OXFORD

Structural bioinformatics

RBind: computational network method to predict RNA binding sites

Kaili Wang[1,†], Yiren Jian[2,†], Huiwen Wang[1], Chen Zeng[1,2] and Yunjie Zhao[1,*]

[1]Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, China and [2]Department of Physics, The George Washington University, Washington, DC 20052, USA

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: Alfonso Valencia

Abstract

Motivation: Non-coding RNA molecules play essential roles by interacting with other molecules to perform various biological functions. However, it is difficult to determine RNA structures due to their flexibility. At present, the number of experimentally solved RNA–ligand and RNA–protein structures is still insufficient. Therefore, binding sites prediction of non-coding RNA is required to understand their functions.

Results: Current RNA binding site prediction algorithms produce many false positive nucleotides that are distance away from the binding sites. Here, we present a network approach, RBind, to predict the RNA binding sites. We benchmarked RBind in RNA–ligand and RNA–protein datasets. The average accuracy of 0.82 in RNA–ligand and 0.63 in RNA–protein testing showed that this network strategy has a reliable accuracy for binding sites prediction.

Availability and implementation: The codes and datasets are available at https://zhaolab.com.cn/RBind.

Contact: yjzhaowh@mail.ccnu.edu.cn

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

1 Introduction

Non-coding RNA molecules are involved in various biological processes. For example, the discovery of 1445 RNA viruses enriched our knowledge of RNA virus biodiversity (Shi *et al.*, 2016). Previous research also showed that RNAs had been implicated in tumorigenesis, neurological, cardiovascular, developmental and many other diseases (Esteller, 2011). RNA functions always depend on their structures. The non-coding aptamer portion of a riboswitch, for example, can form a specific pocket and directly bind small molecules. This binding induces downstream RNA structural change (allosteric effect) that regulates the functions of its coding portion (Garst *et al.*, 2011; Gong *et al.*, 2011, 2012, 2014). Therefore, both local and long-range structural features are important toward function. In addition, the rapidly expanding sequence (NONCODE, www.noncode.org) (Zhao *et al.*, 2016) and structural data (RCSB Protein

Data Bank, www.rcsb.org) (Berman *et al.*, 2000) have shown that RNAs have many more diverse functions than expected. The structural information of RNAs is needed for their function characterization and therapeutic application development.

Most RNA molecules need to interact with other molecules to perform biological functions. However, it remains challenging to determine the high-resolution RNA structures due to their flexibility, especially for RNA–ligand and RNA–protein structures (Magnus *et al.*, 2014). Some computational methods have been proposed to predict the RNA structure, which can serve as distance constraints to guide the complex structural modeling for much better accuracy (Krokhotin *et al.*, 2015; Miao *et al.*, 2017; Popenda *et al.*, 2012; Sun *et al.*, 2017; Wang *et al.*, 2015; Yesselman and Das, 2016; Zhao *et al.*, 2011, 2012, 2013). For RNA–protein complex, many algorithms, like RNABindR (Terribilini *et al.*, 2007), BindN (Wang and Brown, 2006), PiRaNhA (Murakami *et al.*, 2010),

BindUP (Paz *et al.*, 2016) and SPalign (Yang *et al.*, 2012), are developed to identify the binding sites on protein structure to interact with RNAs. The algorithms of RNA binding sites prediction from RNA side are still limited. Rsite and Rsite2 are distance based computational methods to identify the functional sites of noncoding RNAs (Zeng and Cui, 2016; Zeng *et al.*, 2015). It calculates the Euclidean distance from tertiary structural (Rsite) or hamming distance from secondary structural (Rsite2) between each nucleotide and all other nucleotides and then associates the nucleotides at the extreme points on the distance curve as the functional sites. However, solely distance calculations without distinguishing between highly and reliable connections may cause false positive predictions. In their case studies, some of the extreme points lead to false positive predictions. Moreover, the evaluation of the performance in large-scale studies remains excessively demanding in computational terms. There is one additional algorithm for RNA binding site prediction via sequence deep learning (Alipanahi *et al.*, 2015). However, DeepBind is specifically designed for identifying the critical transcription factor binding sites. Therefore, it is still a challenge for accurate prediction of the RNA binding sites and new algorithms are highly desired.

In this paper, we provide a structural network approach to predict the binding sites of RNA molecules. More specifically, the RNA tertiary structures were transformed into a network, where nucleotides are nodes, and their non-covalent interactions with each other are the edges. Then, we calculated the degree values for short-range binding cavity and closeness values for long-range allosteric effect to identify the binding sites. Furthermore, we applied our method on the RNA–ligand dataset (Philips *et al.*, 2013) and the RNA–protein dataset (Huang and Zou, 2013) for large-scale testing. The results showed that RBind has a reliable accuracy for RNA binding sites prediction.

2 Materials and methods

2.1 Network construction

We predict the binding sites by using the following procedures (Fig. 1). First, the RNA structures are transformed into a network. The main components of network are nodes and edges. In the biomolecular structure network, a node usually denotes a single nucleotide/residue. RBind denotes a single nucleotide as a node and the non-covalent interactions as edges. Previous research reported a cutoff range of 6.5–8 Å (residue-residue distance) to be used in the edge construction of protein structural network (Greene and Higman, 2003). Unlike protein, we define the cutoff distance in the RNA network based on the following two considerations: (i) Previous work indicated that 8 Å can serve as a reliable contact cutoff for RNA tertiary structural study (De Leonardis *et al.*, 2015; Weinreb *et al.*, 2016) and (ii) A-form RNA rises 2.6 Å per base pair, the stacking interaction is thus small if the distance of two nucleotides is larger than 8 Å. Therefore, our cutoff threshold is chosen as follows. Two nonconsecutive nucleotides in a sequence are connected by an edge if they contain a pair of heavy atoms, one from each nucleotide, less than 8 Å apart. For example, Supplementary Figure S1 shows the constructed network from an RNA tertiary structure. We used the hamming distance in the network and removed the covalent connections. Therefore, the shortest distances between nucleotide 1 and nucleotide 2 is 2. Cytoscape was used for the network visualization in Supplementary Figure S1 (Shannon, 2003).

Second, the closeness and degree values of each node in RNA network are calculated to identify the binding sites. In the construct
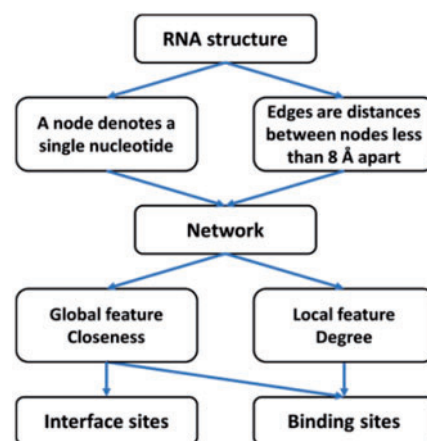

Fig. 1. Workflow chart of RBind to predict interface and binding sites of RNA molecules. RBind first transforms RNA tertiary structure into a connected network, then uses both global (closeness) and local (degree) features to identify binding sites

of the network, the closeness of a node is defined as the inverse of the sum of its shortest distances to all other $(n-1)$ nodes as the following:

$$C(x) = \frac{n-1}{\sum d(x,y)} \tag{1}$$

where $d(x, y)$ is the distance of the shortest path between the node x and any other node y (Amitai *et al.*, 2004). The shortest paths between all pairs of nodes are found using the Floyd-Warshall algorithm. High closeness nodes act as network hubs. The degree of a node is defined as the number of edges attached to the node, a measure that describes the local pocket connections.

Third, we identified those nodes as ligand binding sites when their closeness and degree values are both higher than the corresponding average values by a standard deviation, i.e. higher than a cutoff chosen as cutoff = average + standard deviation. Our previous study suggested that this closeness cutoff is able to identify critical residues for complex formation (Zhao *et al.*, 2017). Therefore, we performed closeness analysis for RNA–protein interface binding sites prediction.

2.2 The testing dataset

2.2.1 RNA–ligand testing set

In previous structure-based RNA–ligand docking study, Philips et al. used an RNA–ligand complex structure benchmark for testing (Philips *et al.*, 2013). Here, we excluded the stranded RNA helix structure due to simple helix topology. In our test, we included 22 RNA–ligand complex structures with RNA length of 20–94 nucleotides. The experimental binding sites are defined as the nucleotides within 4 Å distance from the ligand.

2.2.2 RNA–protein testing set

To validate the accuracy of this network prediction strategy for RNA–protein binding, we applied our method to a non-redundant RNA–protein structure dataset (Huang and Zou, 2013). The detailed description of this RNA–protein structure dataset can be found in refs (Huang and Zou, 2013). Here we just briefly describe the information used in the present paper. The testing dataset consists of 72 diverse RNA–protein structures with RNA length of 20–157 nucleotides. Since the prediction is needed when complexes are unknown, we used the unbound RNA structures for binding site

prediction. Dataset can be downloaded from http://zoulab.dalton.
missouri.edu/RNAbenchmark/.

2.3 Tertiary structure prediction

We predicted RNA tertiary structures using RNAComposer, an automated RNA structure model program to build RNA three-dimensional structures based on sequence and secondary structure information (Biesiada *et al.*, 2016; Popenda *et al.*, 2012) by RNA FRABASE. For each RNA structure prediction, we use the corresponding sequence and secondary structure on the RNA structure modeling server (http://rnacomposer.cs.put.poznan.pl). All tertiary structures are predicted automatically.

2.4 Comparison with existing method

We compared RBind with existing computational RNA binding sites prediction methods Rsite and Rsite2 (Zeng and Cui, 2016; Zeng *et al.*, 2015). It calculates the Euclidean distance from tertiary structural (Rsite) or distance from secondary structural (Rsite2) between each nucleotide and all other nucleotides and then associates the nucleotides at the extreme points on the distance curve as the functional sites. We used the downloaded Rsite program (http://www.cuilab.cn/rsite) and Rsite2 online server (http://www.cuilab.cn/rsite2/start) to analyze the RNA–ligand and RNA–protein data.

2.5 Nucleotide–nucleotide coevolution analysis

Direct Coupling Analysis (DCA) was performed to infer the nucleotide–nucleotide coevolution among base pair interactions, binding site interactions and other interactions, respectively (de Juan *et al.*, 2013; De Leonardis *et al.*, 2015; Marks *et al.*, 2012; Morcos *et al.*, 2011; Weinreb *et al.*, 2016; Xing *et al.*, 2016). The details of DCA are described in references (Zhao *et al.*, 2017) and (Morcos *et al.*, 2011). The homology sequences were extracted from Rfam (Nawrocki *et al.*, 2015).

2.6 Prediction accuracy analysis

We calculated the positive predictive value (PPV) and sensitivity (STY) as follows:

$$PPV = \frac{|TP|}{|TP| + |FP|} \qquad (2)$$

$$STY = \frac{|TP|}{|TP| + |FN|} \qquad (3)$$

TP (FP) stands for true (false) positives, i.e. the number of true (false) binding sites among the predicted binding sites. FN stands for false negatives, i.e. the number of true binding sites the method fails to predict. PPV thus measures the fraction of true binding sites in the prediction and STY the fraction of true binding sites that can be predicted. Together, PPV and STY provide a full picture on the goodness of binding site prediction.

3 Results

We benchmarked our method in two datasets. The first one is an RNA–ligand dataset containing 22 RNAs with small molecules to test the ability of binding site prediction for potential drug targets. The second one is a non-redundant structure RNA–protein dataset including 72 RNAs with the length of 20–157 nucleotides to test the prediction accuracy of RNA binding sites for interface interactions.
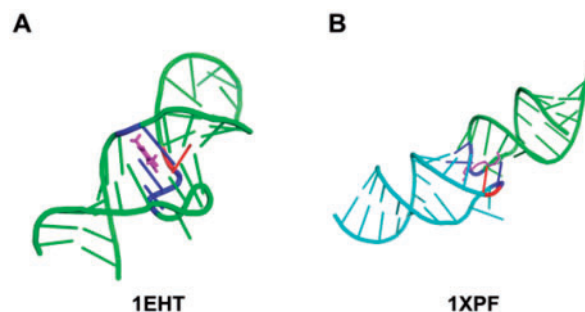

Fig. 2. Cartoon representation of (**A**) RNA structural motif (PDB code: 1EHT) and (**B**) RNA complex (PDB code: 1XPF) with small molecules (colored in magenta). The different chains are colored in green and cyan. The predicted binding sites are colored in blue (experimental binding nucleotides) and red (false positive binding nucleotides) (Color version of this figure is available at *Bioinformatics* online.)

3.1 The small molecule binding sites prediction of RNA–ligand dataset

RNAs have recently been considered as potential drug targets due to the ability to bind small molecules for essential biological processes (Philips *et al.*, 2013). The increasing number of experimentally determined RNA–ligand complex structures allows us to design inhibitors that block specific binding sites for regulating RNA functions. To test the ability to identify potential drug binding pockets, we applied our network approach to the RNA–ligand dataset for small molecule binding sites prediction.

We can identify the binding sites in RNA structural motifs. Many RNA structural motifs form a binding pocket with hydrogen bonds and stacking interactions. RNA 1EHT forms a binding pocket with eight nucleotides (U6, A7, C8, C22, U23, U24, G26, A28, Fig. 2A). Our method successfully identified three such binding sites (A7, C22 and U23) but with one false positive site (C21). The accuracy of PPV is 0.75 for the binding site prediction of this structural motif. We also applied our method to RNAs with two chains. The RNA Kissing Complex (PDB code: 1XPF) is involved in human immunodeficiency virus type 1 (HIV-1). The study on binding sites may help design new drugs disrupting the RNA dimerization step in the viral life cycle. The crystal complex structure of 1XPF shows that three nucleotides from chain A (G9, G10 and U11) and one nucleotide from chain B (G10) form a pocket (Fig. 2B). We applied our network approach to predict the pseudoknot-containing aptamer binding sites and successfully identified two nucleotides in chain A (G10 and U11) and one nucleotide in chain B (G10) but with one false positive nucleotide U11 in chain B. The accuracy of PPV is 0.75 for this binding site prediction. It is noted that the false positive prediction is located next to the binding site. Figure 3 shows the prediction results for the entire RNA–ligand dataset with an average accuracy of 0.82. Taken together, these results indicate that our method can successfully identify the RNA binding sites for small molecule binding.

3.2 The RNA–protein dataset testing

RNA–protein interactions are critical for many biological processes. To predict the RNA binding sites for RNA–protein complexes formation would be valuable for RNA–protein complex structural modeling and understanding their biological mechanisms. Therefore, we applied our network prediction method to a non-redundant RNA–protein structure dataset for binding sites prediction testing (Huang and Zou, 2013).
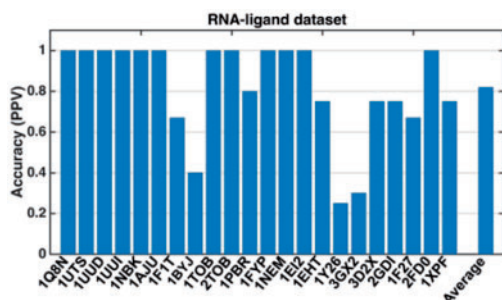
Fig. 3. The RNA–ligand benchmark testing results. RBind is able to successfully identify the binding sites with an average PPV around 0.82

Fig. 5. The RNA–protein benchmark testing results. RBind is able to successfully identify the binding sites with an average PPV around 0.63

Fig. 4. Cartoon representation of tRNA (2ZUE). The predicted interface binding sites are colored in blue (interface nucleotides) and red (not interface nucleotides). Ten out of eleven predicted binding sites are the correct interface binding sites reported by RNA–protein dataset (http://zoulab.dalton.mis souri.edu/RNAbenchmark/)

tRNAs are the essential translation component for producing new proteins and need to interact with other proteins to perform this biological function. For a tRNA (2ZUE) in our RNA–protein dataset, RBind successfully identified 14 interface binding sites (G906, G907, U908, A909, C912, U913, A914, G915, C916, A921, G923, G924, C948, C949) but with one false positive site (G946, Fig. 4). The accuracy of PPV is 0.93 for the nucleotide prediction of this structural interface. Overall, the accuracy value of predicted nucleotides is 0.63 for the entire RNA–protein dataset (Fig. 5). The results suggest that network closeness calculation can successfully identify the interface sites when conformational changes are small between unbound and bound structures. The prediction accuracy may decrease if an RNA–protein complex formation induces considerable change in the unbound RNA conformation.

3.3 Binding sites prediction without experimental tertiary structures

RBind is a tertiary structure based binding site prediction method. Its application is largely restricted to the available RNA tertiary structures, which are very difficult to obtain experimentally. However, the accuracy for modeling RNA 3D structure computationally has been dramatically improved over the past five years. Recent RNA tertiary structure prediction methods can construct accurate atom-level structures for RNAs less than 100 nucleotides. This makes it feasible to use RBind to predict RNA binding sites without their experimentally determined tertiary structures.

To build the RNA tertiary structures, which we call model structures as opposed to the experimentally determined structures, we removed RNAs with more than one chain or pseudoknot interactions. This resulted in 19 RNAs in the RNA–ligand dataset and 47 RNAs in the RNA–protein dataset. Figure 6 shows the accuracy
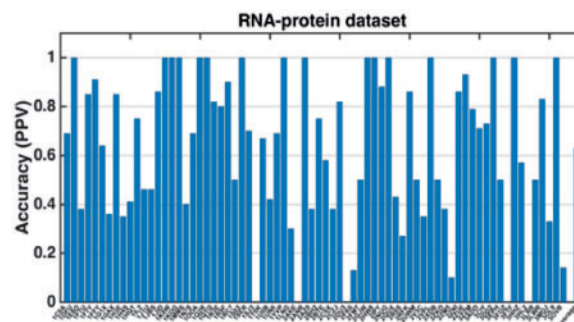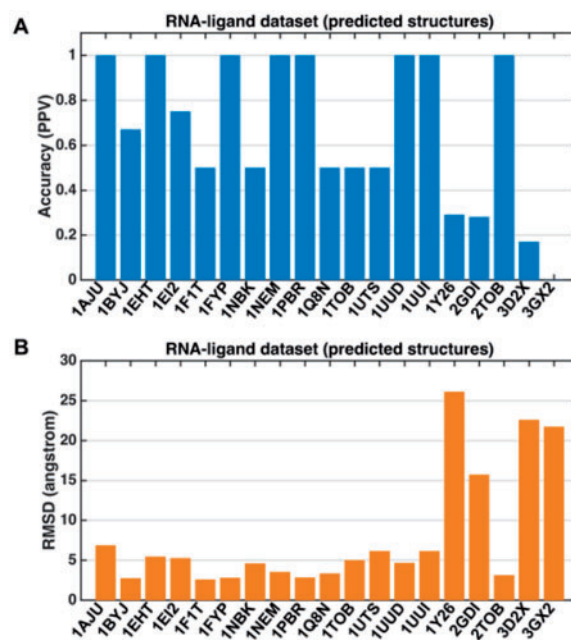

Fig. 6. Relation between the accuracy (**A**) of binding sites prediction from computationally modeled structures and the accuracy (**B**) of these model structures compared to their native structures. As shown in (A), binding site predictions by RBind achieved a reasonable accuracy for most of the model RNA structures from the RNA–ligand dataset. The four cases of low accuracy for binding site predictions with PPV values less than 0.3 also have low accuracy for structure modeling (or high RMSD compared to the native states) as shown in (B) for 1Y26, 2GDI, 3D2X and 3GX2. The correlation coefficient between the accuracy of binding site predictions and RMSDs is −0.73

(PPV) for binding site prediction from the model structures as well as the accuracy (RMSD) of these model structures compared to their native structures. Using only model structures, RBind is able to achieve a reasonable accuracy (PPV > 0.4) for all 19 RNAs except four cases (PDB code: 1Y26, 2GDI, 3D2X and 3GX2). It is interesting to note that only these four model structures deviate significantly from their native structures with RMSD values larger than 15 Å. It is thus not surprising that the accuracy for the binding site prediction also dropped significantly with PPV < 0.3. Overall, the average prediction accuracy (PPV value) is 0.67 and there is a clear correlation between prediction accuracy and RMSD values with the correlation coefficient of −0.73 (see Fig. 6). Therefore, more accurate model structure leads to higher accuracy for binding site prediction. Similarly, the interface nucleotides identification for RNAs in the RNA–protein dataset also reached an average PPV of 0.66.
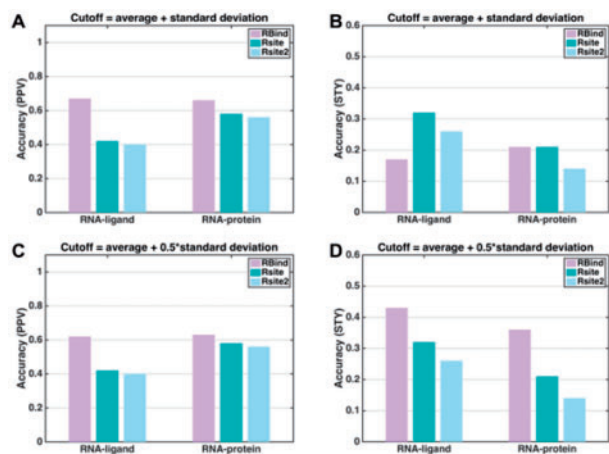
Fig. 7. Comparison of RBind, Rsite and Rsite2 methods. The PPV and STY results of RBind, Rsite and Rsite2 are colored in light purple, green and light blue. RBind performs consistently better PPV results using two cutoff values (**A** and **C**) than the existing Rsite and Rsite2 method for both RNA–ligand and RNA–protein datasets. For STY calculation, RBind predicted less experimental positions than Rsite and Rsite2 for RNA–ligand dataset but performs similar as Rsite and predicted more experimental positions than Rsite2 using a strict cutoff (**B**). However, RBind performs consistently better STY results than the existing Rsite and Rsite2 method for both RNA–ligand and RNA–protein datasets using a soft cutoff values (**D**)

3.4 Comparison with existing methods

Rsite and Rsite2 are computational methods to predict the potential RNA binding sites using tertiary structure (Rsite) or secondary structure (Rsite2) information (Zeng and Cui, 2016; Zeng *et al.*, 2015). To compare our method with existing methods, we also reproduced the prediction on binding sites by both Rsite and Rsite2 for 19 RNAs in the RNA–ligand dataset and 47 RNAs in the RNA–protein dataset. We performed Rsite and RBind using same model structures built from RNA sequences without any input from the experimentally determined structures.

As shown in Figure 7, RBind predictions (average PPV = 0.67) are significantly more accurate than Rsite (average PPV = 0.42) and Rsite2 predictions (average PPV = 0.40) for the RNA–ligand dataset (Fig. 7A). Similarly, for the RNA–protein dataset, RBind (average PPV = 0.66) also outperforms Rsite (average PPV = 0.58) and Rsite2 (average PPV = 0.56) (Fig. 7A). We calculated the sensitivity (STY) which measures the fraction of true binding sites that can be predicted. RBind showed less accuracy (average STY = 0.17) than Rsite (average STY = 0.32) and Rsite2 (average STY = 0.26) for RNA–ligand dataset (Fig. 7B). However, RBind (average STY = 0.21) performs similar as Rsite (average STY = 0.21) and better than Rsite2 (average STY = 0.14) for RNA–protein (Fig. 7B). In the prediction, RBind identified the potential binding sites by using a strict cutoff (average + standard deviation). We also predicted the potential binding sites by using a soft cutoff (average + 0.5*standard deviation). The results showed that RBind performs consistently better PPV and STY results than the existing Rsite and Rsite2 for both datasets (Fig. 7C and D).

4 Discussion and conclusion

The current Rsite and Rsite2 algorithms for RNA binding site prediction produce many false positive predictions that are located far away from the binding sites. The spatial diversity of these false positive positions only exacerbates wrong complex structures explored by docking. As demonstrated above, RBind achieves reliable
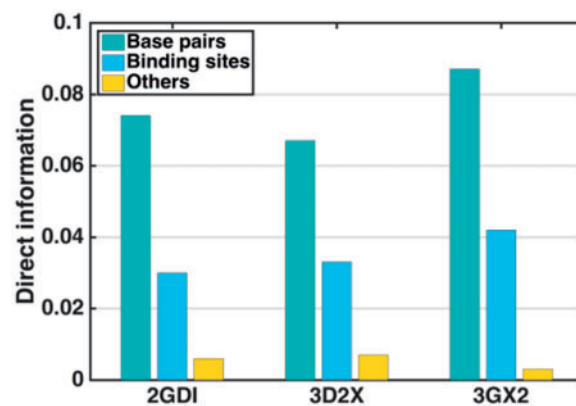

Fig. 8. The nucleotide–nucleotide coevolution analysis. We performed direct coupling analysis on nucleotide–nucleotide coevolution patterns for base pair interactions (green), binding sites interactions (blue) and other interactions (yellow). The results show that the DI values for base pair interactions and binding sites interactions are several times higher than other interactions indicating coevolution patterns in binding site regions

predictions for both ligand and interface binding sites, especially when the binding induces only small conformational changes, but the accuracy decreases when the induced conformation changes are significant. Moreover, the false positive predictions by RBind are typically located in the binding site area next to the catalytic pocket. This spatial proximity will reduce the impact produced by the false positives towards dock simulations.

The ligands first recognize the binding cavity, and then impact the global structure for functions. Therefore, both short-range binding cavity and long-range allosteric effect are important toward function. We analyzed the positions of the high degree nucleotides (cutoff = 1* standard deviation) in RNA–ligand dataset (Supplementary Fig. S2). The 98% nucleotides locate at or near the loop regions (less than 5 base pairs distance from loop regions) suggest the degree calculation is able to identify the RNA binding cavity. Our previous research showed that the closeness cutoff is able to identify critical residues for long-range allosteric effect (Zhao *et al.*, 2017). Taken together, RBind provides a unified and quantitative network framework to use graph theory such as degree and closeness that are exactly the right quantity to describe the binding cavity and long-range allosteric effect of an RNA structure.

The present study only utilizes a static network, i.e. the closeness and degree of nucleotides are computed from a static RNA complex structure. Prediction may be improved by probing conformational flexibility via dynamical network analysis on a series of structural configurations generated from molecular dynamic simulations (Chen *et al.*, 2014; Sethi *et al.*, 2009; Zhao *et al.*, 2015). In addition, most nucleotides involved in binding or catalytic pocket are either conserved or co-evolving. Other methods such as direct coupling analysis (DCA) that can detect co-evolving contacts from sequence analysis may complement the present network binding site prediction (de Juan *et al.*, 2013; De Leonardis *et al.*, 2015; Marks *et al.*, 2012; Morcos *et al.*, 2011; Weinreb *et al.*, 2016; Xing *et al.*, 2016). In fact, we performed DCA for nucleotide–nucleotide coevolution analysis of base pair interactions, binding sites interactions and other interactions. The results in Figure 8 show that the direct information values for both base pair interactions and binding sites interactions are several times higher than that for other interactions. This strongly suggests the presence of coevolution patterns in binding site regions that may be taken advantage of. Another limitation of the present network strategy is that the RNA tertiary structure is

required for prediction. However, we demonstrated that the current computation methods for modeling RNA tertiary structure are capable of building reasonably accurate RNAs with complex topology (Biesiada *et al.*, 2016; Boniecki *et al.*, 2016; Das and Baker, 2007; Jonikas *et al.*, 2009; Krokhotin *et al.*, 2015; Parisien and Major, 2008; Popenda *et al.*, 2012; Wang *et al.*, 2015; Xu *et al.*, 2014; Zhao *et al.*, 2012). We can thus first build an RNA tertiary structure from its sequence and then predict its binding sites from the structure.

In summary, we introduced a network approach, RBind, to identify critical nucleotides for binding. Our results suggest that the high-closeness nucleotides act as hubs for inter-network communication and thus serve as the binding sites to interact with other molecules, while the high-degree nucleotides provide local pocket information. In the future, we will incorporate additional such features as the dynamical motions and co-evolution contacts to enhance the prediction accuracy and make RBind a reliable and user-friendly tool for predicting RNA binding sites even for the hard targets.

Funding

Conflict of Interest: none declared.

References

Alipanahi,B. *et al.* (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.*, **33**, 831–838.

Amitai,G. *et al.* (2004) Network analysis of protein structures identifies functional residues. *J. Mol. Biol.*, **344**, 1135–1146.

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

Biesiada,M. *et al.* (2016) Automated RNA 3D structure prediction with RNAComposer. *Methods Mol. Biol.*, **1490**, 199–215.

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

Chen,H. *et al.* (2014) Break CDK2/Cyclin E1 interface allosterically with small peptides. *PLoS One*, **9**, e109154.

Das,R. and Baker,D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. USA*, **104**, 14664–14669.

de Juan,D. *et al.* (2013) Emerging methods in protein co-evolution. *Nat. Rev. Genet.*, **14**, 249–261.

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.

Esteller,M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.

Garst,A.D. *et al.* (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb. Perspect. Biol.*, **3**, a003533.

Gong,Z. *et al.* (2014) Insights into ligand binding to PreQ1 Riboswitch Aptamer from molecular dynamics simulations. *PLoS One*, **9**, e92247.

Gong,Z. *et al.* (2011) Role of ligand binding in structural organization of add A-riboswitch aptamer: a molecular dynamics simulation. *J. Biomol. Struct. Dyn.*, **29**, 403–416.

Gong,Z. *et al.* (2012) Computational study of unfolding and regulation mechanism of preQ1 riboswitches. *PLoS One*, **7**, e45239.

Greene,L.H. and Higman,V.A. (2003) Uncovering network systems within protein structures. *J. Mol. Biol.*, **334**, 781–791.

Huang,S.Y. and Zou,X. (2013) A nonredundant structure dataset for benchmarking protein–RNA computational docking. *J. Comput. Chem.*, **34**, 311–318.

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

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

Magnus,M. *et al.* (2014) Computational modeling of RNA 3D structures, with the aid of experimental restraints. *RNA Biol.*, **11**, 522–536.

Marks,D.S. *et al.* (2012) Protein structure prediction from sequence variation. *Nat. Biotechnol.*, **30**, 1072–1080.

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

Morcos,F. *et al.* (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. USA*, **108**, E1293–E1301.

Murakami,Y. *et al.* (2010) PiRaNhA: a server for the computational prediction of RNA-binding residues in protein sequences. *Nucleic Acids Res.*, **38**, W412–W416.

Nawrocki,E.P. *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.

Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.

Paz,I. *et al.* (2016) BindUP: a web server for non-homology-based prediction of DNA and RNA binding proteins. *Nucleic Acids Res.*, **44**, W568–W574.

Philips,A. *et al.* (2013) LigandRNA: computational predictor of RNA–ligand interactions. *RNA*, **19**, 1605–1616.

Popenda,M. *et al.* (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.*, **40**, e112.

Sethi,A. *et al.* (2009) Dynamical networks in tRNA: protein complexes. *Proc. Natl. Acad. Sci. USA*, **106**, 6620–6625.

Shannon,P. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Shi,M. *et al.* (2016) Redefining the invertebrate RNA virosphere. *Nature*.

Sun,L.Z. *et al.* (2017) Theory and Modeling of RNA structure and interactions with metal ions and small molecules. *Annu. Rev. Biophys.*, **46**, 227–246.

Terribilini,M. *et al.* (2007) RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res.*, **35**, W578–W584.

Wang,J. *et al.* (2015) 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res.*, **43**, e63.

Wang,L. and Brown,S.J. (2006) BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Res.*, **34**, W243–W248.

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

Xing,S. *et al.* (2016) Tcf1 and Lef1 transcription factors establish CD8(+) T cell identity through intrinsic HDAC activity. *Nat. Immunol.*, **17**, 695–703.

Xu,X. *et al.* (2014) Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One*, **9**, e107504.

Yang,Y. *et al.* (2012) A new size-independent score for pairwise protein structure alignment and its application to structure classification and nucleic-acid binding prediction. *Proteins*, **80**, 2080–2088.

Yesselman,J.D. and Das,R. (2016) Modeling small noncanonical RNA motifs with the Rosetta FARFAR server. *Methods Mol. Biol.*, **1490**, 187–198.

Zeng,P. and Cui,Q. (2016) Rsite2: an efficient computational method to predict the functional sites of noncoding RNAs. *Sci. Rep.*, **6**, 19016.

Zeng,P. *et al.* (2015) Rsite: a computational method to identify the functional sites of noncoding RNAs. *Sci. Rep.*, **5**, 9179.

Zhao,Y. *et al.* (2011) Improvements of the hierarchical approach for predicting RNA tertiary structure. *J. Biomol. Struct. Dyn.*, **28**, 815–826.

Zhao,Y. *et al.* (2012) Automated and fast building of three-dimensional RNA structures. *Sci. Rep.*, **2**, 734.

Zhao,Y. *et al.* (2017) Network analysis reveals the recognition mechanism for dimer formation of bulb-type lectins. *Sci. Rep.*, **7**, 2876.

Zhao,Y. *et al.* (2016) NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.*, **44**, D203–D208.

Zhao,Y. *et al.* (2013) A new role for STAT3 as a regulator of chromatin topology. *Transcription*, **4**, 227–231.

Zhao,Y.J. *et al.* (2015) Molecular dynamics simulation reveals insights into the mechanism of unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain. *Plos One*, **10**, e0124377.