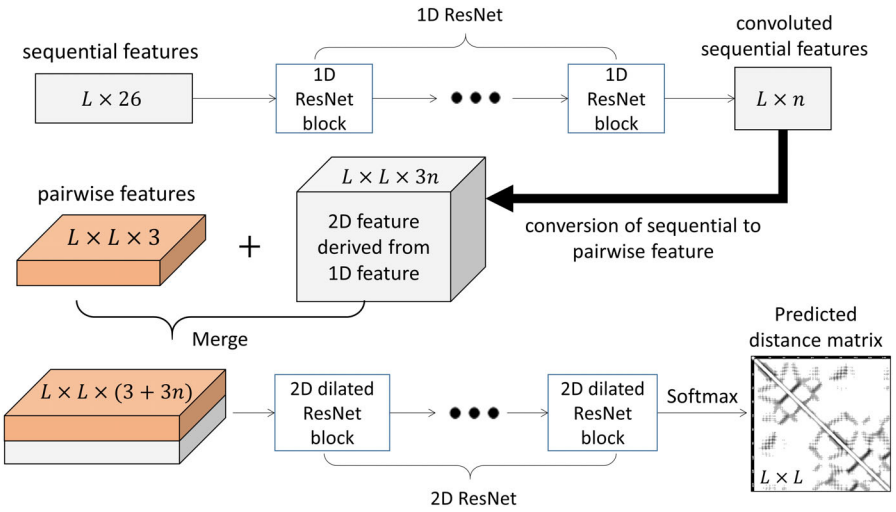

FIGURE 1 The overall architecture of deep dilated ResNet used by RaptorX servers in CASP13

2.2 | Multiple sequence alignment and input features

We generated four different multiple sequence alignments (MSAs) by running HHblits[25] with three iterations and *E*-value set to 0.001 and 1 and running Jackhmmer[26] with *E*-value set to 0.001 and 0.00001, respectively. HHblits and Jackhmmer search through the uniclust30 library released in October 2017 and the UniProt protein sequence database released in March 2018, respectively. We have not used metagenomics sequence database for MSA generation during CASP13. From an MSA, we derive protein sequential and pairwise features. Sequential features include sequence profile, secondary structure, and solvent accessibility predicted by RaptorX-Property.[27] Pairwise features include standard and APC-corrected mutual information, pairwise contact potential and coupling score calculated by CCMpred.[28] In summary, for each target we generated four sets of input features and accordingly four different distance predictions, which are then averaged to obtain the final prediction.

2.3 | Training data

We constructed our training and validation sets from PDB25 created early in 2018, which in total has 11 410 proteins. No two proteins in this set share more than 25% sequence identity. We randomly selected about 900 proteins to form the validation set and used the remaining to form the training set. We have trained three models for each atom pair, which are then combined to form the final model.

2.4 | RaptorX-TBM

RaptorX-TBM used our new threading program DeepThreader[22] to build sequence-template alignments and identify templates and then employed Rosetta-CM[29] to build 3D models from alignment. DeepThreader greatly outperforms previous threading methods by integrating CNFpred[30] with protein-specific distance potential predicted by our deep ResNet model. CNFpred is our old in-house threading program that aligns sequence to templates by integrating sequence profile, predicted secondary structure and solvent accessibility via conditional neural fields,[31] which is a combination of shallow convolutional neural network and linear-graph-based conditional random fields. CNFpred is on average more sensitive than HHpred,[32] but much worse than DeepThreader. RaptorX-TBM and CNFpred used PDB90 as the template database while our deep ResNet models for distance and angle prediction were trained by PDB25.

2.5 | RaptorX-Contact

In CASP13 we registered RaptorX-Contact for both contact prediction and distance-based template-free modeling. RaptorX-Contact converts predicted distance distribution, secondary structure, and backbone torsion angles into CNS restraints[33] and builds 3D models by running CNS, a software program for experimental protein structure determination. Given a matrix of $L \times L$ distance probability distributions (L is sequence length), we select 7L atom pairs with the highest predicted likelihood (probability) having distance <15 Å and assume they have distance <15 Å. From the predicted distance distribution of one atom pair, we estimate the mean distance m and SD s, and then use $m - s$ and $m + s$ as its distance lower and upper bounds. We use the mean degree and variance predicted by our 1D deep ResNet as torsion angle restraints and the same method as CONFOLD[34] to derive hydrogen-bond restraints from predicted alpha helices.

For each protein, we ran CNS to generate 200 possible 3D models and then choose five with the least violation of distance restraints as the final models. CNS uses distance geometry to build initial 3D models from distance restraints and then employs simulated annealing to refine bonds and angles so that the resultant models are protein-like. CNS can generate a 3D model very quickly. We generated multiple models for a protein since the CNS solution may not be globally optimal.

2.6 | RaptorX-DeepModeller

RaptorX-DeepModeller is also a distance-based folding server, differing from RaptorX-Contact in that the ResNet model used by RaptorX-DeepModeller has a few additional input features extracted from sequence-template alignment generated by RaptorX-TBM. The additional features include sequence-template similarity score (eg, amino acid similarity, sequence profile similarity, and secondary structure similarity) and an initial distance matrix extracted from the weakly similar template according to the alignment. Supposing two target residues i and j are aligned to two template residues k and l, we assign the distance between k and l as the initial distance of i and j. When one target residue is not aligned, the corresponding row and column in the initial distance matrix is empty.

RaptorX-DeepModeller is developed to study if we can improve protein structure modeling by integrating alignment, template and coevolutionary information through deep learning. Due to the limit of computing resources, RaptorX-DeepModeller used for most CASP13 targets was trained by the training set described in Wang et al.,[6] which is smaller than what was used for RaptorX-Contact in CASP13. Only at the end of CASP13, we had a new ResNet model for RaptorX-DeepModeller that was trained by ~10 000 proteins. To generate training alignments, we constructed a template database from PDB40 and then threaded each training protein to all the templates in PDB40. Afterwards, for each training protein, we randomly picked up one alignment from the top 30 list to form a set of training alignments, from which we constructed alignment-based input features for our ResNet model. To deal with gaps in the alignment, at each target residue we used one flag to indicate if this residue is covered by the randomly-selected template or not and in the case that it is not covered, we set the corresponding target-template similarity scores to 0.

3 | RESULTS

3.1 | Contact prediction accuracy in CASP13

RaptorX-Contact was officially ranked first among 46 human and server contact predictors, in terms of a combination of several metrics. When top $L/5$, $L/2$, and L long-range predicted contacts are

evaluated, on the FM targets RaptorX-Contact has precision 70.054%, 57.787%, and 44.731%, respectively, and F1 values 0.233, 0.362, and 0.411, respectively. As a control, MetaPSICOV[35] (the CASP11 winner built upon a shallow neural network) ran by the CASP13 organizers has top $L/5$ long-range precision = 25.16% and F1 = 0.078, respectively, and GaussDCA[36] (the only DCA method blindly tested in CASP13) has precision = 21.757% and F1 = 0.067, respectively. Figure 2A,B shows that there is certain, but not very strong correlation between our long-range F1 and precision with MSA depth (correlation coefficient <0.56, trendline $R^2 < 0.32$).

When a target has ln(MSA depth) <3 (ie, MSA depth <40), the top $L/2$ long-range contact precision generated by RaptorX is usually <40%. The only exception is T1010-D1, which has ln(MSA depth) around 2 and top $L/2$ long-range precision 57%. When ln(MSA depth) is between 3 and 4, RaptorX predicted long-range contacts with top $L/2$ precision >50% for 8 out of 11 targets. For most targets with ln(MSA depth) >4, RaptorX predicted long-range contacts with top $L/2$ precision>50%. The exceptions are T0968s2-D1, T0989-D1, and T0953s1. T0953s1 is a special target that does not have native long-range contacts. Although RaptorX did not have good contact precision for T0968s2-D1, RaptorX predicted a good 3D model with TMscore ~0.7, possibly because it generated a good distance prediction. RaptorX did not do well on T0989-D1 possibly because T0989 has two domains with very different MSA depth and RaptorX did not split T0989 into two domains and biased towards the 2nd one with a deeper MSA.

Note that for each target, we generated four different MSAs and thus, there may be four different MSA depths for a target. In this paper we use the depth of the MSA generated by HHblits with *E*-value = 0.001 and the uniclust30 library created in October 2017. This MSA depth is very similar to what is shown on the CASP13 contact assessment web page, and their correlation coefficient is ~0.955 and average absolute difference is ~0.374. Note that the MSA depth resulting from different homology search criteria may not be directly comparable. For example, when *E*-value = 1 is used with HHblits, often we may generated a deeper MSA, but the resultant contact prediction may not be better than when *E*-value = 0.001 is used.

AlphaFold did not submit contact prediction, according to its presentation at the seventh CAPRI meeting, it has similar F1 values as RaptorX-Contact on FM targets. On the top $L/5$, $L/2$, and L long-range contacts predicted for the FM targets, AlphaFold has F1 values 0.227, 0.369, and 0.419, respectively. That is, in terms of contact prediction performance there is almost no difference between the deep network architectures used by AlphaFold and RaptorX-Contact.

In addition to precision and F1, entropy score is introduced by the contact prediction assessors in CASP12 to measure the spread-out of predicted contacts. A contact prediction with a large F1 value may not result in good 3D structure modeling if the predicted contacts are mainly located in a small contact submatrix. As reported by CASP13, when top 10, $L/5$ and $L/2$ long-range contacts are considered, RaptorX-Contact has entropy score 0.311, 0.643, and 1.255, respectively, much larger than GaussDCA,[36] which has entropy score 0.151, 0.332, and 0.553, respectively. Even when only top 10 contacts are

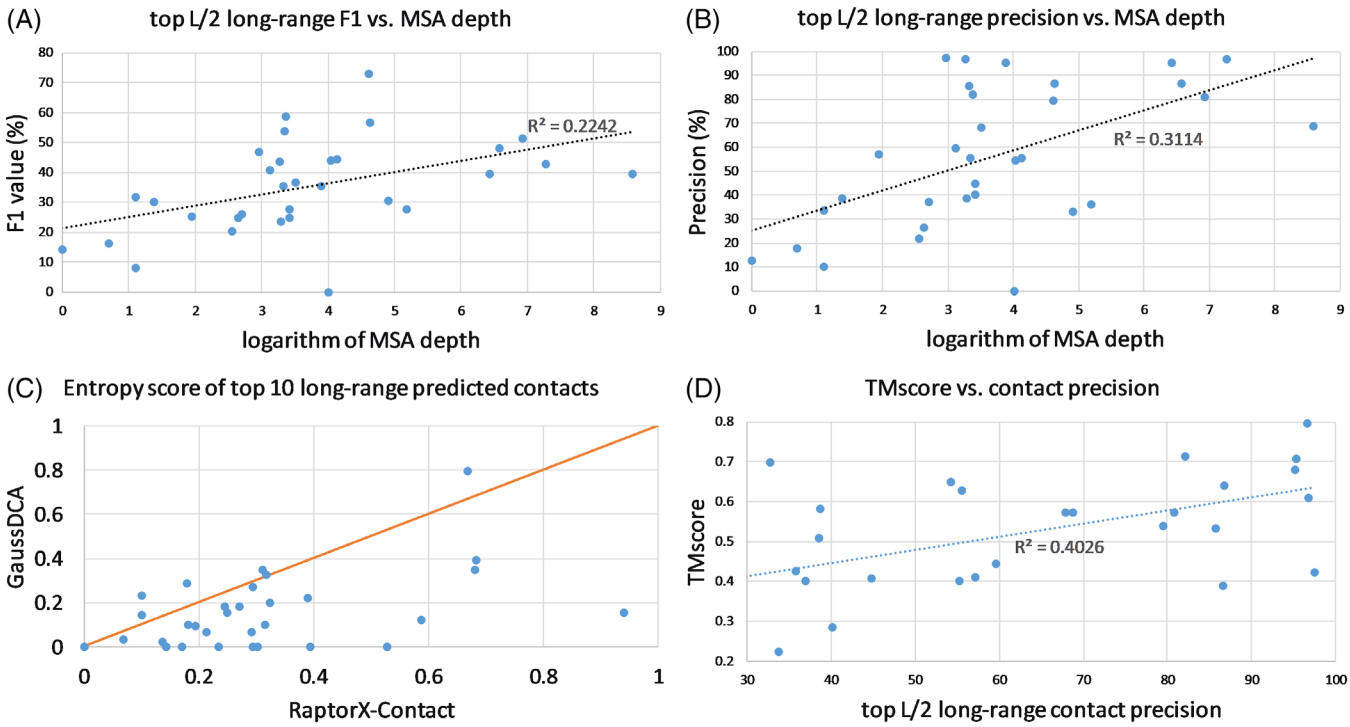

FIGURE 2 Contact prediction analysis. (A) relationship between contact prediction F1 value and MSA depth; (B) relationship between contact prediction precision and MSA depth; (C) entropy score of top 10 long-range contacts predicted by our method versus GaussDCA; (D) relationship between the quality of 3D models predicted by RaptorX-Contact and contact precision. MSA, multiple sequence alignment

considered, on most targets RaptorX-Contact has better entropy score than GaussDCA (Figure 2C). This result indicates that contacts predicted by deep ResNet may contain more information content for 3D structure modeling than contacts predicted by DCA.

3.2 | Distance prediction accuracy in CASP13

We have employed a few metrics to evaluate distance prediction accuracy, including absolute error, relative error, precision, recall, F1, pairwise distance test (PDT) and high-accuracy pairwise distance test (PHA). While evaluating distance prediction, we consider only those long- and/or medium-range atom pairs with predicted distance <15 Å. Absolute error is calculated as the absolute difference between predicted and native distance and relative error is the absolute error normalized by the average of predicted and native distance. Recall is calculated as the ratio of atom pairs with native distance <15 Å that are predicted to have distance <15 Å. Precision is calculated as the ratio of atom pairs with predicted distance <15 Å that have native distance <15 Å. PDT and PHA are analogous to GDT and GHA. To calculate them, we first calculate the fraction, denoted as $R(i)$, of predicted distance with absolute error less than i (i = 0.5, 1, 2, 4, and 8 Å). Then we calculate PDT as the average of $R(1)$, $R(2)$, $R(4)$, and $R(8)$ and PHA as the average of $R(0.5)$, $R(1)$, $R(2)$, and $R(4)$.

Table 1 lists the average distance prediction accuracy in terms of the above-mentioned metrics as well as their correlation with the quality (TMscore) of the 3D model produced by RaptorX-Contact. When long- and medium-range C_β-C_β distance prediction is considered, our prediction has average absolute error 3.76 Å, relative error 0.258, precision 0.678, recall 0.540, F1 0.588, PHA 0.405, and PDT 0.589. The average absolute prediction error for most targets is less than 4 Å (Figure 3C) and it is not strongly correlated with MSA depth (coefficient = −0.36, trendline R^2 = 0.129). On average, ~70% of predicted distance (<15 Å) has absolute prediction error no more than 4 Å. The correlation coefficients between absolute error, relative error, precision, recall, F1, PHA, and PDT and the quality (TMscore) of the 3D model are −0.674, −0.745, 0.657, 0.695, 0.852, 0.799, and 0.789, respectively. When only long-range C_β-C_β distance prediction is considered, the absolute error, relative error, precision, recall, F1, PHA and PDT are 4.05, 0.268, 0.638, 0.527, 0.555, 0.385, and 0.567 Å, respectively, and their correlation coefficients with TMscore of 3D model are −0.675, −0.716, 0.695, 0.673, 0.851, 0.781, and 0.766, respectively. In summary, F1 has the best correlation coefficient with the quality of 3D models, followed by PHA and PDT (Figure 3A). In contrast, the correlation between absolute distance prediction error and the quality of 3D models is not as strong (Figure 3D). Nevertheless, these distance metrics are more strongly correlated with the quality of 3D models than contact precision (Figure 2D). Figure 3B shows that F1, PDT, and PHA are not highly correlated with the MSA depth. Similar to contact prediction, entropy score shall be a good metric for distance prediction, but we have not calculated it since it will take some time to develop the code.

When ln(MSA depth) ≥ 4, RaptorX-Contact predicted distance ≤3 Å for most targets except T0953s1 and T0989-D1, both having MSA depth ~50. RaptorX did well not only on targets with a deep MSA (eg, T0969-D1 with MSA depth > 1000), but also on some targets with a very shallow MSA (eg, T0957s2-D1 with MSA depth 28). RaptorX-Contact did well on T0957s2-D1 possibly because it does not have many extra long-range distance<15 Å. T0953s1 has only

TABLE 1 The long- and/or medium-range distance prediction accuracy in terms of a variety of metrics

	Abs. error	Rel. error	Prec.	Recall	F1	R0.5	R1	R2	R4	R8	PHA	PDT
Long- and medium-range distance prediction accuracy												
CaCa	3.84	0.253	0.666	0.532	0.577	0.163	0.293	0.491	0.719	0.875	0.416	0.594
CaCg	3.84	0.262	0.671	0.512	0.567	0.143	0.272	0.472	0.709	0.8797	0.399	0.583
CbCb	3.76	0.258	0.678	0.540	0.588	0.149	0.278	0.480	0.713	0.8859	0.405	0.589
CgCg	4.02	0.278	0.656	0.532	0.573	0.127	0.244	0.439	0.684	0.8735	0.374	0.560
NO	3.75	0.253	0.674	0.505	0.566	0.165	0.296	0.499	0.722	0.877	0.421	0.598
CC-TM	−0.674	−0.745	0.657	0.695	0.852	0.739	0.769	0.804	0.774	0.668	0.799	0.789
CC-Meff	−0.359	−0.455	0.323	0.525	0.555	0.691	0.692	0.636	0.484	0.295	0.618	0.529
Long-range distance prediction accuracy												
CaCa	4.14	0.262	0.624	0.53	0.546	0.152	0.275	0.470	0.694	0.854	0.398	0.573
CaCg	4.12	0.271	0.630	0.490	0.532	0.134	0.256	0.450	0.683	0.860	0.381	0.562
CbCb	4.05	0.268	0.638	0.527	0.555	0.137	0.259	0.455	0.687	0.868	0.385	0.567
CgCg	4.30	0.287	0.620	0.523	0.540	0.121	0.231	0.416	0.659	0.854	0.357	0.540
NO	4.01	0.260	0.634	0.484	0.533	0.154	0.281	0.476	0.697	0.858	0.402	0.578
CC-TM	−0.675	−0.716	0.695	0.673	0.851	0.725	0.754	0.790	0.759	0.670	0.781	0.766
CC-Meff	−0.328	−0.381	0.353	0.491	0.492	0.630	0.621	0.570	0.436	0.280	0.550	0.472

See the main text for the explanation of the metrics. Row "CC-TM" shows the correlation coefficient between one specific metric and the quality (TMscore) of the 3D produced by RaptorX-Contact. Row "CC-Meff" shows the correlation coefficient between one metric and the logarithm of MSA depth.

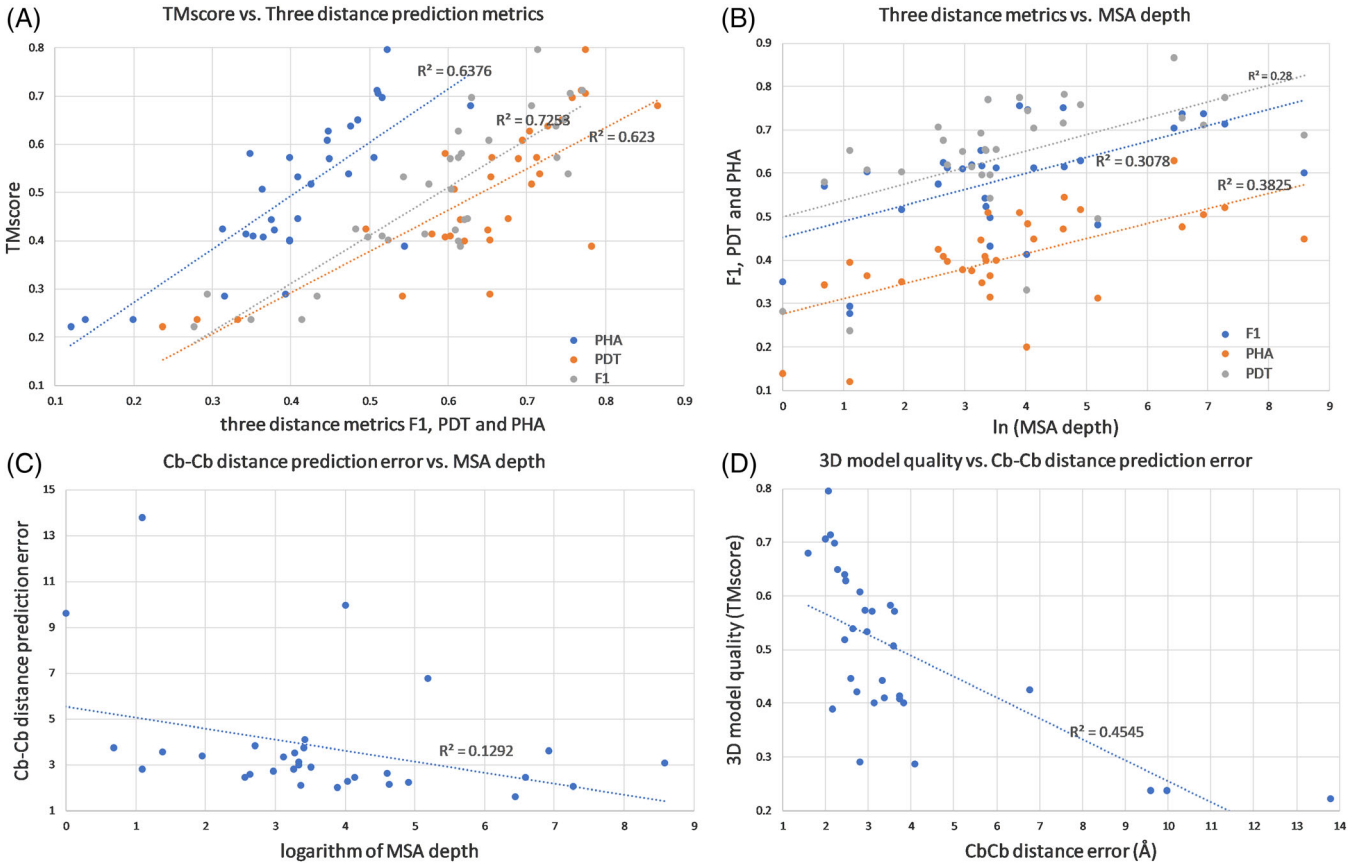

FIGURE 3 Distance prediction analysis. (A) relationship between the quality of 3D models predicted by RaptorX-Contact and distance accuracy in terms of three metrics F1, PDT and PHA; (B) relationship between distance accuracy (measured by F1, PDT, and PHA) and MSA depth; (C) relationship between absolute distance prediction error and MSA depth; (D) relationship between the quality of 3D models predicted by RaptorX-Contact and absolute distance error. MSA, multiple sequence alignment; PHA, high-accuracy pairwise distance test; PDT, pairwise distance test

34 native long-range distance <15 Å, which is much smaller than typical. While estimating distance bounds from predicted distance distribution, we assumed each target had about 7L long-range pairs with distance<15 Å, which resulted in a big prediction error. RaptorX-Contact did not do well on T0989 maybe because it has two domains, among which one has much better coevolution signal than the other. We did not split T0989 into two domains, which resulted in many more C_β-C_β pairs in D1 being assumed to have distance<15 Å and thus, led to a big prediction error. When T0989-D1 is predicted independently, RaptorX has absolute distance error only 4.89 Å.

3.3 | Distance-based tertiary structure modeling accuracy in CASP13

When all the CASP13 targets are considered, RaptorX-DeepModeller has the best 3D modeling performance among the three RaptorX servers. When only FM targets are considered, the average quality (TMscore) of the first models predicted by RatporX-DeepModeller, RaptorX-Contact, and RaptorX-TBM is 0.471, 0.474, and 0.402, respectively. When the best of five models are considered for each target, the average quality of the models generated by RatporX-DeepModeller, RaptorX-Contact, and RaptorX-TBM is 0.498, 0.501, and 0.420, respectively, and they predicted correct folds (TMscore > 0.5) for 17, 17, and 9 targets, respectively. When combined together, RaptorX-DeepModeller and RaptorX-Contact predicted correct folds for 18 of the 32 FM targets. Note that due to limit of computing power, RaptorX-DeepModeller was trained by a smaller set of proteins than RaptorX-Contact.

Figure 4 studies the relationship among the three RaptorX servers and an old in-house threading program CNFpred in terms of their performance on the FM targets. RaptorX-DeepModeller modeling accuracy is highly correlated with RaptorX-Contact (Figure 4A) and RaptorX-TBM (Figure 4B) since they mainly depend on pairwise distance information predicted by deep ResNet, but the correlation between RaptorX-Contact and RaptorX-TBM is not that strong (Figure 4C). Although both used the same template database, RaptorX-TBM greatly outperformed CNFpred (Figure 4D) since the former can recognize structurally similar templates even if they are not evolutionarily related to a target. This indicates the importance of pairwise distance in protein threading with remotely related

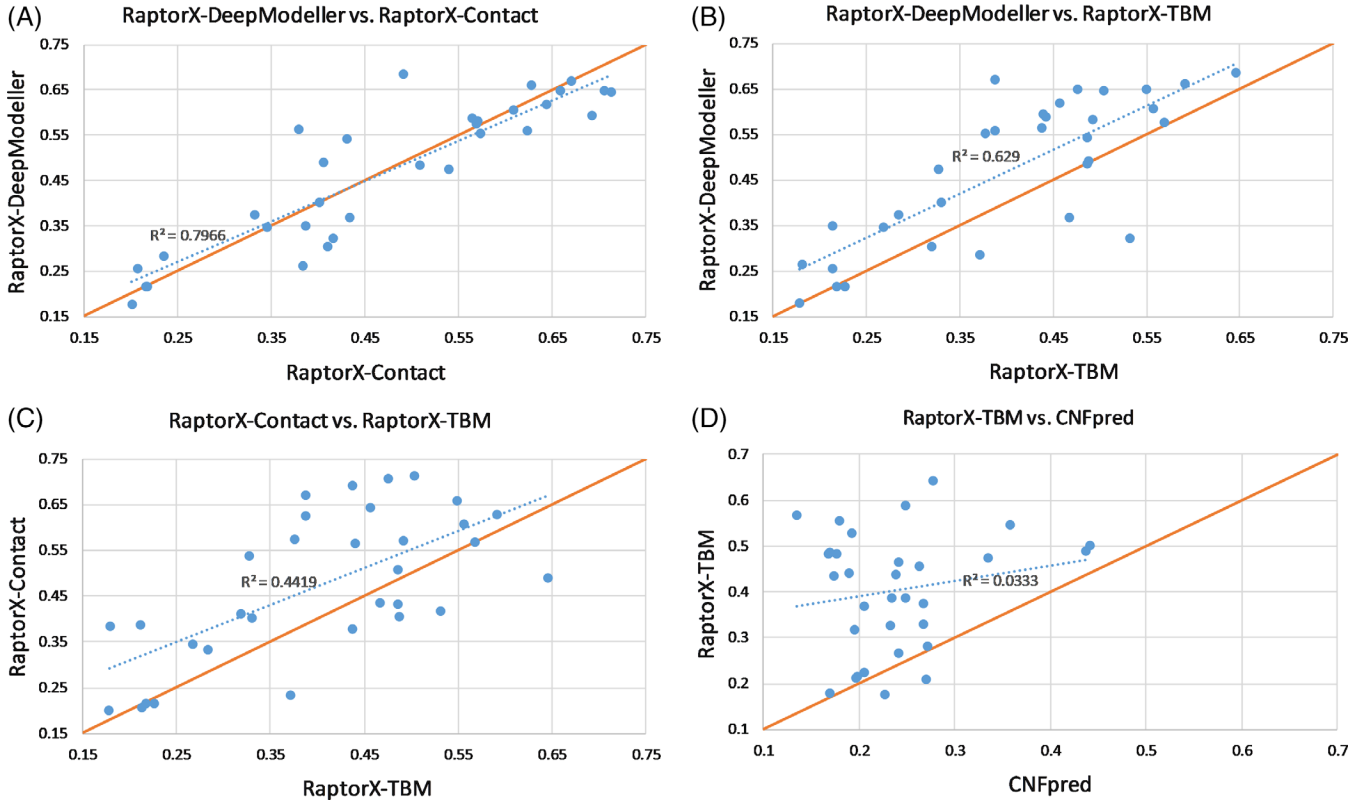

FIGURE 4 Relationship between RaptorX-DeepModeller, RaptorX-Contact, and RaptorX-TBM and an old in-housing threading program CNFpred, in terms of their performance on the CASP13 FM targets

templates. We did not have a contact-based threading program, but Zhang's CEthreader[37] is such a program and was tested in CASP13. RaptorX-TBM outperformed CEthreader by about 14% in terms of the TMscore of the first models on FM targets. This may suggest that predicted distance is much more informative than contacts for template-based modeling.

To further compare the three RaptorX servers, we display the quality (TMscore) of their best models for all the FM, FM/TBM, and TBM-hard targets, as shown in Figure 5. On FM targets RaptorX-DeepModeller and RaptorX-Contact clearly outperformed RaptorX-TBM (also shown in Figure 3C), which implies that for FM targets template-based modeling is insufficient even if predicted distance information is applied to template selection and sequence-template alignment. On FM/TBM targets, RaptorX-DeepModeller has a larger advantage over both RaptorX-Contact and RaptorX-TBM. In particular, RaptorX-DeepModeller predicted better models for eight FM/TBM targets than RaptorX-Contact and RaptorX-TBM combined. For TBM-hard targets, RaptorX-DeepModeller and RaptorX-TBM performed similarly and outperformed RaptorX-Contact. However, for a couple of TBM-hard targets (T0979-D1 and T1021s1-D1), RaptorX-DeepModeller (and RaptorX-Contact) did worse than RaptorX-TBM possibly because our distance-geometry-based folding algorithm is not as good as sampling-based energy minimization method used by RaptorX-TBM. For one TBM-hard target T0981-D5, RaptorX-DeepModeller and RaptorX-TBM did worse than RaptorX-Contact, even if this target does not have a deep MSA (ln[Meff] = 3.84). In summary, it is useful to integrate both template and coevolutionary information for most FM/TBM targets and such an integration on average does not decrease modeling accuracy of FM targets.

3.4 | Progress in contact prediction and 3D structure modeling

We did not keep our old ResNet models developed in 2016, so cannot measure their performance on the CASP13 FM targets. Here we measure our progress using the 37 CASP12 FM targets. We compare contact prediction and folding accuracy of the two old versions of our ResNet model (denoted as CASP12-submit and CASP12-postdict) and the version trained right before CASP13 (denoted as CASP13-submit). CASP12-submit represents our model used during CASP12 in 2016. As explained before,[11] our ResNet method was under development during CASP12 and CASP12-submit was updated from time to time, so CASP12-submit is not a single complete version of our ResNet model. CASP12-postdict was trained right after CASP12, representing a full implementation of our deep ResNet method described by Wang et al.[6] CASP13-submit was trained right before CASP13, improving over CASP12-postdict in the following aspects: (a) dilated convolution is used in CASP13-submit; and (b) 25 discrete distance bins are used in CASP13-submit as labels while three distance bins (0-8, 8-15, and >15 Å) are used in CASP12-postdict. To compare these three versions

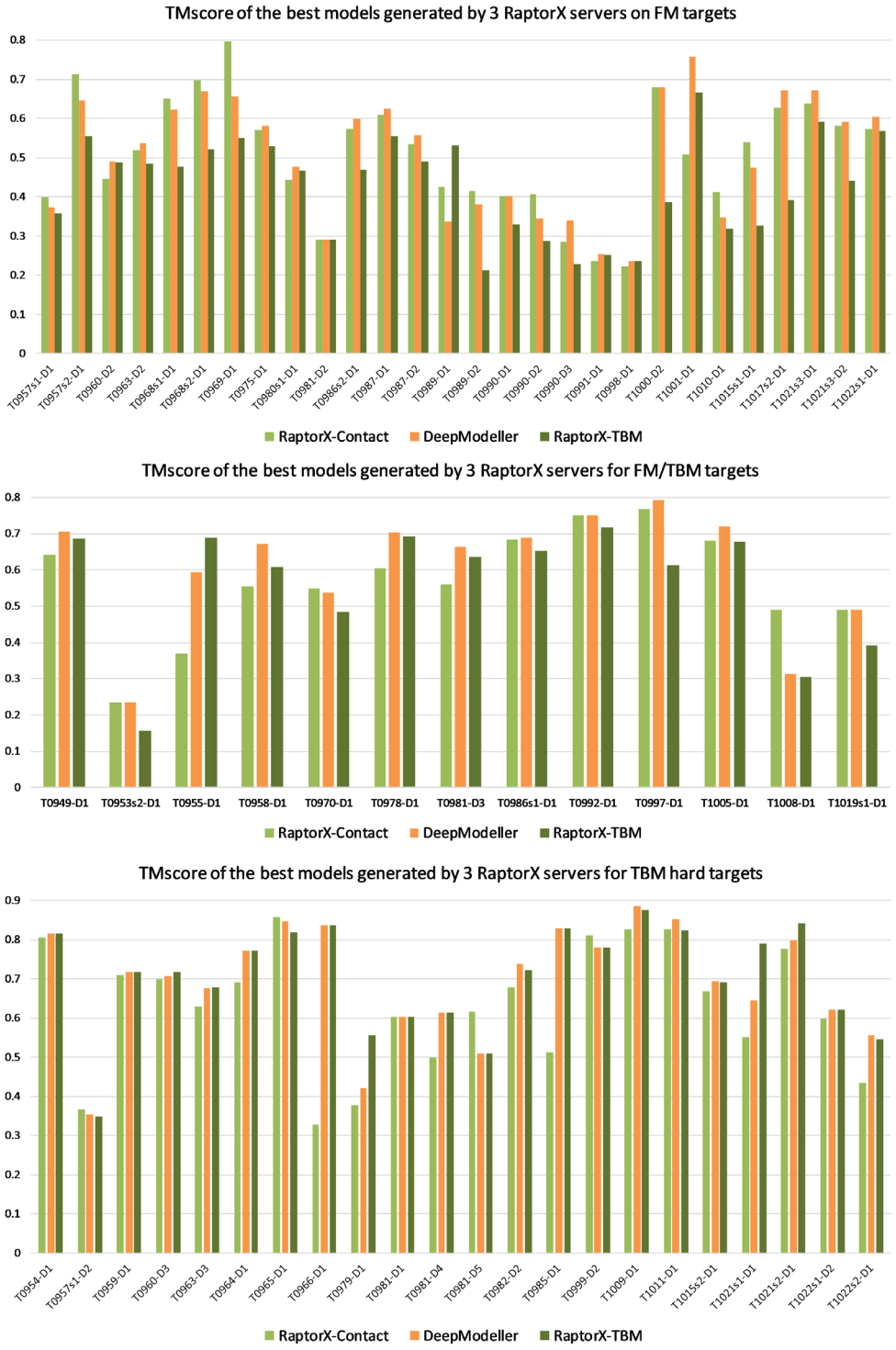

FIGURE 5 Detailed comparison of RaptorX-DeepModeller, RaptorX-Contact, and RaptorX-TBM on CASP13 FM (top), FM/TBM (middle), and TBM-hard targets (bottom)

of our method, we use the same input features for the 37 CASP12 FM targets, which were generated in CASP12. Note that for fair comparison, both CASP12-postdict and CASP13-submit were trained by the same set of ~10 000 training proteins. Nevertheless, while used in CASP13, CASP13-submit was trained by a slightly larger set of 11 410 proteins.

CASP12-postdict yields much better contact prediction and structure modeling than CASP12-submit because the former is a complete implementation of our ResNet method while the latter is not (Table 2). The results of both CASP12-postdict and CASP12-submit are taken from Tables 1 and 3 of our previous paper.[11] For contact prediction, on average CASP13-submit outperforms CASP12-postdict by about 6%-7%. That is, compared to our CASP12 submission, we have greatly improved contact prediction, but compared to the ResNet model we have fully implemented in 2016, the improvement is only 6%-7%. The major improvement of our CASP13 method over

TABLE 2 Progress in terms of long-range contact prediction precision and 3D structure modeling on the 37 CASP12 FM targets

	Long-range contact precision (%)				Folding accuracy (TMscore)	
	L	*L*/2	*L*/5	*L*/10	Top 1	Top 5
CASP12-submit	28.63	36.42	46.76	51.50	0.274	0.307
CASP12-postdict	40.18	50.20	58.87	63.93	0.354	0.397
CASP13-submit	43.10	56.90	66.90	73.80	0.466	0.476

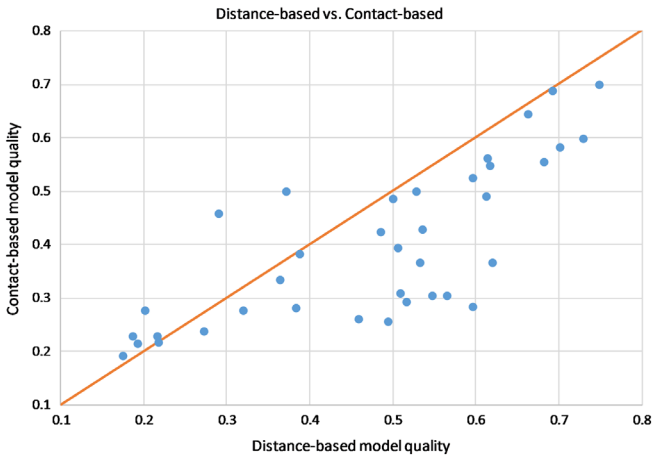

FIGURE 6 Distance-based folding accuracy versus contact-based folding accuracy, measured on the 37 CASP12 FM targets

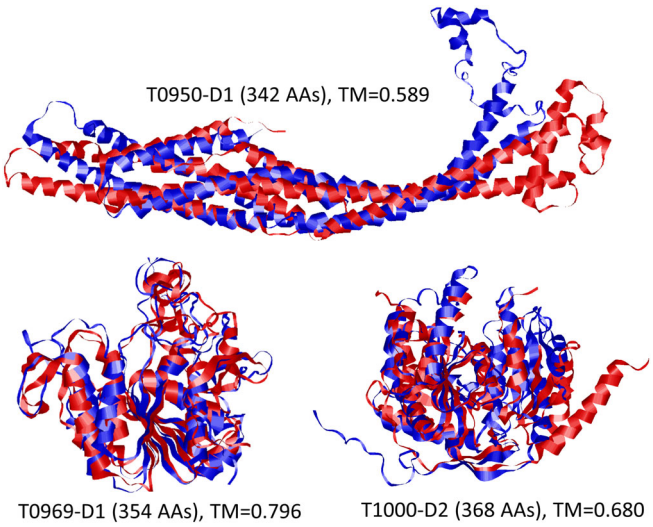

FIGURE 7 Superimposition between the best 3D models (blue) predicted by RaptorX servers and the native structure (red) for three largest CASP13 FM targets (T0950-D1, T0969-D1 and T1000-D2)

CASP12-postdict lies in 3D structure modeling. Table 2 shows that by using distance-based instead of contact-based folding, we may improve 3D structure model quality by 0.1 in terms of TMscore. For most targets, distance-based folding generated better 3D models than contact-based folding (Figure 6).

3.5 | Case study

Here we study the models predicted by our servers for the three largest CASP13 FM targets: T0950-D1, T0969-D1, and T1000-D2. They have 342, 354, and 368 residues with valid coordinates, respectively, and MSA depth 103, 1132, and 877, respectively. The fourth largest target has only ~280 residues with valid coordinates. Our servers predicted correct folds for all three targets and the best models for T0950-D1 and T0969-D1 among all human and server groups. In contrast, AlphaFold predicted correct folds for two of them, although on average AlphaFold has better modeling accuracy on all the FM targets.

3.5.1 | T0950-D1

This is a very hard target and its MSA depth normalized by sequence length is only ~0.3. Among all the 3D models accepted by CASP13, only seven have a correct fold (ie, TMscore > 0.5), including five models generated by RaptorX-DeepModeller, one by Zhang's QUARK[38] and one by Cheng's human group MULTICOM. The best model was generated by RaptorX-DeepModeller, which has TMscore = 0.589 (Figure 7). MULTICOM's best model is a copy of RaptorX-DeepModeller's first model with TMscore = 0.564. Zhang's and AlphaFold's best models have TMscore 0.506 and 0.443, respectively. Note that although this is a server-only target, some human groups such as AlphaFold and MULTICOM still submitted their predictions. Structure alignment by DeepAlign[39] shows that the most similar training protein in our training set has TMscore = 0.542 with this target. The template-based models predicted by RaptorX-TBM and CNFpred have TMscore = 0.437 and 0.173, respectively, much worse than RaptorX-DeepModeller models. This implies that RaptorX-DeepModeller predicted models for this target not by copying from a single template, although RaptorX-DeepModeller used alignments generated by RaptorX-TBM as input. That is, RaptorX-DeepModeller is able to generate better models than both RaptorX-Contact and RaptorX-TBM by combining template and coevolution information.

3.5.2 | T0969-D1

This target has the largest MSA depth among all CASP13 FM targets. Quite a few groups predicted models with a correct fold. RaptorX-Contact predicted a model with TMscore = 0.796 (Figure 7), better than all the other server models and all but four human models (MESHI, MUFold, Seder3mm and Elofsson), which have similar quality as the RaptorX-Contact model simply because they are derived from

this server model. For this target, AlphaFold's and Zhang's best models have TMscore = 0.730 and 0.680, respectively. The most similar training protein in our training set has TMscore = 0.452 with this target. The template-based models predicted by RaptorX-TBM and CNFpred have TMscore = 0.549 and 0.357, respectively. This implies that our server model is not copied from individual proteins in our training set or template database.

3.5.3 | T1000-D2

This is a very large protein domain with MSA depth being 877. Quite a few groups predicted models of a correct fold. In particular, AlphaFold and Zhang predicted very good models with TMscore = 0.880 and 0.851, respectively. Our servers have also predicted a correct fold (Figure 7), but with much lower model quality (TMscore = 0.680). The most similar training protein in our training set has TMscore = 0.347 with the target. The template-based models predicted by RaptorX-TBM and CNFpred have TMscore = 0.387 and 0.233, respectively. That is, our server model is not copied from individual proteins in our training set or template database. For this target, our contact prediction is ranked very top, better than Zhang's contact prediction regardless of ranking metrics, but our 3D modeling has much lower quality than Zhang's. This may indicate that we did not do a good job in building 3D model from predicted distance distribution.

3.6 | Running time

Our distance-based folding method runs very fast and may produce a single decoy from seconds to minutes. Typically, it takes minutes to generate MSA, seconds to predict distance, and from minutes to hours to generate 200 decoys from predicted distance. Since in CASP13 we ran different jobs (eg, feature generation, folding, and threading) simultaneously on a single machine and different targets on different machines, it is hard for us to accurately measure the relationship between running time and sequence length based upon our CASP13 data. However, by chance we have run all 37 CASP12 FM targets on a single Linux workstation (20 processors and 512G RAM) without interference of other jobs, so here we show the running time (in minutes) of the folding step (excluding feature generation and distance prediction) with respect to sequence length using the CASP12 data. Roughly speaking, the running time has a very good linear correlation (coefficient = 0.885, trendline R^2 = 0.783) with sequence length although a 2nd-degree polynomial may fit the running time slightly better (trendline R^2 = 0.826) (Figure 8).

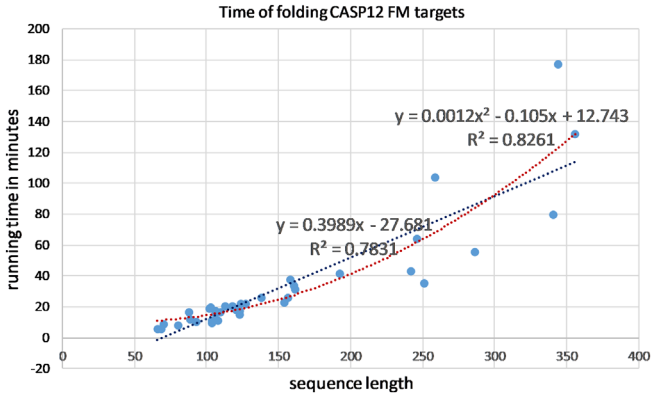

FIGURE 8 Folding time of RaptorX-Contact with respect to sequence length, measured on CASP12 FM targets

4 | DISCUSSION AND CONCLUSION

Our CASP13 result confirms that by using deep ResNet to predict interatom or interresidue distance, we may fold proteins much more accurately than ever before and within minutes to hours even on a Linux workstation. In particular, our distance-based folding algorithm predicted the best models for two of the three largest FM targets with ~350 residues. Our analysis shows that predicted distance is useful for both template-free and template-based modeling, and that protein modeling can be further improved by combining template and coevolutionary information.

4.1 | What went right?

The CASP13 result is consistent with our previous findings:[8,9] (a) protein distance matrix can be predicted very well by a fully deep ResNet, (b) predicted distance is very useful for both template-based and template-free modeling, (c) predicted distance is more informative than contacts for protein structure modeling, and (4) protein structure modeling can be further improved by integrating template and coevolutionary information. Our protocol for contact prediction works very well even if we did not use as many layers as AlphaFold or as many input features as other top-performing groups. Our distance geometry method for building 3D models from predicted distance works fine on very large targets even without extensive conformation sampling.

4.2 | What went wrong?

Although our protocol for predicting contacts from predicted distance distribution worked very well, our protocol for building 3D models from predicted distance distribution was not optimal. Our contact prediction has similar F1 value as AlphaFold, but our folding protocol underperformed AlphaFold on quite a few FM targets even if on very large targets our method has favorable performance. This may imply that for many FM targets our distance-geometry-based folding method may not work as well as the energy minimization methods and/or we did not do a good job in deriving distance bounds from predicted distance distribution. There are several potential issues with our bound estimation method. First, some information could be lost while converting distance distribution to distance bounds. Second, while estimating mean distance and SD from predicted distance distribution, we ignored the predicted probability of distance >16 Å for those atom pairs which are likely to have native distance less than 15 Å. Third, we assumed that each target has at least 7L C_β-C_β pairs

with native distance less than 15 Å. Such an assumption leads to a big distance prediction error for T0953s1, which has only a small number of C_β-C_β pairs with native distance <15 Å.

We did not handle some multidomain targets very well, especially when domains have very different MSA depth. In this case some domains of a target may have much stronger coevolution signal than the others. Since we did not split a multidomain target into domains when none of them have reasonable templates, the atom pairs chosen by us to build 3D models may not spread uniformly across different domains or segments. As such, some domains (or segments) may be covered by very few selected atom pairs while others may be covered by many more selected atom pairs. Both scenarios may lead to bad 3D modeling.

Our current folding protocol does not use any energy terms, fragment assembly, or extensive conformation sampling, which may prevent us from generating very high-resolution models. For example, beta sheets in some models do not form very well since we did not use any hydrogen-bonding energy to promote their formation. Finally, we have not used the metagenomics sequence database for MSA generation, which may be helpful for few targets.

4.3 | Why did deep learning perform much better in CASP13 than before?

Deep learning such as Deep Belief Networks (DBN) has been attempted by Cheng group for protein contact prediction in 2012,[40,41] but it drew little attention from the community. This is mainly because that Cheng's DBN method has almost the same performance as traditional machine learning methods such as Random Forests, as reported.[40,41] In CASP10, the DBN method outperformed Random Forests by only 1.2% in terms of average long-range contact precision. In CASP11, DBN underperformed MetaPSICOV (a traditional neural network method) by a good margin and in CASP12, Cheng group still underperformed MetaPSICOV even if this group used both DBN and MetaPSICOV. The main reason why deep learning becomes effective for protein folding in the past 2 to 3 years is not the enlargement of protein sequence databases, but the introduction of a totally new formulation of contact prediction (ie, simultaneous prediction of all contacts in the whole matrix or at least a large submatrix instead of predicting contacts one-by-one) and new deep network architecture (ie, fully deep ResNet). In fact, the CASP13 FM targets have similar MSA depth as the CASP12 FM targets, but the contact prediction and folding accuracy on CASP13 FM targets are much higher due to community-wide adoption of fully deep ResNet.

ResNet is not derived from DBN, instead it is a variant of convolutional neural networks, which were invented much earlier than DBN. For contact/distance prediction, the difference between fully ResNet and DBN is analogous to that between DCA and mutual information. Both DCA and deep ResNet are global methods while mutual information and DBN are local methods. While predicting the label (ie, contact or distance) of two residues, global methods look at the labels of other residue pairs, but local methods do not. By applying a global convolutional operation to the whole contact/distance matrix, deep ResNet may learn protein structure patterns easily and yield much better contact/distance prediction. As pointed out by us before,[11] when one contact is predicted independent of the others, even if ResNet is applied, the resultant contact prediction accuracy is not very good.

Our ResNet method for contact prediction (ie, RaptorX-Contact) was not fully implemented during CASP12, not to mention the folding protocol. Because of this, in CASP12 contact prediction, RaptorX-Contact was not much better than MetaPSICOV, although RaptorX-Contact was ranked first. As shown in Table 2 and previous blind CAMEO test,[42] a full implementation of our ResNet method right after CASP12 has much better contact prediction and folding accuracy.

4.4 | The major difference among top groups in CASP13

Many CASP13 groups have incorporated deep convolutional neural network into their structure prediction pipelines. A nature question to ask is what are their major difference? Here we analyze four representative groups: AlphaFold, Zhang, RaptorX, and Baker.[43] We do not consider some human groups such as Cheng's MULTICOM since they heavily relied on consensus analysis of server models, and it is unclear if there are any major methods underlying their results other than consensus analysis.

AlphaFold used deep ResNet to predict interresidue distance distribution, converted this distribution to protein-specific distance potential using a similar method as described by Zhao and Xu,[21] and then minimized it by sampling- and gradient-based methods to build 3D models. Zhang used ResNet to predict interresidue contacts and then used them to guide extensive fragment-based conformation sampling. RaptorX used ResNet to predict interatom distance distribution, converted it to mean distance and deviation and then used this as distance bounds to build 3D models by CNS. Baker used DCA to predict contacts for some targets and then conducted contact-assisted and fragment-based conformation sampling. On average Baker's Rosetta underperformed the other three groups due to lack of the deep learning module, although Rosetta may do extensive conformation sampling. AlphaFold and RaptorX have a similar contact prediction performance, which is slightly better than Zhang. In terms of 3D modeling, RaptorX has similar performance as Zhang's servers, although RaptorX did not use extensive conformation sampling. For 3D modeling, RaptorX's strength lies in distance prediction, which is more informative than contact prediction, but Zhang made it up by incorporating predicted contacts into a well-developed folding engine. Compared to RaptorX, AlphaFold has a much better folding protocol, that is, building 3D models by sampling- and gradient-based energy minimization methods. Compared to Zhang, AlphaFold has more informative distance prediction. In summary, AlphaFold did well in both the deep learning and model building steps while the other three

groups did well in only one of them, which is why AlphaFold stands out in modeling FM targets.

ACKNOWLEDGEMENTS

This work is supported by National Institutes of Health grant R01GM089753 to J.X. and National Science Foundation grant DBI-1564955 to J.X. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors thank CASP13 organizers and assessors and all contributors of the experimental data.

AUTHOR CONTRIBUTIONS

J.X. conceived, designed, and implemented the major algorithms and wrote the paper. S.W. built the initial pipeline for MSA generation and template-based modeling and helped draw Figure 1.

ORCID

Jinbo Xu https://orcid.org/0000-0001-7111-4839

REFERENCES

1. Lapedes AS, Giraud B, Liu L, Stormo GD. In: Seillier-Moiseiwitsch F, ed. *Statistics in Molecular Biology and Genetics*. Vol 33. Hayward, CA: Institute of Mathematical Statistics; 1999:236-256.
2. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein–protein interaction by message passing. *Proc Natl Acad Sci U S A*. 2009;106:67-72.
3. Burger L, van Nimwegen E. Disentangling direct from indirect co-evolution of residues in protein alignments. *PLoS Comput Biol*. 2010;6:e1000633.
4. Balakrishnan S, Kamisetty H, Carbonell JG, Lee SI, Langmead CJ. Learning generative models for protein fold families. *Proteins*. 2011;79:1061-1078.
5. Marks DS, Colwell LJ, Sheridan R, et al. Protein 3D structure computed from evolutionary sequence variation. *PloS One*. 2011;6:e28766.
6. Wang S, Sun S, Li Z, Zhang R, Xu J. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput Biol*. 2017;13:e1005324.
7. Liu Y, Palmedo P, Ye Q, Berger B, Peng J. Enhancing evolutionary couplings with deep convolutional neural networks. *Cell Syst*. 2018;6:65-74 e63.
8. Zhu JW, Wang S, Bu DB, Xu JB. Protein threading using residue co-variation and deep learning. *Bioinformatics*. 2018;34:263-273.
9. Xu J. Distance-based protein folding powered by deep learning. *PNAS*. 2019;116:16856-16865.
10. Wang S, Li Z, Yu Y, Xu J. Folding membrane proteins by deep transfer learning. *Cell Syst*. 2017;5:202-211.e203.
11. Wang S, Sun SQ, Xu JB. Analysis of deep learning methods for blind protein contact prediction in CASP12. *Proteins*. 2018;86:67-77.
12. Zeng H, Wang S, Zhou T, et al. ComplexContact: a web server for inter-protein contact prediction using deep learning. *Nucleic Acids Res*. 2018;46:W432-W437.
13. Wang S, Fei S, Wang Z, et al. PredMP: a web server for de novo prediction and visualization of membrane proteins. *Bioinformatics*. 2019;35:691-693.
14. Adhikari B, Hou J, Cheng J. DNCON2: improved protein contact prediction using two-level deep convolutional neural networks. *Bioinformatics*. 2017;34:1466-1472.
15. Jones DT, Kandathil SM. High precision in protein contact prediction using fully convolutional neural networks and minimal sequence features. *Bioinformatics*. 2018;34:3308-3315.
16. Hanson J, Paliwal K, Litfin T, Yang Y, Zhou Y. Accurate prediction of protein contact maps by coupling residual two-dimensional bidirectional long short-term memory with convolutional neural networks. *Bioinformatics*. 2018;34:4039-4045.
17. Aszodi A, Gradwell MJ, Taylor WR. Global fold determination from a small number of distance restraints. *J Mol Biol*. 1995;251:308-326.
18. Kloczkowski A, Jernigan RL, Wu Z, et al. Distance matrix-based approach to protein structure prediction. *J Struct Funct Genomics*. 2009;10:67-81.
19. Pietal MJ, Bujnicki JM, Kozlowski LP. GDFuzz3D: a method for protein 3D structure reconstruction from contact maps, based on a non-Euclidean distance function. *Bioinformatics*. 2015;31:3499-3505.
20. Kukic P, Mirabello C, Tradigo G, Walsh I, Veltri P, Pollastri G. Toward an accurate prediction of inter-residue distances in proteins using 2D recursive neural networks. *BMC Bioinformatics*. 2014;15:6.
21. Zhao F, Xu J. A position-specific distance-dependent statistical potential for protein structure and functional study. *Structure*. 2012;20:1118-1126.
22. Ma JZ, Wang S, Wang ZY, Xu JB. MRFalign: protein homology detection through alignment of Markov random fields. *PLoS Comput Biol*. 2014;10:e1003500.
23. Wang Z. Vol. PhD 110. Chicago: Toyota Technological Institute at Chicago; 2016.
24. Fisher Y, Vladlen K. Multi-scale context aggregation by dilated convolutions. In: *International Conference on Learning Representations (ICLR)*, Scottsdale, Arizona. 2016.
25. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. 2012;9:173-175.
26. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*. 2010;11:431.
27. Wang S, Li W, Liu S, Xu J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res*. 2016;44:gkw306.
28. Seemayer S, Gruber M, Söding J. CCMpred: fast and precise prediction of protein residue–residue contacts from correlated mutations. *Bioinformatics*. 2014;30:3128-3130.
29. Song YF et al. High-resolution comparative modeling with RosettaCM. *Structure*. 2013;21:1735-1742.
30. Ma JZ, Peng J, Wang S, Xu JB. A conditional neural fields model for protein threading. *Bioinformatics*. 2012;28:I59-I66.
31. Peng J, Bo L, Xu J. Conditional neural fields. In: Bengio Y, Schuurmans D, Lafferty JD, Williams CKI, Culotta A, eds. *Advances in Neural Information Processing Systems 22*. New York: Curran Associates, Inc.; 2012:1419-1427. http://papers.nips.cc/paper/3869-conditional-neural-fields.pdf.
32. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33:W244-W248.
33. Brunger AT. Version 1.2 of the crystallography and NMR system. *Nat Protoc*. 2007;2:2728-2733.
34. Adhikari B, Bhattacharya D, Cao R, Cheng J. CONFOLD: residue-residue contact-guided ab initio protein folding. *Proteins*. 2015;83:1436-1449.
35. Jones DT, Singh T, Kosciolek T, Tetchner S. MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinformatics*. 2015;31:999-1006.
36. Baldassi C, Zamparo M, Feinauer C, et al. Fast and accurate multivariate Gaussian modeling of protein families: predicting residue contacts and protein-interaction partners. *PloS One*. 2014;9:e92721.

37. Zheng W, Zhang C, Wuyun Q, Pearce R, Li Y, Zhang Y. LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. *Nucleic Acids Res*. 2019;47:W429-W436.
38. Zhang CX, Mortuza SM, He BJ, Wang YT, Zhang Y. Template-based and free modeling of I-TASSER and QUARK pipelines using predicted contact maps in CASP12. *Proteins*. 2018;86:136-151.
39. Wang S, Ma J, Peng J, Xu J. Protein structure alignment beyond spatial proximity. *Sci Rep*. 2013;3:1448.
40. Eickholt J, Cheng J. Predicting protein residue-residue contacts using deep networks and boosting. *Bioinformatics*. 2012;28:3066-3072.
41. Eickholt J, Cheng JL. A study and benchmark of DNcon: a method for protein residue-residue contact prediction using deep networks. *BMC Bioinformatics*. 2013;14:S12.
42. Haas J, Barbato A, Behringer D, et al. Continuous automated model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. *Proteins*. 2018;86:387-398.
43. Ovchinnikov S, Park H, Kim DE, DiMaio F, Baker D. Protein structure prediction using Rosetta in CASP12. *Proteins*. 2018;86:113-121.

How to cite this article: Xu J, Wang S. Analysis of distance-based protein structure prediction by deep learning in CASP13. *Proteins*. 2019;87:1069–1081. https://doi.org/10.1002/prot.25810