OXFORD

Structural bioinformatics

RNA inter-nucleotide 3D closeness prediction by deep residual neural networks

Saisai Sun[1], Wenkai Wang[1], Zhenling Peng[2] and Jianyi Yang ⬡ [1,*]

[1]School of Mathematical Sciences, Nankai University, Tianjin 300071, China and [2]Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

*To whom correspondence should be addressed.

Associate Editor: Yann Ponty

Abstract

Motivation: Recent years have witnessed that the inter-residue contact/distance in proteins could be accurately predicted by deep neural networks, which significantly improve the accuracy of predicted protein structure models. In contrast, fewer studies have been done for the prediction of RNA inter-nucleotide 3D closeness.

Results: We proposed a new algorithm named RNAcontact for the prediction of RNA inter-nucleotide 3D closeness. RNAcontact was built based on the deep residual neural networks. The covariance information from multiple sequence alignments and the predicted secondary structure were used as the input features of the networks. Experiments show that RNAcontact achieves the respective precisions of 0.8 and 0.6 for the top $L/10$ and L (where L is the length of an RNA) predictions on an independent test set, significantly higher than other evolutionary coupling methods. Analysis shows that about 1/3 of the correctly predicted 3D closenesses are not base pairings of secondary structure, which are critical to the determination of RNA structure. In addition, we demonstrated that the predicted 3D closeness could be used as distance restraints to guide RNA structure folding by the 3dRNA package. More accurate models could be built by using the predicted 3D closeness than the models without using 3D closeness.

Availability and implementation: The webserver and a standalone package are available at: http://yanglab.nankai.edu.cn/RNAcontact/.

Contact: yangjy@nankai.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

It is well accepted that the 3D structure of RNA determines its biological function. As revealed in the RNA-Puzzles experiment (Miao *et al.*, 2017), it is urgent to develop computational algorithms to predict RNA 3D structures. Most *de novo* modeling methods use a coarse-grained representation and simplify various aspects of RNA structures, while simulating RNA structures with some restraints, such as secondary structure and atomic distance (Boniecki *et al.*, 2016; Jonikas *et al.*, 2009; Krokhotin *et al.*, 2015). While homology modeling and fragment assembly approaches build the RNA structure using the atomic coordinates of known structures in the Protein Data Bank (PDB) (Berman, 2000), some recent fragment assembly methods also add secondary structure, inter-nucleotide interaction as restraints during structure optimization (Antczak *et al.*, 2017; Wang *et al.*, 2017a). It was shown that inter-nucleotide interaction could be predicted with direct coupling (also called co-evolution) analysis, which can also be used as restraints to improve the accuracy of RNA structure modeling (De Leonardis *et al.*, 2015; Wang *et al.*, 2017a; Weinreb *et al.*, 2016).

In recent years, deep learning was proved to be very powerful for improving the accuracy of inter-residue contact/distance prediction and protein structure prediction (Abriata *et al.*, 2019; Xu, 2019; Yang, 2020). The precision of inter-residue contact prediction in proteins was almost doubled by replacing direct coupling methods with deep residual neural networks (Kandathil *et al.*, 2019; Li *et al.*, 2019; Wang *et al.*, 2017b; Wu *et al.*, 2020). However, to the best of our knowledge, deep learning has not been explored for improving RNA inter-nucleotide interaction prediction. Note that the word 'contact' from the RNA community is usually used to describe various base-pairing geometries (Leontis and Westhof, 2001; Leontis and Zirbel, 2012), which is different with protein's inter-residue contact. Thus rather than using the word 'contact' (Weinreb *et al.*, 2016), we use '3D closeness' to describe the tertiary inter-nucleotide interactions to avoid confusion.

In this work, we developed RNAcontact, a deep learning-based algorithm for the prediction of RNA inter-nucleotide 3D closeness. Benchmark tests revealed that the predicted 3D closeness were more accurate than those predicted by the direct coupling methods.

The predicted 3D closeness were used as restraints to guide RNA structure prediction, generating significantly more accurate models than the models without restraints.

2 Materials and methods

2.1 Overview of the proposed method
Figure 1 gives an overview of the proposed method RNAcontact. The only input to our method is the nucleotide sequence, which is submitted to Infernal (Nawrocki and Eddy, 2013) and PETfold (Seemann *et al.*, 2008) to collect the homologous sequences and predict the secondary structure, respectively. The homologous sequences are used to construct a multiple sequence alignment (MSA), from which covariance features are extracted. The feature channels are then fed into a deep residual neural network to predict the 3D closeness for all nucleotide pairs.

2.2 Benchmark datasets
Our benchmark datasets were constructed based on a set of non-redundant RNA 3D structures from Leontis and Zirbel (2012) (Version 3.99, 2019-11-06), which is updated weekly. First, a representative dataset containing 1786 entries with resolution < 4 Å was downloaded. Then RNA chains with length < 32 nt or >1000 nt were removed and 511 RNA chains remained. In addition, to reduce the similarity between these RNA chains, cd-hit-est (Li and Godzik, 2006) and BLASTclust (Altschul, 1997) were used together to remove redundant sequences at 30% sequence identity cutoff, yielding 336 RNA chains. 70% of these chains were randomly selected for training and the rest for test. The length distribution of RNAs in the training and test sets are available in the Supplementary Figure S1, which suggests that most RNAs have less than 100 nucleotides. In addition, the datasets from the work of Weinreb *et al.* (2016) and Jian *et al.* (2019), consisting of 22 RNAs and 6 RNAs, respectively, were used as additional test sets.

According to the work of Weinreb *et al.* (2016), for each two nucleotides in the whole RNA, they have tertiary interaction (i.e. positive sample) if their minimal atomic distance is less than 8 Å. Otherwise, they do not have tertiary interaction (i.e. negative sample). RNA chains with too few positive samples (the number of positive samples <5) were removed from the above datasets. So the remaining number of RNAs are: 221 in the training set (denoted by TR221), 80 in the test set (denoted by TE80), 19 from Weinreb *et al.* (2016) (denoted by W19) and 6 from Jian *et al.* (2019) (denoted by J6).

According to Jian *et al.* (2019), the 3D closeness in RNA can be classified into two classes based on sequence separation: short range ($5<|i - j|<24$) and long range ($|i - j|\geq 24$), where i and j are the ith and jth nucleotides in the sequence.
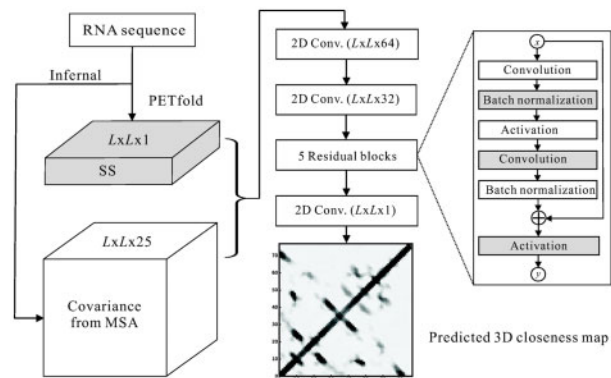

Fig. 1. The flowchart of RNAcontact. From an input RNA sequence of length L, 26 2D feature maps ($L \times L \times 26$) are calculated based on the secondary structure (SS) predicted by PETfold and the covariance signal inferred from an MSA by Infernal (shown on the left). The feature maps are fed into the deep residual network (shown in the middle) with five residual blocks (shown on the right) to predict the 3D closeness map

Note that the 3D closeness defined above is based on pairwise distance, borrowed from previous studies (Jian *et al.*, 2019; Weinreb *et al.*, 2016). On the other hand, the RNA 3D Structures Atlas (Leontis and Zirbel, 2012) has a different definition about inter-nucleotide interaction by considering hydrogen bonds between nucleotides, including base-pair, base-phosphate, base-ribose and base-stacking. We did not re-train our method based on this new definition to reduce computational cost. Instead, the impact of this definition is discussed in Section 3.5, through the illustration of RNA folding.

2.3 Input features
The features we tried, included covariance, predicted secondary structure, predicted relative solvent accessible surface area and single sequence, as detailed below.

Covariance (Cov). For a query sequence, the RNA sequence alignment program Infernal (Nawrocki and Eddy, 2013) was used to search through the NCBI's non-redundant nucleotide sequence database to construct an MSA. Then, the MSA was filtered by removing redundant sequences at sequence identity cutoff 90% and sequences with more than 50% gaps. Finally, covariance features were computed from the filtered MSA with Eq. (1).

$$\text{cov}(i_x, j_y) = p(i_x j_y) - p(i_x)p(j_y) \tag{1}$$

where i and j are the ith and jth columns in the MSA, respectively; i_x (j_y) is one of the four nucleotide types or a gap at the ith (jth) column; $p()$ is the frequency of observing the corresponding nucleotide or nucleotide pair in the MSA. Thus, each MSA was transferred into a feature map of three dimensional matrix of $L \times L \times 25$ ($25 = 5 \times 5$, where 5 represents four types of nucleotides and gap; L is the length of a sequence).

Secondary structure (SS). According to the systematic comparison in the work of CompaRNA (Puton *et al.*, 2014), PETfold (Seemann *et al.*, 2008) is one of the best RNA secondary structure (SS) prediction algorithms. Thus PETfold was used here to predict the RNA SS. The predicted SS by PETfold was converted into a 2D feature map ($L \times L \times 1$, L is the length of a sequence), with elements being 1 (paired) or 0 (un-paired).

Relative solvent accessible surface area (RSA). Predicted relative solvent accessible surface area was also transferred into an input feature. RSA was predicted by RNAsol (Sun *et al.*, 2019), where the value is between 0 and 1 representing the exposure level of a nucleotide. This 1D feature was tiled horizontally and vertically to yield two 2D feature maps ($L \times L \times 2$, L is the length of a sequence).

Single sequence (Seq). Each nucleotide in a query sequence was transferred into a one-hot vector (with dimension 4). Similar to the RSA conversion, this 1D feature was tiled horizontally and vertically to yield eight 2D feature maps ($L \times L \times 8$, L is the length of a sequence).

2.4 Neural network architecture
In this work, we used the residual neural network (ResNet) to construct our network. The network architecture is shown in Figure 1, it mixed three 2D convolution layers with five residual blocks, each containing two convolution layers. It adds a shortcut connection in each residual block. The output layer is a 2D convolution layer, which outputs a 2D probability matrix of 3D closeness. We built our network using the deep learning library Keras (https://keras.io/) with Tensorflow (Abadi *et al.*, 2016) as a backend. We optimized the following parameters based on the training set: learning rate, dropout rate, kernel size, filter size and the number of layers.

2.5 Training parameters
We selected the ReLU function as the activation function, the binary cross-entropy as the loss function and the Adamax as the optimizer function. To avoid over-fitting, we added dropout in each layer and used early-stopping. Due to the different lengths of RNA chains, the batch size was set to 1 instead of using padding. All other parameters in the residual neural network, including the kernel size, filter

size, number of layers, learning rate and dropout rate, were tuned to maximize the precision on a validation set randomly selected from the training set. Parameters were tuned based on grid search. The ranges of the parameters are listed in Supplementary Table S1. After optimization, the values of the parameters are as follows: kernel size: 9; filter size: 32; number of layers: 13; learning rate: 0.0005; and dropout rate: 0.15. Due to the random effects of the training procedure, 100 models were trained. For each nucleotide pair, the average of the probabilities predicted from all these models was used as the final prediction.

2.6 Performance evaluation

Similar to the assessment in protein inter-residue contact prediction, the performance of RNA inter-nucleotide 3D closeness prediction is measured by precision (the number of true positives divided by the number of predicted positives). A predicted 3D closeness is regarded as a true positive (TP) if the two nucleotides are close in the native structure. In general, the top L/n ($n = 1$, 2, 5 or 10) of the ranked predictions are used to evaluate the performance.

3 Results and discussion

3.1 Impact of feature sets

We compared the performance of our method built with different sets of features. The results on the independent test set TE80 are listed in Table 1 (long range) and Supplementary Table S2 (short range). As we can see, the precision is the lowest when only the sequence information was used. The precision from predicted RSA features is about 10% higher than from sequence information. The predictions using the covariance and predicted secondary structure features have similar precisions, both outperforming the predicted RSA features. When combining these two features, the highest precisions are achieved, i.e. 0.89, 0.88, 0.81 and 0.66 for the top $L/10$, $L/5$, $L/2$ and L predicted long-range 3D closeness, respectively. This suggests that the secondary structure and covariance features are largely complementary. We also tried to combine them with other features but did not see an improvement, which might be due to much redundant or wrong predicted information in features.

3.2 Impact of secondary structure prediction algorithms

As shown above, the secondary structure plays an important role for improving 3D closeness prediction. Thus, we investigated more about the impact of the predicted SS to the 3D closeness prediction. Besides the PETfold, we tried the single-sequence-based RNAfold and the MSA-based RNAalifold programs (Gruber *et al.*, 2015). Both minimum free energy (MFE) and probability matrix versions from RNAfold and RNAalifold were tried here. We re-trained our network after replacing the SS prediction method PETfold by RNAfold and RNAalifold.

The results of the predicted long-range 3D closeness with different SS predictions on the test set TE80 are listed in Supplementary Table S3. It suggests that compared with other SS prediction, the SS predicted by PETfold yields 3D closeness prediction with the highest precision. The precision by the MFE version of RNAfold is slightly

lower than PETfold. It is unanticipated that the RNAalifold leads to the lowest precision, probably because the MSA we provided is not optimal for it. For both RNAfold and RNAalifold, the MFE version outperforms the probability version. In order to understand the above data, we further computed the accuracy of the predicted SS, measured by the average Matthews Correlation Coefficient (MCC). The MCCs for the SS predictions by the above programs are listed in Supplementary Table S3. The average MCC for the PETfold prediction is indeed higher than others, consistent to the corresponding 3D closeness predictions. When the native SS is used, the highest precision was achieved, suggesting that one of the possible way to improve our method is to include more accurate SS prediction in future.

3.3 Comparison with other methods

We compared our method with Weinreb's method (denoted by PLMC) on two test sets (TE80 and W19) and Jian's method (called DIRECT) on the dataset J6. Both PLMC and DIRECT used evolutionary coupling for 3D closeness prediction. The precisions for PLMC were calculated by running its standalone packages with MSA input. While for DIRECT, the precisions were obtained by running its standalone package with the input features provided at its website https:// zhaolab.com.cn/DIRECT/, which is available for the dataset J6 only.

Table 2 suggests that RNAcontact has the highest long-range precision on all test sets, outperforming other compared methods. The precision for each RNA in the test sets is listed in Supplementary Tables S4–S6. The precisions of the short-range prediction are listed in Supplementary Table S7. Moreover, the standard deviations of the precisions are provided in Supplementary Tables S8 (long range) and S9 (short range). On the test set TE80, RNAcontact achieves precisions 0.89—0.66 for the top $L/10$—the top L long-range predictions, about 15% higher than PLMC. On the test set W19, the precisions of the top $L/10$ are similar for RNAcontact and PLMC. While from the top $L/5$ to the top L prediction, our method outperforms PLMC significantly. On the test set J6, which contains 6 RNAs only, RNAcontact has significant improvement over DIRECT for all the assessed top predictions. Detailed head-to-head comparisons between RNAcontact and PLMC on the top L long-range predictions of the test sets W19 and TE80 are presented in Figure 2. It shows that on the W19 dataset, RNAcontact outperforms PLMC for 16 out of 19 RNAs. While on the TE80 dataset, RNAcontact has higher precision for 75 out of 80 RNAs.

Moreover, we compared the set of correctly predicted long-range 3D closeness (i.e. true positives, TPs) out of the top L predicted 3D closeness for RNAcontact and PLMC with a Venn diagram in Supplementary Figure S2. As we can see, the numbers of TPs of RNAcontact are 3-4 times higher than that of PLMC on both datasets. Interestingly, the overlap between both methods is small. For example, there are only 446/259 3D closeness in common on the TE80/W19 dataset. This is possibly because different methodologies are adopted by RNAcontact and PLMC.

The native and predicted 3D closeness maps from RNAcontact and PLMC on three example RNAs (PDB ID: 4r4v_A, 6ol3_C and 6fyy_1) are presented in Figure 3, which shows RNAcontact predicts most 3D closeness correctly. For these examples, the precisions of the

Table 1. Precision of the long-range 3D closeness prediction with different features on the independent test set TE80

Features	$L/10$	$L/5$	$L/2$	L
Seq	0.48	0.45	0.40	0.33
RSA	0.57	0.54	0.46	0.36
Cov	0.81	0.80	0.73	0.59
SS	0.79	0.78	0.71	0.58
Cov + SS	**0.89**	**0.88**	**0.81**	**0.66**
Cov + SS + RSA	0.68	0.66	0.55	0.41
Cov + SS + RSA + Seq	0.48	0.46	0.40	0.34

Note: The highest value on each column is highlighted in bold type.

Table 2. Precision of the long-range 3D closeness prediction on different test sets

Dataset	Method	$L/10$	$L/5$	$L/2$	L
TE80	RNAcontact	**0.89**	**0.88**	**0.81**	**0.66**
	PLMC	0.78	0.75	0.65	0.28
W19	RNAcontact	**0.97**	**0.95**	**0.88**	**0.72**
	PLMC	0.96	0.87	0.65	0.48
J6	RNAcontact	**0.98**	**0.90**	**0.84**	**0.77**
	DIRECT	0.67	0.61	0.44	0.31

Note: The highest value on each dataset is highlighted in bold type. The *P*-values for the statistical tests on the differences between RNAcontact and other methods are listed in Supplementary Table S10.

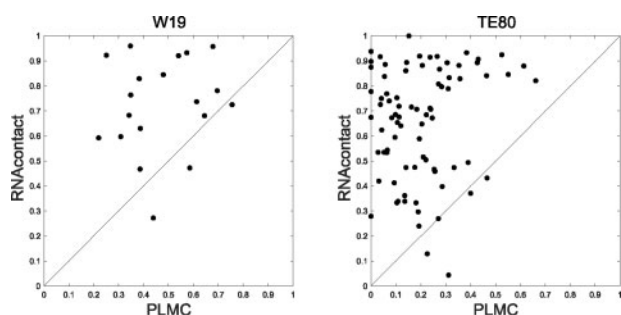**Fig. 2.** Head-to-head comparisons between RNAcontact and PLMC on the test sets W19 and TE80. The x- and y-axis are the precisions of the top L 3D closeness predictions by the corresponding methods

top L predictions by RNAcontact are 0.62, 0.91 and 0.93, respectively. In comparison, the precisions for PLMC are 0.04, 0.04 and 0.38, respectively. Especially, RNAcontact predicts some key 3D closenesses playing critical roles for RNA structure folding. These regions are highlighted in red circles in the 3D closeness maps, which map to the red cartoons in the structures. Meanwhile, RNAcontact correctly detected some long-range 3D closenesses, which are shown in blue circles in 3D closeness maps mapping to the blue cartoons in the structures. In comparison, PLMC only provided partial segment prediction and most of them are located near the diagonal line, which means PLMC mainly predicted some short-range 3D closenesses.

3.4 Secondary structure versus 3D closeness

We divided the 3D closeness defined in Section 2.2 into two categories: SS pairs and non-SS pairs. The SS pairs were generated by the software RNAview (Yang, 2003) from the native structures, which
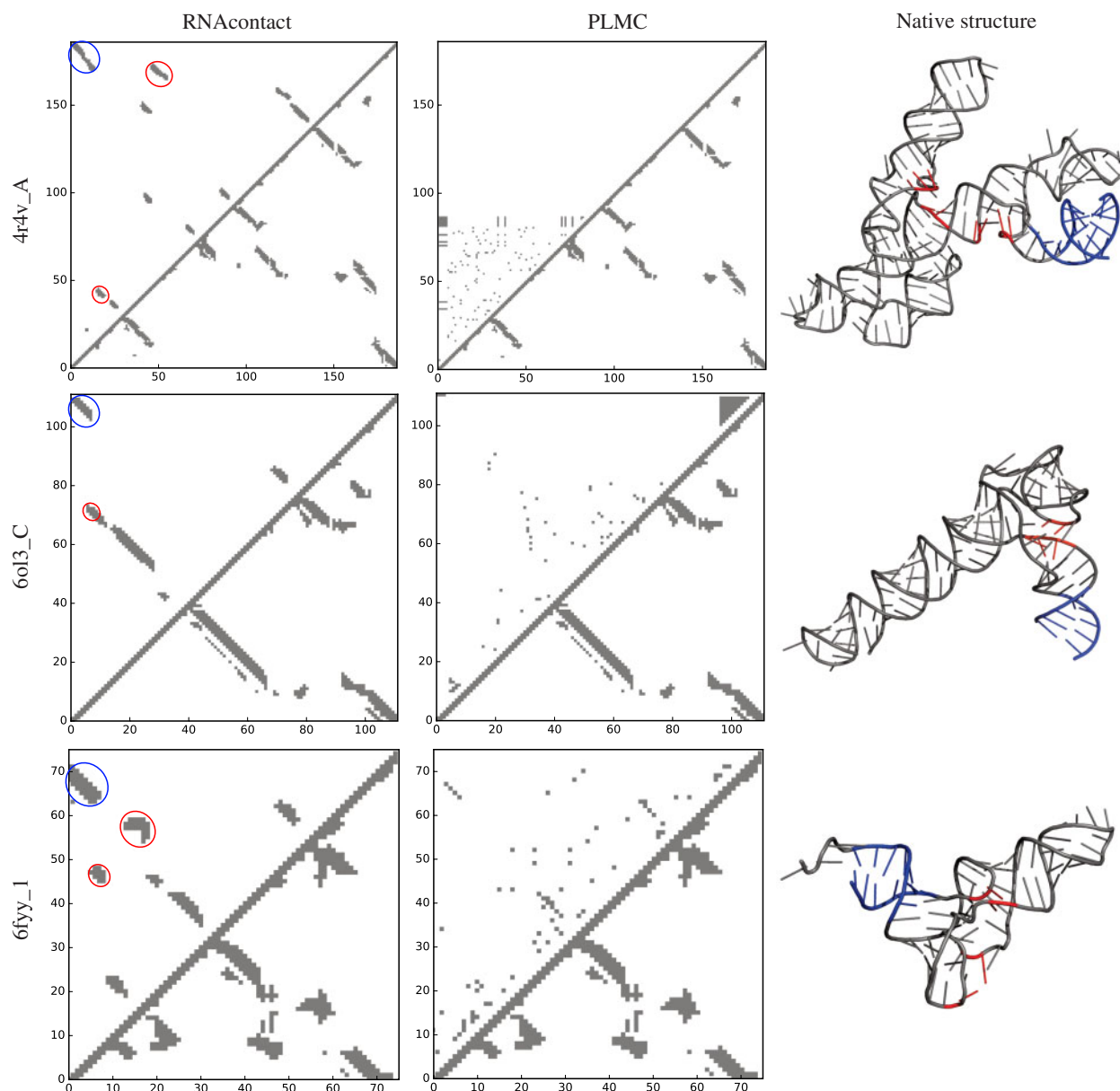

Fig. 3. Comparison of the predictions by RNAcontact and PLMC for three example RNAs. The upper/lower triangle is for the predicted/native 3D closeness maps. The red circles are some key 3D closeness playing critical roles in RNA structure folding predicted by RNAcontact, which correspond to the red regions in the structures. The blue circles highlight some correctly predicted long-range 3D closeness by RNAcontact, which correspond to the blue regions in the structures

included both canonical (e.g. Watson-Crick and Wobble pairs) and non-canonical base pairings. The non-canonical base pairings are SSs with tertiary interactions. Other nucleotide pairs with tertiary interactions but do not form SSs are defined as non-SS pairs.

We compared the proportions of SS and non-SS pairs in the native structures and the correctly predicted 3D closeness from the top L predictions on the dataset TE80. Table 3 shows that 52% native 3D closeness are SS pairs, which means that the remaining 48% non-SS pairs are out of the object of SS prediction, though these nucleotides are close in the structure. Therefore, it is necessary to predict other 3D closeness that are not covered by the SS predictions, to provide more distance information for the subsequent 3D folding. Table 3

Table 3. The division of 3D closeness into SS and non-SS pairs

3D closeness	SS pairs (%)	Non-SS pairs (%)
Native	52	48
RNAcontact	73	27
PETfold	86	14

Note: 'Native' means 3D closeness calculated from the native structures. 'RNAcontact' means the correctly predicted 3D closeness from the top L prediction by RNAcontact. 'PETfold' means the correctly predicted 3D closeness by PETfold.

aslo suggests that 73% predicted 3D closeness are SS pairs, probably because SS-based features were used in RNAcontact. However, a significant proportion (27%) of prediction is non-SS pairs, which is compared with the 14% of PETfold's prediction. This higher proportion of non-SS pairs may be attributed to the contribution from the covariance-based features in the model training.

3.5 Application of predicted 3D closeness in RNA folding

The predicted 3D closeness were used to guide RNA 3D structure modeling to demonstrate the usefulness of predicted 3D closeness. Here, we tested on three RNAs listed in Figure 3 (4r4v_A and 6ol3_C are RNA-Puzzles targets). The 3D modeling package 3dRNA (Wang *et al.*, 2019) was used for RNA structure folding. For a given RNA, we first predicted its 3D closeness by RNAcontact and SS by PETfold. The top L of the predicted long-range 3D closeness were then converted into distance restraints. Details of running 3dRNA was given in the Supplementary Material. The top $L/5$ 3D closeness were also tried, which resulted to models with similar or lower accuracy (Supplementary Fig. S3). The distance restraints and the predicted SS were used to guide 3dRNA to generate 10 structure models. The one with the lowest energy was selected as the final structure model. The modeling by 3dRNA takes about 0.5 h for an RNA with <100 nucleotides.
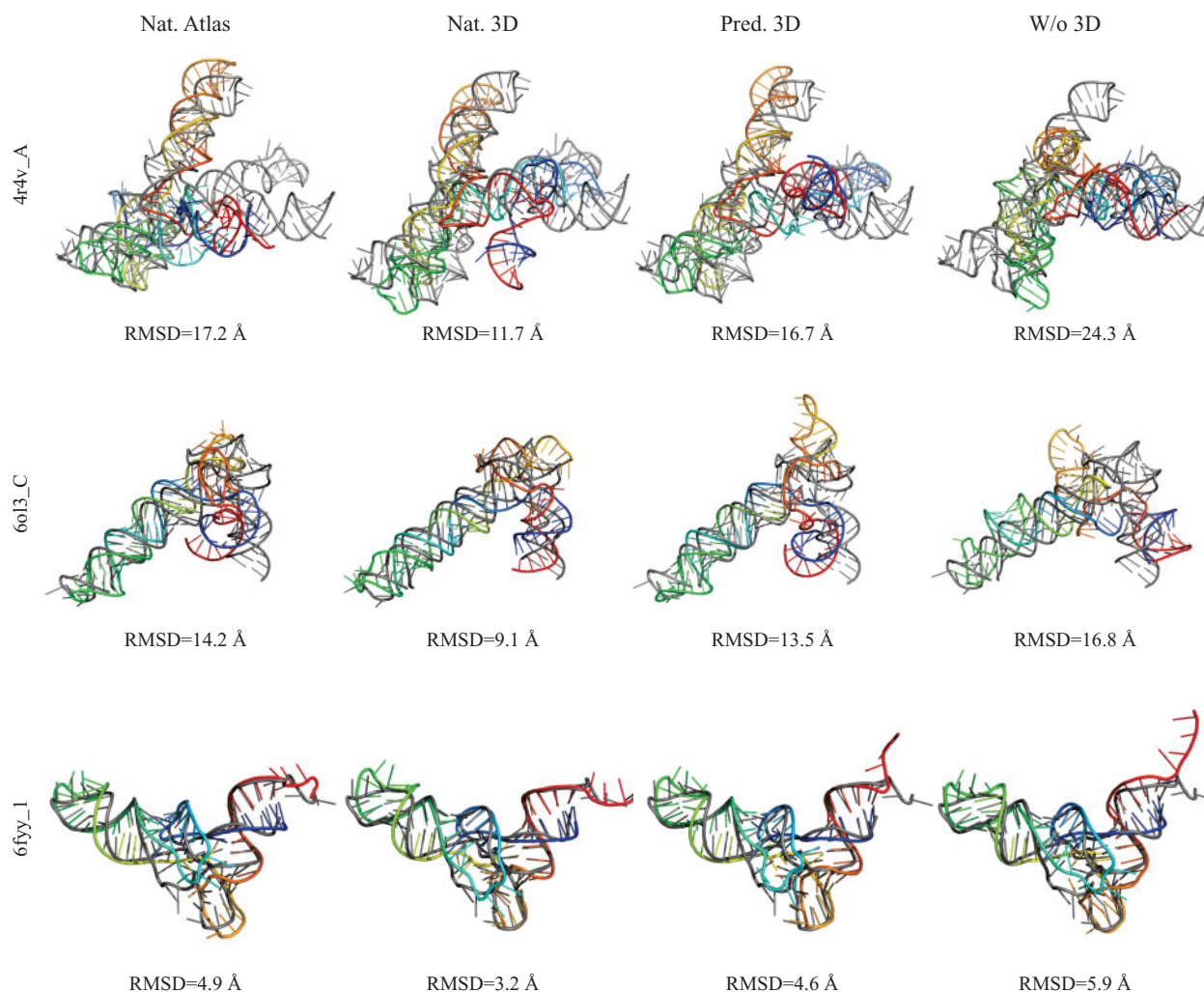

Fig. 4. Three examples of native structures (in gray cartoon) and predicted models (in rainbow cartoon) from the test set TE80. Every column shows structures with different restraints, in which 'Nat. Atlas' means models predicted with 3D closeness defined by RNA Structures Atlas, 'Nat. 3D' means models with native 3D closeness defined based on distance, 'Pred. 3D' means models with predicted 3D closeness and 'W/o 3D' means models predicted without 3D closeness

We compared the effectiveness of restraints from the contact defined by the RNA 3D Structures Atlas (Leontis and Zirbel, 2012) and the 3D closeness defined in Section 2.2. Figure 4 suggests that for these RNAs, the models predicted with native 3D closeness defined in Section 2.2 (the second column) achieve the lower RMSDs than the contact defined by RNA 3D Structures Atlas (the first column). In addition, our analysis suggests the 3D closeness between the nucleotides in red in Figure 3 are not covered in the definition by RNA 3D Structures Atlas. However, these 3D closeness are important to determine the shape of the RNA structure. That is the reason why we defined 3D closeness based on atomic distances rather than RNA 3D Structures Atlas.

The last two columns of Figure 4 shows that models built with predicted 3D closeness are much more accurate than models without 3D closeness. The target 4r4v_A (the first row) is a *varkud satellite ribozyme* with 168 nucleotides, which is a RNA-Puzzles target. Its structure consists of several helical segments with interactions between them. The structure generated with predicted 3D closeness has a similar topology to the native structure with 16.7 Å RMSD, which is more accurate than the best model (20.4 Å RMSD) in the RNA-Puzzles experiment. The target 6ol3_C (the second row) is also from RNA-Puzzles, which is an *adenovirus virus-associated RNA* with 111 nucleotides and two helical segments. The model predicted using 3D closeness has 13.5 Å RMSD, more accurate than the model without 3D closeness (16.8 Å RMSD). 3dRNA also participated to the RNA-Puzzles experiment and predicted structure for the target 6ol3_C with 15.4 Å RMSD. However, the most accurate model from the RNA-Puzzles experiment has 4.8 Å RMSD, suggesting that there are still much room to improve for 3D structure modeling with predicted 3D closeness. The non-RNA-Puzzles target 6fyy_1 (the last row) is from the *yeast 48S ribosome* with 75 nucleotides. This target was predicted with relatively higher accuracy, probably because it is smaller than the other two targets. For this target, the predicted 3D closeness improve the model from 5.9 Å to 4.6 Å in RMSD.

The preliminary tests above shows that the predicted 3D closeness are very helpful to improve RNA 3D structure modeling. However, the way of using predicted 3D closeness in 3D folding has not been systematically optimized and other 3D modeling software such as Rosetta may be tried [such as we demonstrated in the recent work of protein structure prediction (Yang., 2020)], which will be investigated in future.

4 Conclusions

Predicted inter-nucleotide 3D closeness can assist in solving the difficult problem of RNA tertiary structure determination. With covariance features derived from MSAs, we developed a deep learning-based algorithm for predicting the inter-nucleotide 3D closeness. Experiments show that our method outperforms the coevolution-based methods by a large margin. Based on 3D modeling with the 3dRNA package, we demonstrated that by using the predicted 3D closeness as restraints, more accurate RNA structure models can be built than without restraints. However, the best way of applying the predicted 3D closeness in 3D structure folding remains unknown, which will be investigated in future.

Acknowledgements

Funding

References

Abadi,M. *et al.* (2016) *TensorFlow: a system for large-scale machine learning*. OSDI, Savannah, GA, p. 265.

Abriata,L.A. *et al.* (2019) A further leap of improvement in tertiary structure prediction in CASP13 prompts new routes for future assessments. *Proteins*, **87**, 1100–1112.

Altschul,S.F. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Antczak,M. *et al.* (2017) New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochim. Pol.*, **63**, 737–744.

Berman,H.M. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Boniecki,M.J. *et al.* (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, **44**, e63–e63.

De Leonardis,E. *et al.* (2015) Direct-coupling analysis of nucleotide coevolution facilitates RNA secondary and tertiary structure prediction. *Nucleic Acids Res.*, **43**, 10444–10455.

Gruber,A.R. *et al.* (2015) The ViennaRNA web services. *Methods Mol. Biol.*, **1269**, 307–326.

Jian,Y. *et al.* (2019) DIRECT: RNA contact predictions by integrating structural patterns. *BMC Bioinformatics*, **20**, 497.

Jonikas,M.A. *et al.* (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*, **15**, 189–199.

Kandathil,S.M. *et al.* (2019) Prediction of interresidue contacts with DeepMetaPSICOV in CASP13. *Proteins*, **87**, 1092–1099.

Krokhotin,A. *et al.* (2015) iFoldRNA v2: folding RNA with constraints. *Bioinformatics*, **31**, 2891–2893.

Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.

Leontis,N.B. and Zirbel,C.L. (2012) Nonredundant 3D structure datasets for RNA knowledge extraction and benchmarking. In: Leontis,N. and Westhof,E. (eds.) *RNA 3D Structure Analysis and Prediction*. Springer, Berlin, Heidelberg, pp. 281–298.

Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Li,Y. *et al.* (2019) ResPRE: high-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks. *Bioinformatics*, **35**, 4647–4655.

Miao,Z. *et al.* (2017) RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*, **23**, 655–672.

Nawrocki,E.P. and Eddy,S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.

Puton,T. *et al.* (2014) CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. *Nucleic Acids Res.*, **42**, 5403–5406.

Seemann,S.E. *et al.* (2008) Unifying evolutionary and thermodynamic information for RNA folding of multiple alignments. *Nucleic Acids Res.*, **36**, 6355–6362.

Sun,S. *et al.* (2019) Enhanced prediction of RNA solvent accessibility with long short-term memory neural networks and improved sequence profiles. *Bioinformatics*, **35**, 1686–1691.

Wang,J. *et al.* (2017a) Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide-nucleotide interactions from direct coupling analysis. *Nucleic Acids Res.*, **45**, 6299–6309.

Wang,S. *et al.* (2017b) Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput. Biol.*, **13**, e1005324.

Wang,J.et al. (2019) 3dRNA v2.0: An Updated Web Server for RNA 3D Structure Prediction. *Int. J. Mol. Sci.*, **20**, 4116.

Weinreb,C. *et al.* (2016) 3D RNA and functional interactions from evolutionary couplings. *Cell*, **165**, 963–975.

Wu,Q. *et al.* (2020) Protein contact prediction using metagenome sequence data and residual neural networks. *Bioinformatics*, **36**, 41–48.

Xu,J. (2019) Distance-based protein folding powered by deep learning. *Proc. Natl. Acad. Sci. USA*, **116**, 16856–16865.

Yang,H. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.

Yang,J. *et al.* (2020) Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. USA*, **117**, 1496–1503.