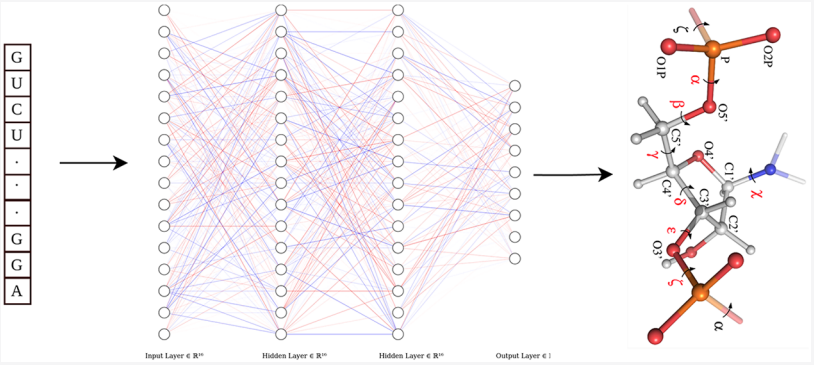
JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

RNA Backbone Torsion and Pseudotorsion Angle Prediction Using Dilated Convolutional Neural Networks

Jaswinder Singh,* Kuldip Paliwal, Jaspreet Singh, and Yaoqi Zhou*

Cite This: J. Chem. Inf. Model. 2021, 61, 2610−2622 Read Online



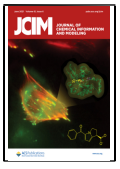
ABSTRACT: RNA three-dimensional structure prediction has been relied on using a predicted or experimentally determined secondary structure as a restraint to reduce the conformational sampling space. However, the secondary-structure restraints are limited to paired bases, and the conformational space of the ribose-phosphate backbone is still too large to be sampled efficiently. Here, we employed the dilated convolutional neural network to predict backbone torsion and pseudotorsion angles using a single RNA sequence as input. The method called SPOT-RNA-1D was trained on a high-resolution training data set and tested on three independent, nonredundant, and high-resolution test sets. The proposed method yields substantially smaller mean absolute errors than the baseline predictors based on random predictions and based on helix conformations according to actual angle distributions. The mean absolute errors for three test sets range from 14°−44° for different angles, compared to 17°−62° by random prediction and 14°−58° by helix prediction. The method also accurately recovers the overall patterns of single or pairwise angle distributions. In general, torsion angles further away from the bases and associated with unpaired bases and paired bases involved in tertiary interactions are more difficult to predict. Compared to the best models in RNA-puzzles experiments, SPOT-RNA-1D yielded more accurate dihedral angles and, thus, are potentially useful as model quality indicators and restraints for RNA structure prediction as in protein structure prediction.

INTRODUCTION

RNA structures are the key for understanding their versatile functions. However, due to the low cost of high-throughput sequencing and the high cost and difficulty of solving RNA structures experimentally, computational prediction of RNA structures becomes increasingly important. Because RNAs fold hierarchically with their secondary structures (stacked base pairs) formed first, secondary structures are often predicted to use as the restraints for modeling tertiary structures. Many computational methods have been developed over the past three decades for RNA secondary structure prediction by folding-based algorithms.[1,2] Recent employment of deep contextual learning leads to a leap in prediction accuracy not only in canonical base pairs but also in the base pairs stabilized by tertiary interactions, including pseudoknots and non-canonical base pairs.[3,4]

Deep learning has also contributed to the recent advance in protein structure prediction.[5−7] This advance was attributed to the use of the restraints from much-improved accuracy in predicted contact maps[8,9] and backbone structures.[10−13] RNA secondary structure prediction corresponds to contact map prediction in proteins.[3] Thus, an interesting question is whether backbone structures can also be predicted with reasonable accuracy. Unlike protein backbone structures determined by two torsional angles, often with regular helical and sheet patterns, RNA ribose-phosphate backbone structures are characterized by six torsion angles without obvious regular structural patterns. The six backbone torsion angles (see Figure

Received: February 9, 2021
Published: May 26, 2021



https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

1) are rotations about bonds P_i-O5_i' (α), $O5_i'-C5_i'$ (β), $C5_i'-C4_i'$ (γ), $C4_i'-C3_i'$ (δ), $C3_i'-O3_i'$ (ϵ), and $O3_i'-P_{i+1}$

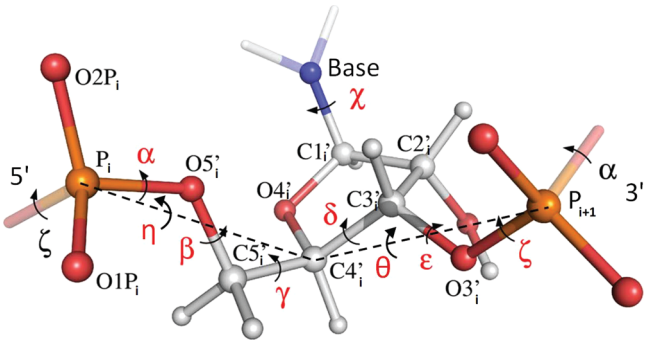


Figure 1. RNA backbone torsion and pseudotorsion angles. Diagram adapted with permission from Frellsen et al.[15] Copyright 2009 The Authors under Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

(ζ). In addition to these angles, there is another angle chi (χ) that describes the rotation of the base relative to the sugar ($C1_i'$−Base) as shown in Figure 1. The RNA backbone can also be simplified by two pseudotorsion angles (η and θ) only without losing much information,[14] as shown by the dotted lines in Figure 1. Angle eta (η) is rotation about pseudobond P_i-C4_i', and angle theta (θ) is rotation about pseudobond bond $C4_i'-P_{i+1}$. All these angles vary from −180° and 180°.

In this work, we developed the first method to predict RNA backbone torsion and pseudotorsion angles by using deep contextual learning based on single-sequence information only, inspired by previous work for the RNA secondary structure[3] and solvent accessibility prediction.[16−18] The method SPOT-RNA-1D employed a dilated convolutional neural network architecture, which can predict seven torsions and two pseudotorsion angles. It achieves mean absolute error (MAE) between 14° and 44° for nine angles that are substantially lower than the random-baseline predictor (17°−62°) and the helix-baseline predictor (14°−58°) according to three independent test sets. Moreover, the predicted angles are even more accurate than those angles provided by the best RNA-puzzles models. Thus, these predicted angles will likely be useful as new restraints for RNA structure prediction as well as structure-quality assessment.

METHODOLOGY

Data Sets. The data sets used for SPOT-RNA-1D training, validation, and testing were obtained from the protein data bank (PDB).[19] More specifically, we downloaded all the RNA structures with X-ray resolution <3.5 Å from PDB on October 3, 2020. All the PDB structures were split into individual chains using a program from Biopython.[20] All these PDB chain sequences were clustered using CD-HIT-EST[21] with the lowest allowed identity cutoff of 80%. RNA sequences without any clusters were kept for a noncluster set, and the remaining sequences were kept for the training set (TR).

As an 80% sequence identity cutoff may not be strict enough, we further employ the BLAST-N[22] tool on noncluster sequences against the training set and within itself with an e-value cutoff of 10. Any sequences in the training set having hits with the noncluster sequences were removed. Furthermore, any sequences having hits within themselves were also removed.

The noncluster RNA sequences obtained after CD-HIT-EST and BLAST-N processing were randomly divided into one validation (VL) and two test sets (TS1 and TS2). To make VL and TS2 nonredundant from the training set and within itself even at the remote homologue level, we built the covariance models for RNAs in VL and TS2. To build a covariance model, each sequence from VL and TS2 was searched against NCBI's database[23] for homologous sequences using the BLAST-N tool with the e-value cutoff of 10 and the maximum number of 50,000 aligned sequences. Next, the covariance model was built using the cmbuild program in the INFERNAL[24] tool from the BLAST-N alignment along with a consensus secondary structure derived from the PDB files. These covariance models were searched against the training data to find any remote homologues using the cmsearch program in INFERNAL with an e-value cutoff of 0.1 for VL and 10 for TS2 set. We employed 10 for TS2 to purposely make this test set as hard as possible by removing all potential remote homologues while using 0.1 for VL is to avoid removal of too many training RNAs because any hits with the training data were removed from the training set. The same procedures were used to make VL and TS2 RNA sequences nonredundant within themselves and from TS1.

The final training (TR), validation (VL), and two test sets (TS1 and TS2) have 286, 30, 63, and 30 RNA chains, respectively. The number of RNAs in each data set, median and maximum sequence lengths, and the number of base pairs in each data set are shown in Supplementary Table S1. The DSSR[25] program was used to calculate native torsion angles and identify base pairs and secondary structures from 3D structures.

While preparing the PDB data sets, we purposely kept the maximum number of RNA chains in test sets (TS1 and TS2) that belong to the RNA-Puzzles benchmarking test set.[26−29] The RNA-Puzzles data set is widely used for benchmarking RNA tertiary structure prediction tools. There were 12 RNA-Puzzles chains in TS1 and TS2 that were nonredundant from the training (TR) and validation (VL) sets. We used these 12 RNAs as a separate test set (named RNA-Puzzles) for benchmarking SPOT-RNA-1D.

In addition to the above three test sets (TS1, TS2, and RNA-Puzzles), we prepared another test set (TS3) by downloading all the NMR structures (707 RNAs) from the PDB on April 5, 2021. After removing redundancy within TS3 and from all other RNA structures (TR, VL, TS1, and TS2) using the exact same specifications as TS1, we obtained 54 nonredundant RNA chains in TS3. The native angles for TS3 were obtained from their 3D Model-1 structures using the DSSR program. The length and base pair information for TS3 is shown in Supplementary Table S1.

Besides the comparison to native angles, we also compare predicted angles with the angles of the models predicted for RNA-puzzles. We downloaded 3D structural files of the best five (based on RMSD) RNA 3D structure predictors from the RNA-Puzzles repository (https://github.com/mmagnus/RNA-Puzzles-Standardized-Submissions, Release-1.2). Among 12 RNAs, the structural files for four RNAs were missing, and there was a chain mismatch in two RNAs. Thus, we obtained complete structural files for only six RNAs (out of 12). There were up to ten predicted structural models for each puzzle per predictor, and we considered only the most accurate model based on the RMSD (*w.r.t.* native structure) for each predictor to compare with SPOT-RNA-1D.

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

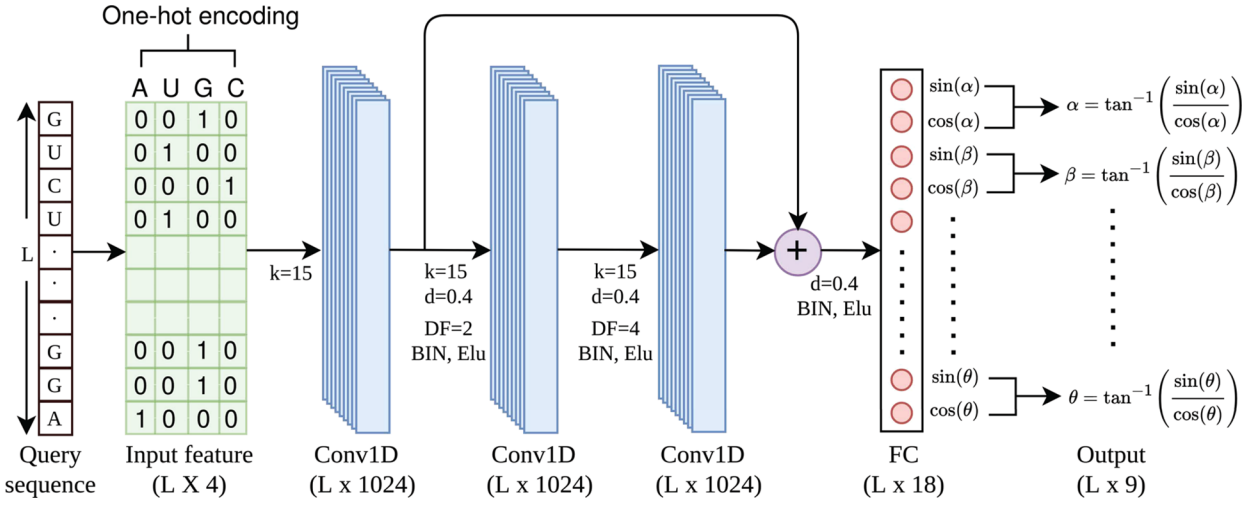


Figure 2. Network architecture of SPOT-RNA-1D. *k*, *d*, *DF*, and *BIN* are the size of the filter, dropout rate, dilation factor, and batch instance normalization, respectively, and *L* is the length of the input RNA. Scalar 4 and 1024 represent the number of features per nucleotide and the number of filters in each convolutional layer, respectively.

We also downloaded the RNAPOT[30] data set of 85 RNA structures from http://melolab.org/supmat/RNApot/Sup._Data.html. In addition to native structures, each RNA has 500 non-native structures (decoys) built by MODELLER[31] with a set of Gaussian restraints for dihedral angles and atom distances. From 85 RNAs, there were only ten nonredundant RNAs from our training data with the exact same specification as TS1 and TS3. These ten RNAs were also a subset of TS1 and TS3. We used these ten RNAs to examine the relation between the errors of predicted angles and the model accuracy according to the root-mean-square deviation (RMSD) and global distance test (GDT)-score compared to the native structure.

Dilated Convolutional Neural Network. The neural network architecture of SPOT-RNA-1D (shown in Figure 2) was inspired by our previous work of RNA solvent accessibility predictor RNAsnap2.[16] Similar to RNAsnap2, we employed a residual dilated convolutional neural network,[32,33] which is better at learning long-range interactions between nucleotides as compared to standard ResNets[32] and LSTMs[34] neural networks.

The SPOT-RNA-1D neural network architecture (shown in Figure 2) consists of an initial 1D (one-dimensional) convolutional layer (Conv1D) with 1024 filters and a kernel size of 15. After the initial convolutional layer, we employed a dilated residual neural network (ResNet), which consists of two 1D convolutional layers (Conv1D) with 1024 filters and a kernel size of 15. Dilation rates/factors (*DF*) of two and four were utilized in the first and second layers of the ResNet block, respectively. The fully connected (FC) layer was used for output with 18 nodes.

The input to each layer (except for the initial convolutional layer) was activated with an exponential linear unit (*Elu*)[35] activation function and normalized using the batch instance normalization (*BIN*) technique.[36] To avoid overfitting of model parameters on the training data, a dropout rate (*d*) of 40% was used before each layer except for the initial convolutional layer during training. The order of operations was normalization (*BIN*), activation (*Elu*), and, finally, dropout (*d*) before each layer.

SPOT-RNA-1D was implemented in Google's deep learning framework TensorFlow Version 1.15.[37] The model parameters (shown in Figure 2) were learned using the RMSProp[38] back-propagation algorithm with a learning rate of 0.001 and the Mean Squared Error (MSE) as a loss function. The model was trained using a minibatch size of 8 for 25 epochs. Due to the variable-length sequences, minibatches were padded with zeroes up to the length of the longest sequence in the minibatch. To minimize the effect of padding, we mask the padded zeroes from the output of every layer of the neural network architecture during the training. The model hyper-parameters such as the kernel size (*k*), the number of filters, the activation function, the normalization technique, the dropout rate (*d*), the dilation factor (*DF*), the number of convolutional layers (Conv1D), the choice of the back-propagation algorithm, and the learning rate were optimized based on the model's performance on the validation set (VL).

Input. The input to the SPOT-RNA-1D is an RNA sequence represented by a binary one-hot vector of size *L*×4 as shown in Figure 2, where *L* is the length of the RNA sequence, and four corresponds to the number of base types (A, U, G, and C). In one-hot encoding, a value of 1 was assigned to the corresponding base-type position in the vector and 0 elsewhere. One-hot encoded input features were standardized to have a zero mean and unit variance according to the training data before inputting to the neural network model.

Output. The SPOT-RNA-1D model attempts to predict nine backbone angles. Because these angles vary from −180° to 180°, we predicted the sine and cosine of the angles instead of predicting angles directly to remove the effect of angle periodicity. Thus, the output FC layer shown in Figure 2 has 18 nodes, i.e., two nodes per angle. The predicted sine and cosine values from the model were converted back to angles by using $angle = tan^{-1}\left(\frac{\sin(angle)}{\cos(angle)}\right)$. Such transformation is commonly used for protein torsion angle prediction using a deep neural network.[10,11,39]

Performance Evaluation. For the performance evaluation, we used the mean absolute error (MAE), which is the average absolute difference between predicted and experimentally determined angles in degrees. The periodicity of the

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

Table 1. Performance Comparison in Terms of Mean Absolute Errors (MAEs) on VL (30 RNAs), TS1 (63 RNAs), TS2 (30 RNAs), TS3 (54 RNAs), and RNA-Puzzles (12 RNAs) Sets by SPOT-RNA-1D and a Random-Baseline Predictor[a]

	standard backbone torsion angles							virtual torsion angles	
	alpha (α)	beta (β)	gamma (γ)	delta (δ)	epsilon (ϵ)	zeta (ζ)	chi (χ)	eta (η)	theta (θ)
SPOT-RNA-1D									
VL	45.18	20.58	33.88	17.99	20.72	37.50	23.01	33.55	37.02
TS1	43.94	21.94	32.98	14.61	20.69	33.27	19.59	30.25	32.91
TS2	39.50	18.92	29.47	16.01	17.46	28.91	18.20	28.14	30.25
TS3	37.89	21.04	34.68	13.83	22.32	27.87	17.01	25.31	27.22
RNA-Puzzles	39.29	19.71	30.08	16.60	18.73	33.76	19.99	31.58	36.09
Random-Baseline Predictor									
VL	61.59	30.17	48.68	19.32	27.45	46.75	28.83	39.92	44.92
TS1	62.15	31.25	48.79	16.66	28.44	44.21	26.27	39.34	42.63
TS2	60.42	30.06	46.79	17.80	27.17	43.86	26.17	38.20	42.69
TS3	60.17	32.03	50.57	17.80	28.94	43.32	25.56	35.47	40.86
RNA-Puzzles	59.63	29.12	46.45	16.96	25.72	43.63	25.72	38.73	43.29

[a]An average of over 100 random predictions.

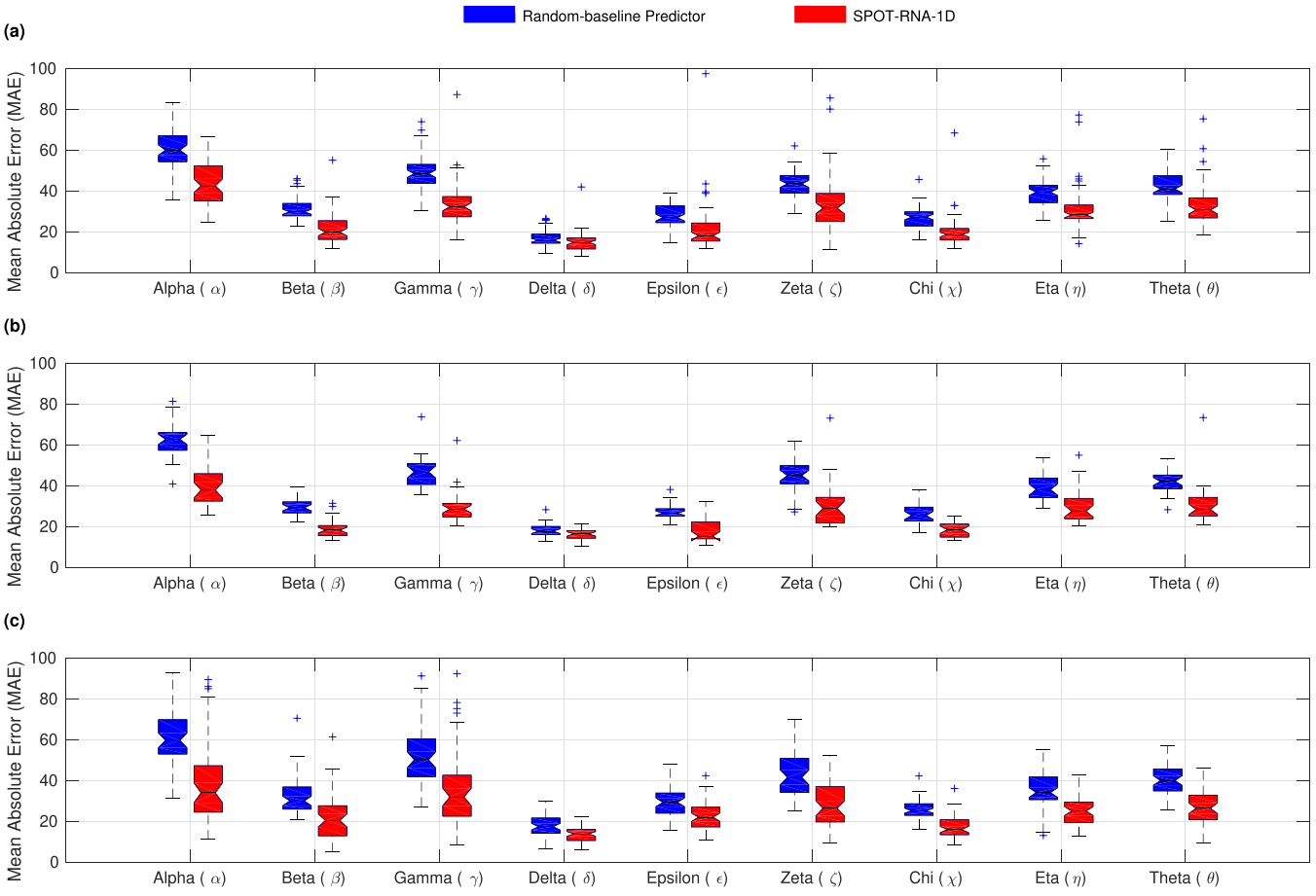


Figure 3. Distribution of mean absolute errors (MAEs) for individual RNA chains on test sets (a) TS1, (b) TS2, and (c) TS3 by a random-baseline predictor (in blue) and by SPOT-RNA-1D (in red). On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The outliers are plotted individually by using the "+" symbol.

angles was taken into consideration by utilizing the smaller value of the absolute difference $d_i = |angle_i^{Pred} - angle_i^{Actual}|$ and $360 - d_i$ for average. To obtain the statistical significance (P-value) of improvement made by SPOT-RNA-1D, we performed a one-tailed paired t-test. The smaller the P-value is, the more significant improvement by SPOT-RNA-1D over the baseline predictor. Here, the baseline predictor is a random prediction according to the native distribution of specific angles generated from training (TR) RNA structures. Each angle was randomly predicted 100 times, and the average mean absolute error of 100 predictions was reported. We also introduced a baseline predictor based on A-form helical conformations by random predictions derived from the distribution of angles in the helical regions of the training data only.

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

Table 2. Performance According to Mean Absolute Errors (MAEs) by SPOT-RNA-1D for Nucleotides in Different Pairing Interactions on Test Sets TS1, TS2, and TS3 along with MAEs by the Random-Baseline Predictor Inside the Parentheses[a]

	standard backbone torsion angles							virtual torsion angles	
	alpha (α)	beta (β)	gamma (γ)	delta (δ)	epsilon (ϵ)	zeta (ζ)	chi (χ)	eta (η)	theta (θ)
				TS1					
unpaired (20.92%)	62.05 (71.43)	28.43 (37.45)	39.76 (52.93)	20.53 (21.69)	29.17 (34.66)	54.10 (56.27)	30.83 (36.88)	54.58 (58.63)	56.70 (58.57)
lone pairs (15.85%)	54.26 (67.12)	27.96 (33.34)	40.61 (53.47)	20.93 (23.38)	28.32 (35.35)	51.64 (56.42)	27.18 (34.40)	42.76 (50.61)	44.52 (52.82)
pseudoknots (5.80%)	54.16 (67.55)	23.19 (30.58)	38.42 (51.00)	20.02 (22.93)	25.76 (32.20)	47.95 (57.13)	21.91 (31.07)	45.04 (53.17)	42.15 (52.45)
multiplets (7.42%)	49.92 (62.41)	27.52 (31.85)	40.73 (52.83)	16.97 (18.07)	24.52 (30.69)	36.84 (44.26)	22.19 (28.65)	34.84 (39.71)	33.80 (43.62)
noncanonical pairs (18.11%)	46.57 (60.71)	23.32 (32.01)	35.08 (50.24)	16.45 (17.82)	22.92 (30.31)	40.33 (48.26)	22.37 (29.47)	33.47 (39.74)	38.93 (46.89)
canonical nested pairs (49.84%)	35.02 (56.41)	18.32 (28.87)	29.17 (47.83)	10.59 (12.78)	15.46 (24.40)	19.63 (34.12)	13.43 (20.08)	18.24 (29.62)	20.02 (33.69)
				TS2					
unpaired (20.54%)	60.26 (75.35)	24.49 (34.26)	37.89 (52.87)	22.74 (24.23)	26.92 (36.73)	48.34 (59.61)	27.67 (36.67)	52.92 (60.72)	52.40 (60.04)
lone pairs (15.76%)	53.62 (68.11)	26.17 (35.11)	39.49 (55.66)	23.39 (27.99)	24.20 (32.01)	50.20 (63.45)	28.00 (34.55)	39.04 (47.03)	44.06 (52.25)
pseudoknots (6.63%)	57.40 (72.32)	20.98 (31.01)	35.46 (55.61)	17.88 (20.31)	18.47 (28.68)	32.74 (52.02)	21.06 (24.98)	42.38 (52.20)	35.54 (47.27)
multiplets (6.93%)	50.43 (71.02)	22.69 (32.94)	36.66 (57.00)	21.13 (24.37)	18.20 (28.03)	41.04 (54.77)	24.32 (28.24)	33.45 (44.38)	34.03 (47.26)
noncanonical pairs (15.89%)	44.32 (63.42)	19.78 (29.59)	31.99 (49.38)	20.08 (22.56)	19.43 (29.90)	43.14 (58.48)	23.74 (30.59)	32.08 (40.75)	38.65 (46.60)
canonical nested pairs (51.68%)	28.03 (55.66)	15.14 (27.36)	23.91 (41.69)	11.43 (13.06)	12.30 (22.50)	14.73 (34.01)	11.72 (19.75)	16.23 (28.24)	17.05 (32.55)
				TS3					
unpaired (18.12%)	68.62 (76.67)	33.74 (41.52)	58.96 (72.33)	23.79 (30.11)	32.83 (37.70)	61.55 (69.92)	27.66 (36.01)	54.43 (58.87)	58.84 (62.14)
lone pairs (9.20%)	51.32 (65.07)	31.53 (40.96)	50.42 (61.16)	17.84 (24.95)	27.01 (32.63)	40.07 (49.95)	23.59 (31.36)	30.11 (40.58)	36.86 (49.90)
pseudoknots (3.27%)	26.06 (56.90)	23.42 (35.84)	32.84 (53.97)	11.67 (15.72)	21.40 (27.24)	29.20 (36.91)	15.42 (20.61)	22.02 (33.30)	21.53 (36.01)
multiplets (6.41%)	46.16 (63.47)	29.95 (39.68)	46.50 (61.00)	16.22 (22.83)	24.73 (31.50)	35.15 (50.11)	20.40 (27.59)	28.60 (36.80)	29.70 (43.33)
noncanonical pairs (13.46%)	52.68 (69.54)	28.72 (38.67)	47.25 (59.41)	15.50 (21.46)	22.94 (28.82)	34.95 (48.30)	21.01 (28.12)	27.14 (35.44)	32.89 (49.01)
canonical nested pairs (60.14%)	25.44 (53.54)	15.40 (27.50)	24.61 (42.49)	10.75 (13.39)	18.68 (26.08)	15.99 (34.54)	12.77 (21.88)	16.01 (28.29)	16.21 (33.03)

[a]The number mentioned with the base pairing type is the percentage of total nucleotides in that region.

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

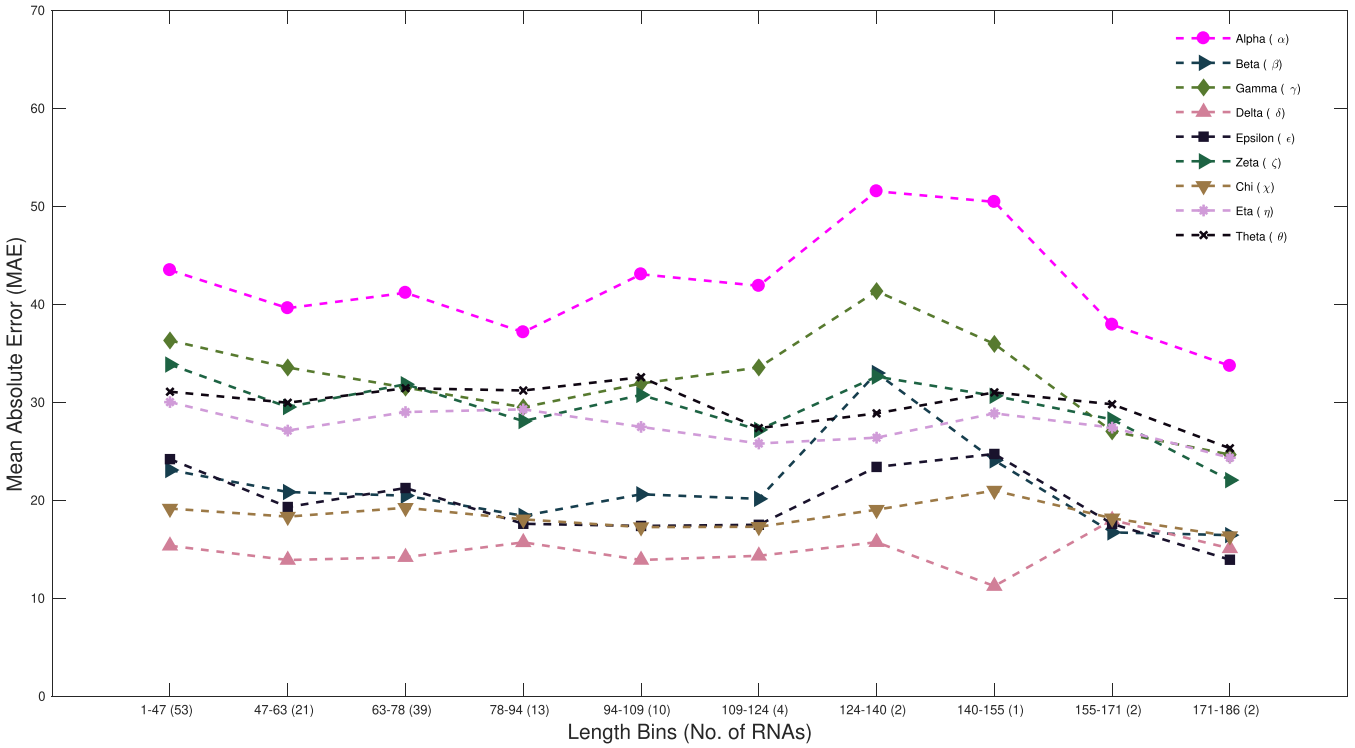


Figure 4. Mean absolute errors (MAEs) as a function of RNA sequence length for nine angles on combined 147 RNAs from three test sets (TS1, TS2, and TS3).

RESULTS

The SPOT-RNA-1D model (shown in Figure 2) was trained using the training set TR, validated using the validation set VL, and tested on three independent test sets TS1, TS2, and TS3. Table 1 and Figure 3 show the performance comparison between SPOT-RNA-1D and the random-baseline on validation (VL) and three test sets (TS1, TS2, and TS3) along with a separate list of RNAs in RNA-puzzles on seven standard torsion angles (α, β, γ, δ, ϵ, ζ, and χ) and two pseudo/virtual torsion angles (η and θ). Both baseline and SPOT-RNA-1D predictors indicate a small error for delta (δ), epsilon (ϵ), beta (β), and chi (χ), followed by gamma (γ) and zeta (ζ), with alpha (α) being the most difficult angle to predict. The same trend indicates that the distribution of an angle largely controls its level of difficulty in prediction. We also observed that torsion angles closer to the base (χ, ϵ) are easier to predict than those away from the base (α, γ, ζ). Exceptions are angles beta (β) and delta (δ). They were easier to predict due to their narrower distributions not completely spanned from $-180°$ to $180°$ (as shown in Supplementary Figure S1). Nevertheless, SPOT-RNA-1D leads to more than a 20% improvement over random prediction except for delta (δ), where about a 10% improvement was observed. For virtual torsion angles (η and θ), an MAE of around 30° by SPOT-RNA-1D was obtained for both test sets (TS1 and TS2), which is also more than a 20% reduction from the random prediction. Virtual torsion angles are relatively more difficult to predict than the standard torsion angles because of the distributions. For the neural network architecture (in Figure 2), wide and evenly distributed angles (α, γ, ζ, η, and θ) are more difficult to predict than uneven and narrowly distributed angles (β, δ, ϵ, and χ) as shown in Table 1 and Supplementary Figure S1. Similar performance trends were observed for the TS3 (NMR structures) and RNA-Puzzles test sets, indicating the model's robustness for different types of test sets. These performance improvements are statistically significant according to the *P*-values obtained through a paired *t*-test for respective angles, as shown in Supplementary Table S2.

One interesting observation from Table 1 is that there is a systematic trend that MAE values for all angles are the largest for VL and the smallest for TS3 (except β, γ, and ϵ) with the second-lowest (or the lowest for β, γ, and ϵ) for TS2. This trend is less obvious for the random prediction. This observation is somewhat counterintuitive because VL was optimized for performance, and both VL and TS2 do not have any remote homologues in the training set. By comparison, we only exclude close homologues in TS1 and TS3 from the training set to maintain a reasonable number of RNAs. If there were overfitting, we would expect that the MAEs for VL were the smallest and followed by TS1/TS3 and TS2. The fact that the opposite is true suggests the robustness of the method trained. An inspection of each data set shows that the VL set has the most tertiary interactions (23.7% noncanonical base pairs and 30.4% multiplets), followed by TS1 (22.7% noncanonical base pairs and 26.4% multiplets), TS2 (20.2% noncanonical base pairs and 25.4% multiplets), and TS3 (17.0% noncanonical base pairs and 21.3% multiplets) as shown in Supplementary Table S1. In other words, it seems that fewer tertiary interactions make angles easier to predict.

To confirm this, we compare the performance of prediction in TS1, TS2, and TS3 according to the nucleotides associated with tertiary interactions (lone pairs, pseudoknots, multiplets, and noncanonical pairs) and those are not (canonical nested base pairs and nucleotides not paired). As shown in Table 2, the angles for the nucleotides involved in the tertiary interaction are indeed more difficult to predict than the nucleotides in canonical nested interactions. Therefore, more tertiary interaction in an RNA increases the difficulty of

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

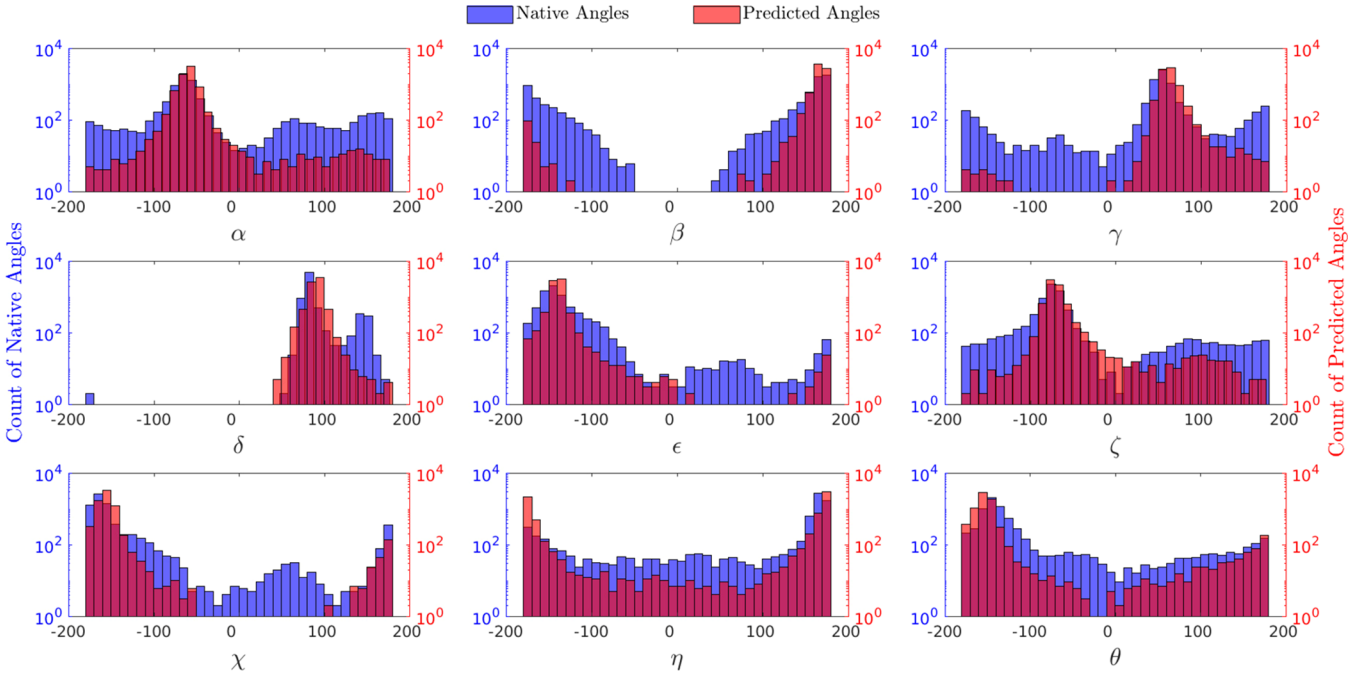


Figure 5. Distribution plots of native (in blue) and predicted (in red) torsion angles on combined test sets TS1 and TS2.

backbone angle prediction. Of course, the angles associated with unpaired bases are the most difficult to predict because of higher flexibility. All three data sets have a similar level of unpaired bases (20.92%, 20.92%, 20.54% for VL, TS1, and TS2, respectively). TS2 also has slightly more pseudoknots (6.6%) than TS1 (5.8%). This difference is compensated by the lower factions of noncanonical base pairs and multiplets in TS2. A similar trend is overserved for TS3 except for the pseudoknot region. This is perhaps because the test set TS3 had a significantly different distribution of nucleotides in structural regions as compared to VL, TS1, and TS2 sets. TS3 structures are derived from distance restraints obtained by the NMR technique, whereas the other data sets' structures are from X-ray crystallography.

Table 2 also demonstrates MAEs of the random-baseline predictor for different interactions. As shown by the numbers specified inside the parentheses of Table 2, a similar difficulty is observed for the baseline predictors with the maximum for the unpaired nucleotides and minimum for the nucleotides in canonical nested pairs. Overall, SPOT-RNA-1D has consistently outperformed the random-baseline predictor for different pairing interactions and unpaired nucleotides.

Because the maximum sequence length (64) of RNA in the validation set (VL) is significantly less than those in the training (418) and test sets (186), it is of interest to know if this difference would lead to poorer performance for long RNAs. Figure 4 plotted the performance of SPOT-RNA-1D for nine angles as a function of sequence length. Some angles have the largest MAE at RNA lengths between 120 and 150. Others have essentially the same MAE across all lengths. Thus, different lengths of RNA between the validation set and the test sets do not lead to a systematic bias.

The distribution of native and predicted angles from SPOT-RNA-1D on combined test sets TS1 and TS2 is shown in Figure 5. The predicted angles (in red) follow the distribution of native angles (in blue). For instance, the distribution of native α angle peaks at around $-70°$, and predicted angles distribution also peaks around the same point. Similarly, the distribution of other predicted angles is also in good agreement with native angles. In addition to a one-dimensional (1D) distribution plot, Supplementary Figures S2 and S3 show two-dimensional (2D) Ramachandran-like plots of two standard torsion angles (α and β) and pseudotorsion angles (η and θ). In both plots, SPOT-RNA-1D was able to detect angles in different regions of native distributions.

Although SPOT-RNA-1D can detect angles in most of the regions of their native distributions, there were still a few regions in angles beta (β), gamma (γ), epsilon (ϵ), and chi (χ) that are not in their respective predicted distributions as shown in Figure 5. There may be a possibility that these predicted angles are more like those in an A-form helix region only. As most RNA helices are of the A-form, we compared the performance of SPOT-RNA-1D with random predictions derived from the distribution of angles in the helix regions of training data only (the helix-baseline predictor). As shown in Supplementary Table S3, SPOT-RNA-1D consistently outperformed the helix-baseline predictor except for angle delta (δ). In this case, the helix-baseline predictor is slightly more accurate. This may be because the delta angle distribution is narrow and most accurately predicted and helical and nonhelical regions have indistinguishable angles.

To understand the relationship between MAE of torsion angles, we plotted a correlation matrix of torsion angles shown in Figure 6 on combined test sets TS1 and TS2. In general, MAEs of torsion angles are highly correlated with 1−2 neighboring angles and less correlated with those angles further away. Interestingly, the MAEs of angle chi (χ) strongly correlate (CC > 0.67) with essentially all angles except the one furthest away (angle α), and the MAEs of angle beta (β) strongly correlate (CC > 0.66) with essentially all angles except the angle delta (δ), whereas the angle delta (δ) does not show much correlation with other torsion angles except for the angle chi (χ), likely due to the delta angle being restricted due to the ribose ring structure. The correlation network shown in Figure

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

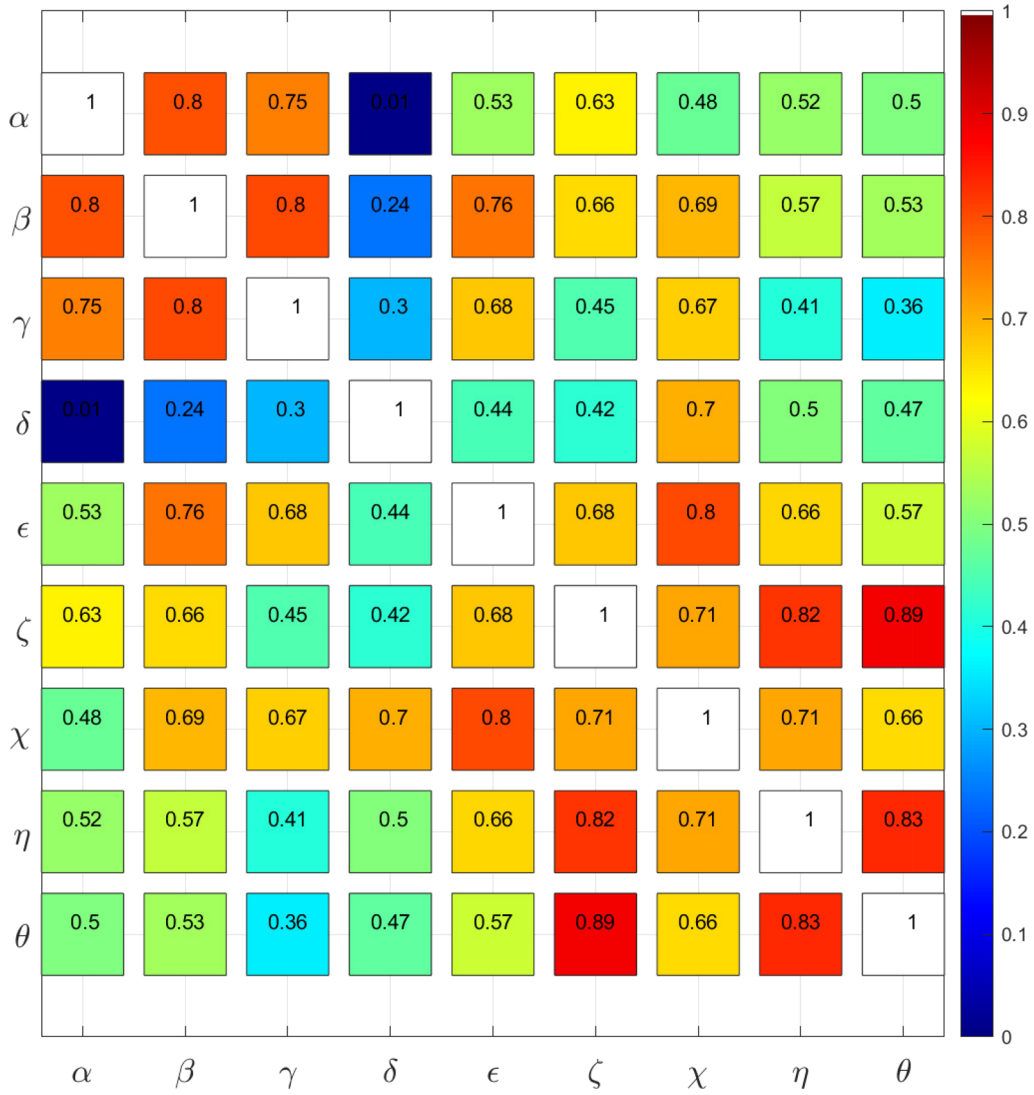


Figure 6. Correlation coefficients (CCs) of mean absolute errors (MAEs) between torsion angles on combined test sets TS1 and TS2.

6 is consistent with the expectation that the base pairing is the dominant determinant of backbone angles. It is of interest to note that the errors for pseudotorsion angles (η and θ) correlate not only with each other but also strongly with zeta (ζ).

To further assess the performance of SPOT- RNA-1D, we obtained the same angles derived from the best models from the best five predictors of 6 RNA-Puzzles that are non-redundant from our training data. As shown in Supplementary Table S4, SPOT-RNA-1D consistently produces the most accurate alpha, beta, gamma, and epsilon angles in most cases and comparable results for other angles.

The fact that most predicted angles are more accurate than the angles from best-predicted models suggests the possibility of using the difference between predicted angles and model angles as a quality assessment score. To demonstrate this possibility, we obtained the RNAPOT data set, which contains a nearly continuous distribution of decoy models with different levels of accuracy. Figure 7a,b plots the mean absolute errors between SPOT-RNA-1D angles and the angles from decoy model structures as a function of structural accuracy represented by RMSD and GDT-score, respectively, for PDB ID 2zni (Chain C). The MEA is consistently increasing with RMSD and decreasing with GDT-score. Similar trends were observed for the remaining nine RNAs of the RNAPOT data set, as shown in Supplementary Figures S4 and S5. This indicates that the deviation from SPOT-RNA-1D predicted angles could be used for model quality assessment.

One interesting question is that at what MAE level a predicted angle will be useful enough to judge the model quality. Figure 8 shows the MAE between the native angles and angles from the RNAPOT models for an rRNA (PDB ID: 1i6u, Chain C) as a function of RMSD. The remaining nine RNAs are shown in Supplementary Figures S6−S14. Except for one RNA (tRNA in PDB ID 2dr2), all results indicate smaller angle errors are closer to native structures. For the exception (tRNA in PDB ID 2dr2), RMSD at 4.32 Å for near-perfect angles is caused by a large error introduced by the last five nucleotides at the terminal (see Supplementary Figure S15). The average MAE for these five nucleotides was 52.47 compared to 13.34 for all nucleotides in this RNA. The majority region of the structure is nearly perfect, as shown in Supplementary Figure S15. The results indicate that all angles are useful for indicating near-native models (5 Å RMSD or less) if its error is <60° for alpha (α), gamma (γ), delta (δ), and zeta (ζ), <40° for beta (β) and epsilon (ϵ), and <20° for

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

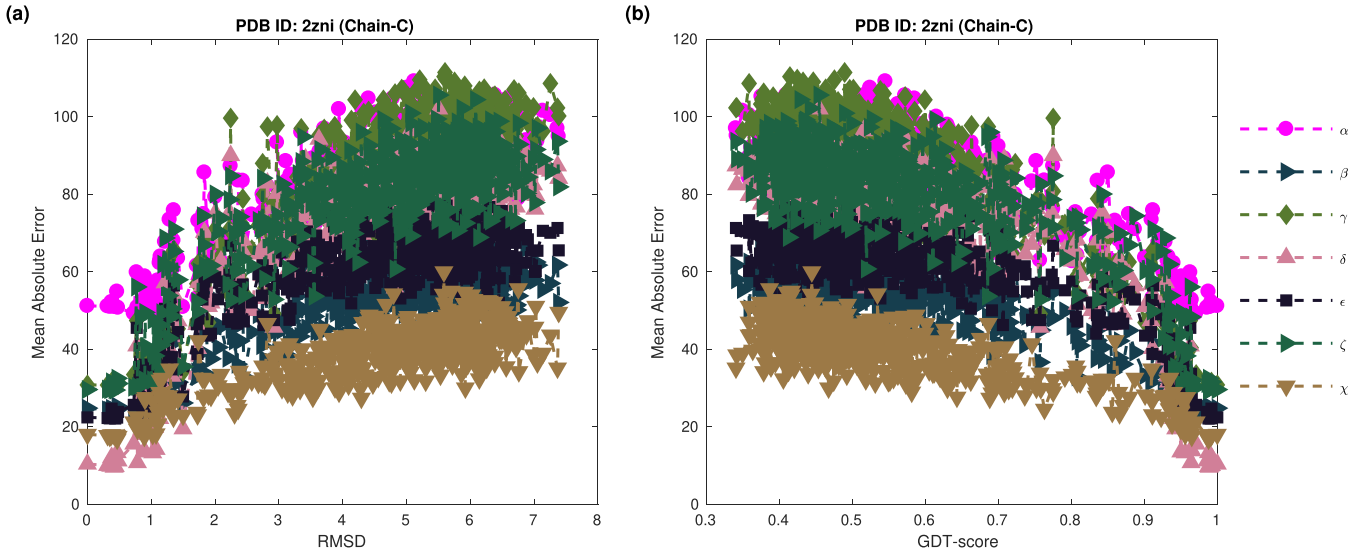


Figure 7. (a) MAE vs RMSD scatterplot of RNA 2zni (Chain C), where MAE is evaluated from SPOT-RNA-1D angles and the angles derived from a 3D model structure for different values of RMSD from the native 3D structure. (b) MAE vs GDT-score scatterplot of RNA 2zni (Chain C), where MAE is evaluated from SPOT-RNA-1D angles and the angles derived from a 3D model structure for different values of GDT-score from the native 3D structure.

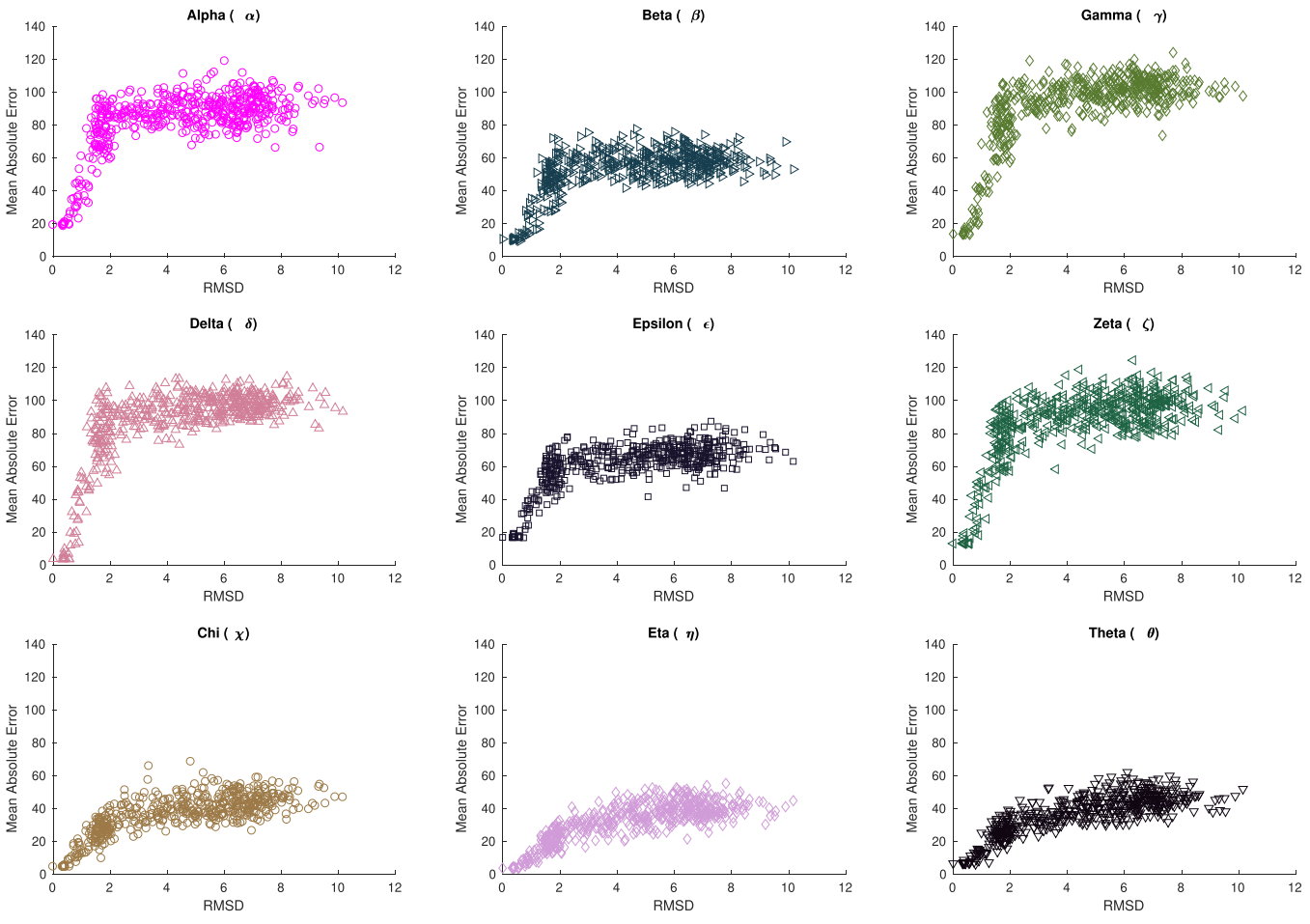


Figure 8. Mean absolute error (MAE) between the native angles and the angle from the RNAPOT models for RNA with PDB ID: 1i6u, Chain C as a function of RMSD.

chi (χ), eta (η), and theta (θ). The average MAEs for all angles except pseudotorsion angles are much less than the above thresholds, confirming the usefulness of using the predicted angles for assessing model quality.

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

DISCUSSION

In this work, we have developed SPOT-RNA-1D to make backbone angle prediction for RNAs. The single-sequence-based method provides a substantial improvement of over 20% for most angles compared to random predictions. The accuracy of predicted angles strongly depends on whether bases are paired or unpaired and paired bases associated with tertiary interactions or not. The backbone angles of unpaired bases are the most difficult to predict, followed by those bases associated with tertiary interactions. This is likely because the most stable canonical base pairs provide the strongest restriction to backbone angles (thus easier to predict).

This dilated convolutional neural network-based architecture can predict RNA torsion angles with reasonable accuracy because of its ability to use a wide receptive field for the predictions. The receptive field is the total number of the surrounding nucleotides' features used by the neural network to predict a given nucleotide's angles. The receptive field for the SPOT-RNA-1D architecture (in Figure 2) was 127, which means the model considered features of 63 nucleotides from either side of a given nucleotide for the prediction. This wide receptive field allows the SPOT-RNA-1D network to efficiently learn both sequence-neighboring and nonlocal interactions. Furthermore, to avoid the model overfitting on the small training data, we did not use the architecture with a very deep neural network. The SPOT-RNA-1D network architecture consists of only three convolutional layers because we found that additional layers would lead to model overfitting. We also trained SPOT-RNA-1D for 25 epochs only to avoid overfitting. Careful training is what makes a SPOT-RNA-1D a generalizable model.

To improve the performance of SPOT-RNA-1D, we also examined the use of predicted secondary structure base pair probability from the single-sequence-based method LinearPartition[40] as an input feature. The use of predicted base pair probability from LinearPartition[40] improved the performance of RNA solvent accessibility prediction in our previous work RNAsnap2.[16] The predicted base pair probability from LinearPartition was a 2D feature of size $L\times L$, where L is the length of the sequence. We converted this 2D base pair probability to 1D features of size $L\times 1$ by taking summation across one dimension. These 1D features ($L\times 1$) were concatenated with one-hot encoding ($L\times 4$) to yield an input feature vector of size $L\times 5$ for SPOT-RNA-1D. Supplementary Table S5 shows the performance comparison between one-hot encoding and one-hot encoding with the LinearPartition base pair probability as an additional input feature. While the addition of base pair probability decrease MAEs significantly for all nine torsion angles (Supplementary Table S5), the distributions of predicted angles (Supplementary Figure S16) are very narrow, dominated by the most populated angles. This indicates that adding predicted secondary structures will overfit the method. Thus, we have limited SPOT-RNA-1D input features to one-hot encoding only in the final version of the tool.

We also attempted to improve delta (δ) angle prediction because it has the smallest improvement (~2°) over the random-baseline prediction. As we noticed that the delta (δ) could be approximated between 0° and 180° according to its distribution (Figure 5), we applied the restriction to delta-angle training and testing. We found that this restriction only leads to a minor improvement in delta prediction (~2%). However, it also leads to a slight reduction in the performance for other angles. Thus, we decided not to restrict the range in predicting the delta (δ) angle.

The SPOT-RNA-1D predictor has a few limitations. First, the mean absolute error (MAE) of SPOT-RNA-1D is relatively high (>30) for the angles alpha (α), eta (η), and theta (θ) for the majority of the test sets (see Table 1). This is because these three angles are more evenly distributed from −180° to +180° compared to other angles (see Figure 5). Thus, there is a need to explore some other neural network architectures, such as the Attention[41] mechanism that has shown the potential to learn from a wide-range distribution space in proteins' structural properties.[11,42,43] Also, SPOT-RNA-1D does not perform well for RNAs with more tertiary interactions and unpaired nucleotides. This problem is intrinsically related to the challenge for predicting 3D structures from sequences and the flexible nature of the loop regions. Bigger data sets and better-designed network architectures may be required for further improving the performance.

A substantial improvement over random prediction by SPOT-RNA-1D and its comparison to existing models predicted for RNA-puzzles indicate that RNA backbone torsion angles are predictable, despite the fact that there are so many more RNA backbone angles than two angles [phi (ϕ) and psi (ψ)] in protein backbones although the MAEs for phi and psi by SPOT-1D[10] are at 17° and 25°, respectively. The MAEs for RNA backbone angles are not far behind except alpha (~40°). However, SPOT-1D has relied on evolutionary information from multiple sequence alignment. If only single-sequence is employed for angle prediction, the MAEs for ϕ and ψ are 24° and 44°, respectively.[44] Thus, the overall accuracy is comparable if only a single sequence is used as input. This result highlights the importance of using evolution information for improving angle prediction as we have demonstrated their importance in solvent accessibility prediction in RNAsnaps[16,18] and in secondary structure prediction in SPOT-RNA2.[4,45] The work for using evolution information for RNA backbone angle prediction is in progress.

It should be noted that directly using predicted angles to build a 3D model will not yield an accurate structure because small deviations in angles would lead to systematic errors in 3D. The purpose of predicting torsion angles in RNA is the same as in proteins: to make predicted angles accurate enough as a starting point for further future improvement and as useful restraints for *ab initio* structure prediction methods along with predicted base pairs. This is feasible based on the result that predicted angles can serve as useful indicators of model quality.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00153.

Table S1, number of RNAs, base pairs, and length distribution of data sets; Table S2, statistical significance of SPOT-RNA-1D improvement over random-baseline predictor; Table S3, performance comparison SPOT-RNA-1D and random-baseline predictor based on helix angle distribution; Table S4, performance comparison of SPOT-RNA-1D with RNA-Puzzles predictors; Table S5, performance comparison of method with single-sequence input and with single-sequence plus predicted secondary structure; Figure S1, distribution plots of

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622

native torsion angles of training data; Figure S2, distribution plots of native and predicted torsion angles; Figure S3, distribution plots of native and predicted pseudotorsion angles; Figure S4, MAE vs RMSD scatterplot of nine RNAs from RNAPOT; Figure S5, MAE vs GDT-score scatterplot of nine RNAs from RNAPOT; Figure S6, MAE between native angles and angle from RNAPOT models for 1kxk-A as function of RMSD; Figure S7, MAE between native angles and angle from RNAPOT models for 1mzp-B as function of RMSD; Figure S8, MAE between native angles and angle from RNAPOT models for 1s03-A as function of RMSD; Figure S9, MAE between native angles and angle from RNAPOT models for 1u63-Chain B as function of RMSD; Figure S10, MAE between native angles and angle from RNAPOT models for 1un6-E as function of RMSD; Figure S11, MAE between native angles and angle from RNAPOT models for 2dr2-B as function of RMSD; Figure S12, MAE between native angles and angle from RNAPOT models for 2oiu-P as function of RMSD; Figure S13, MAE between native angles and angle from RNAPOT models for 2qwy-A as function of RMSD; Figure S14, MAE between native angles and angle from RNAPOT models for 2zni-C as function of RMSD; Figure S15, 3D structural alignment of native structure and decoy-17 of 2dr2-B; and Figure S16, distribution plots of native and predicted torsion angles by method with single sequence and predicted secondary structure as input (PDF)

AUTHOR INFORMATION

Corresponding Authors

Jaswinder Singh − *Signal Processing Laboratory, Griffith University, Brisbane, Queensland 4122, Australia;* orcid.org/0000-0002-0478-5533; Email: jaswinder.singh3@griffithuni.edu.au

Yaoqi Zhou − *Institute for Glycomics and School of Information and Communication Technology, Griffith University, Southport, Queensland 4222, Australia; Institute for Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518055, China; Peking University Shenzhen Graduate School, Shenzhen 518055, P.R. China;* orcid.org/0000-0002-9958-5699; Email: zhouyq@szbl.ac.cn

Authors

Kuldip Paliwal − *Signal Processing Laboratory, Griffith University, Brisbane, Queensland 4122, Australia*

Jaspreet Singh − *Signal Processing Laboratory, Griffith University, Brisbane, Queensland 4122, Australia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c00153

Notes

The authors declare no competing financial interest.

Standalone versions of SPOT-RNA-1D along with data sets used for training, validation, and testing are available at https://github.com/jaswindersingh2/SPOT-RNA-1D. Direct prediction can also be made at https://sparks-lab.org/server/spot-rna-1d.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council (DP210101875 to Y.Z. and K.P.). We gratefully acknowledge the use of the High Performance Computing Cluster Gowonda to complete this research and the aid of the research cloud resources provided by the Queensland Cyber Infrastructure Foundation (QCIF). We also gratefully acknowledge the support of the NVIDIA Corporation with the donation of the Titan V GPU used for this research. The support of the Shenzhen Science and Technology Program (grant no. KQTD20170330155106581) and the Major Program of Shenzhen Bay Laboratory S201101001 is also acknowledged.

REFERENCES

(1) Puton, T.; Kozlowski, L. P.; Rother, K. M.; Bujnicki, J. M. CompaRNA: A Server for Continuous Benchmarking of Automated Methods for RNA Secondary Structure Prediction. *Nucleic Acids Res.* **2013**, *41*, 4307−4323.

(2) Hamada, M. *RNA Bioinformatics*; Springer New York: New York, NY, 2015; pp 17−38, DOI: 10.1007/978-1-4939-2291-8_2.

(3) Singh, J.; Hanson, J.; Paliwal, K.; Zhou, Y. RNA Secondary Structure Prediction using an Ensemble of Two-dimensional Deep Neural Networks and Transfer Learning. *Nat. Commun.* **2019**, *10*, 5407.

(4) Singh, J.; Paliwal, K.; Zhang, T.; Singh, J.; Litfin, T.; Zhou, Y. Improved RNA Secondary Structure and Tertiary Base-pairing Prediction using Evolutionary Profile, Mutational Coupling and Two-dimensional Transfer Learning. *Bioinformatics* **2021**, DOI: 10.1093/bioinformatics/btab165.

(5) Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)-Round XIII. *Proteins: Struct., Funct., Genet.* **2019**, *87*, 1011−1020.

(6) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Zídek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction using Potentials from Deep Learning. *Nature* **2020**, *577*, 706−710.

(7) Cai, Y.; Li, X.; Sun, Z.; Lu, Y.; Zhao, H.; Hanson, J.; Paliwal, K.; Litfin, T.; Zhou, Y.; Yang, Y. SPOT-Fold: Fragment-Free Protein Structure Prediction Guided by Predicted Backbone Structure and Contact Map. *J. Comput. Chem.* **2020**, *41*, 745−750.

(8) 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.

(9) 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.

(10) Hanson, J.; Paliwal, K.; Litfin, T.; Yang, Y.; Zhou, Y. Improving Prediction of Protein Secondary Structure, Backbone Angles, Solvent Accessibility and Contact Numbers by using Predicted Contact Maps and an Ensemble of Recurrent and Residual Convolutional Neural Networks. *Bioinformatics* **2019**, *35*, 2403−2410.

(11) Xu, G.; Wang, Q.; Ma, J. OPUS-TASS: A Protein Backbone Torsion Angles and Secondary Structure Predictor Based on Ensemble Neural Networks. *Bioinformatics* **2020**, *36*, 5021.

(12) Singh, J.; Hanson, J.; Heffernan, R.; Paliwal, K.; Yang, Y.; Zhou, Y. Detecting Proline and Non-Proline Cis Isomers in Protein Structures from Sequences using Deep Residual Ensemble Learning. *J. Chem. Inf. Model.* **2018**, *58*, 2033−2042.

(13) Singh, J. *Detection of Cis-Trans Conformation in Protein Structure using Deep Learning Neural Network Techniques*. M.Sc. thesis, Griffith University, Brisbane, Australia, 2019; DOI: 10.25904/1912/560.

(14) Wadley, L. M.; Keating, K. S.; Duarte, C.; Pyle, A. Evaluating and Learning from RNA Pseudotorsional Space: Quantitative

Validation of a Reduced Representation for RNA Structure. *J. Mol. Biol.* **2007**, *372*, 942−957.
(15) Frellsen, J.; Moltke, I.; Thiim, M.; Mardia, K. V.; Ferkinghoff-Borg, J.; Hamelryck, T. A Probabilistic Model of RNA Conformational Space. *PLoS Comput. Biol.* **2009**, *5*, e1000406.
(16) Hanumanthappa, A. K.; Singh, J.; Paliwal, K.; Singh, J.; Zhou, Y. Single-sequence and Profile-based Prediction of RNA Solvent Accessibility using Dilated Convolutional Neural Network. *Bioinformatics* **2021**, *36*, 5169−5176.
(17) Sun, S.; Wu, Q.; Peng, Z.; Yang, J. Enhanced Prediction of RNA Solvent Accessibility with Long Short-term Memory Neural Networks and Improved Sequence Profiles. *Bioinformatics* **2019**, *35*, 1686−1691.
(18) Yang, Y.; Li, X.; Zhao, H.; Zhan, J.; Wang, J.; Zhou, Y. Genome-scale Characterization of RNA Tertiary Structures and Their Functional Impact by RNA Solvent Accessibility Prediction. *RNA* **2017**, *23*, 14−22.
(19) Rose, P. W.; Prlić, A.; Altunkaya, A.; Bi, C.; Bradley, A. R.; Christie, C. H.; Costanzo, L.i Di; Duarte, J. M.; Dutta, S.; Feng, Z.; Green, R. K.; Goodsell, D.S.; Hudson, B.; Kalro, T.; Lowe, R.; Peisach, E.; Randle, C.; Rose, A. S.; Shao, C.; Tao, Y.-P.; Valasatava, Y.; Voigt, M.; Westbrook, J. D.; Woo, J.; Yang, H.; Young, J. Y.; Zardecki, C.; Berman, H. M.; Burley, S. K. The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information. *Nucleic Acids Res.* **2016**, *45*, D271−D281.
(20) Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422−1423.
(21) Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the Next-generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150−3152.
(22) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403−410.
(23) Coordinators, N. R. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44*, D7−D19.
(24) Nawrocki, E. P.; Eddy, S. R. Infernal 1.1:100-fold Faster RNA Homology Searches. *Bioinformatics* **2013**, *29*, 2933−2935.
(25) Lu, X.-J.; Bussemaker, H. J.; Olson, W. K. DSSR: An Integrated Software Tool for Dissecting the Spatial Structure of RNA. *Nucleic Acids Res.* **2015**, *43*, gkv716.
(26) Cruz, J. A.; Blanchet, M.-F.; Boniecki, M.; Bujnicki, J. M.; Chen, S.-J.; Cao, S.; Das, R.; Ding, F.; Dokholyan, N. V.; Flores, S. C.; Huang, L.; Lavender, C. A.; Lisi, V.; Major, F.; Mikolajczak, K.; Patel, D. J.; Philips, A.; Puton, T.; Santalucia, J.; Sijenyi, F.; Hermann, T.; Rother, K.; Rother, M.; Serganov, A.; Skorupski, M.; Soltysinski, T.; Sripakdeevong, P.; Tuszynska, I.; Weeks, K. M.; Waldsich, C.; Wildauer, M.; Leontis, N. B.; Westhof, E. RNA-Puzzles: A CASP-like Evaluation of RNA Three-dimensional Structure Prediction. *RNA* **2012**, *18*, 610−625.
(27) Miao, Z.; Adamiak, R. W.; Blanchet, M.-F.; Boniecki, M.; Bujnicki, J. M.; Chen, S.-J.; Cheng, C.; Chojnowski, G.; Chou, F.-C.; Cordero, P.; Cruz, J. A.; Ferré-D'Amaré, A. R.; Das, R.; Ding, F.; Dokholyan, N. V.; Dunin-Horkawicz, S.; Kladwang, W.; Krokhotin, A.; Lach, G.; Magnus, M.; Major, F.; Mann, T. H.; Masquida, B.; Matelska, D.; Meyer, M.; Peselis, A.; Popenda, M.; Purzycka, K. J.; Serganov, A.; Stasiewicz, J.; Szachniuk, M.; Tandon, A.; Tian, S.; Wang, J.; Xiao, Y.; Xu, X.; Zhang, J.; Zhao, P.; Zok, T.; Westhof, E. RNA-Puzzles Round II: Assessment of RNA Structure Prediction Programs Applied to Three Large RNA Structures. *RNA* **2015**, *21*, 1066−1084.
(28) Miao, Z.; Adamiak, R. W.; Antczak, M.; Batey, R. T.; Becka, A. J.; Biesiada, M.; Boniecki, M. J.; Bujnicki, J. M.; Chen, S.-J.; Cheng, C. Y.; Chou, F.-C.; Ferré-D'Amaré, A. R.; Das, R.; Dawson, W. K.; Ding, F.; Dokholyan, N. V.; Dunin-Horkawicz, S.; Geniesse, C.; Kappel, K.; Kladwang, W.; Krokhotin, A.; Łach, G. E.; Major, F.; Mann, T. H.; Magnus, M.; Pachulska-Wieczorek, K.; Patel, D. J.; Piccirilli, J. A.; Popenda, M.; Purzycka, K. J.; Ren, A.; Rice, G. M.; Santalucia, J.; Sarzynska, J.; Szachniuk, M.; Tandon, A.; Trausch, J. J.; Tian, S.; Wang, J.; Weeks, K. M.; Williams, B.; Xiao, Y.; Xu, X.; Zhang, D.; Zok, T.; Westhof, E. RNA-Puzzles Round III: 3D RNA Structure Prediction of Five Riboswitches and One Ribozyme. *RNA* **2017**, *23*, 655−672.
(29) Miao, Z.; Adamiak, R. W.; Antczak, M.; Boniecki, M. J.; Bujnicki, J.; Chen, S.-J.; Cheng, C. Y.; Cheng, Y.; Chou, F.-C.; Das, R.; Dokholyan, N. V.; Ding, F.; Geniesse, C.; Jiang, Y.; Joshi, A.; Krokhotin, A.; Magnus, M.; Mailhot, O.; Major, F.; Mann, T. H.; Piątkowski, P.; Pluta, R.; Popenda, M.; Sarzynska, J.; Sun, L.; Szachniuk, M.; Tian, S.; Wang, J.; Wang, J.; Watkins, A. M.; Wiedemann, J.; Xiao, Y.; Xu, X.; Yesselman, J. D.; Zhang, D.; Zhang, Y.; Zhang, Z.; Zhao, C.; Zhao, P.; Zhou, Y.; Zok, T.; Żyła, A.; Ren, A.; Batey, R. T.; Golden, B. L.; Huang, L.; Lilley, D. M.; Liu, Y.; Patel, D. J.; Westhof, E. RNA-Puzzles Round IV: 3D Structure Predictions of Four Ribozymes and Two Aptamers. *RNA* **2020**, *26*, 982−995.
(30) Capriotti, E.; Norambuena, T.; Marti-Renom, M. A.; Melo, F. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics* **2011**, *27*, 1086−1093.
(31) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(32) He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. *Computer Vision âEUR" ECCV 2016*; Cham, 2016; pp 630−645, DOI: 10.1007/978-3-319-46493-0_38.
(33) Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. 2015, arXiv:1511.07122. *arXiv Preprint*. https://arxiv.org/abs/1511.07122 (accessed 2021-05-20).
(34) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9*, 1735−1780.
(35) Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). 2015, arXiv:1511.07289. *arXiv Preprint*. https://arxiv.org/abs/1511.07289 (accessed 2021-05-20).
(36) Nam, H.; Kim, H.-E. Batch-Instance Normalization for Adaptively Style-Invariant Neural Networks. 2018, arXiv:1805.07925. *arXiv Preprint*. https://arxiv.org/abs/1805.07925 (accessed 2021-05-20).
(37) Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; Kudlur, M.; Levenberg, J.; Monga, R.; Moore, S.; Murray, D. G.; Steiner, B.; Tucker, P.; Vasudevan, V.; Warden, P.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: A System for Large-Scale Machine Learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*; Savannah, GA, 2016; pp 265−283.
(38) Tieleman, T.; Hinton, G. Lecture 6.5—RMSProp: Divide the Gradient by a Running Average of its Recent Magnitude. *Neural Networks for Machine Learning*; COURSERA: 2012.
(39) Singh, J.; Litfin, T.; Paliwal, K.; Singh, J.; Hanumanthappa, A. K.; Zhou, Y. SPOT-1D-Single: Improving the Single-Sequence-Based Prediction of Protein Secondary Structure, Backbone Angles, Solvent Accessibility and Half-Sphere Exposures using a Large Training Set and Ensembled Deep Learning. *Bioinformatics* **2021**, btab316.
(40) Zhang, H.; Zhang, L.; Mathews, D. H.; Huang, L. LinearPartition: Linear-time Approximation of RNA Folding Partition Function and Base-pairing Probabilities. *Bioinformatics* **2020**, *36*, i258−i267.
(41) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*; 2017.
(42) Uddin, M. R.; Mahbub, S.; Rahman, M. S.; Bayzid, M. S. SAINT: Self-attention Augmented Inception-inside-inception Network Improves Protein Secondary Structure Prediction. *Bioinformatics* **2020**, *36*, 4599−4608.
(43) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 million

Protein Sequences. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, e2016239118.
(44) Heffernan, R.; Paliwal, K.; Lyons, J.; Singh, J.; Yang, Y.; Zhou, Y. Single-sequence-based Prediction of Protein Secondary Structures and Solvent Accessibility by Deep Whole-sequence Learning. *J. Comput. Chem.* **2018**, *39*, 2210−2216.
(45) Zhang, T.; Singh, J.; Litfin, T.; Zhan, J.; Paliwal, K.; Zhou, Y. RNAcmap: A Fully Automatic Pipeline for Predicting Contact Maps of RNAs by Evolutionary Coupling Analysis. *Bioinformatics* **2021**.

https://doi.org/10.1021/acs.jcim.1c00153
J. Chem. Inf. Model. 2021, 61, 2610−2622