Computational and Structural Biotechnology Journal 18 (2020) 3762–3765

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

RBinds: A user-friendly server for RNA binding site prediction

Huiwen Wang, Yunjie Zhao *

Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, China

ARTICLE INFO

Article history:
Received 17 September 2020
Received in revised form 27 October 2020
Accepted 31 October 2020
Available online 24 November 2020

Keywords:
RNA
Binding site prediction
Network

ABSTRACT

RNA performs various biological functions by interacting with other molecules. The knowledge of RNA binding sites is essential for the understanding of RNA-protein or RNA-ligand complex structures and their mechanisms. However, the RNA binding site prediction study requires tedious programming scripts and manual handling. One user-friendly bioinformatics tool for RNA binding site prediction has been missing. This limitation motivated us to develop the RBinds, a user-friendly web server, to predict the RNA binding site using a simple graphical user interface. Some advanced features implemented in RBinds are (1) transforming the RNA structure to a network automatically; (2) analyzing the structural network properties to predict binding site; (3) constructing one annotated force-directed network; (4) providing a visualization tool for users to scale and rotate the structure; (5) offering the related tools to predict or simulate RNA structures. RBinds web server is a reliable and user-friendly tool and facilitates the RNA binding site study without installing programs locally. RBinds is freely accessible at http://zhaoserver.com.cn/RBinds/RBinds.html.

1. Introduction

RNA performs various biological processes by interacting with other molecules [1–3]. Therefore, the comprehensive knowledge of RNA-protein or RNA-ligand complex structures is essential for the understanding of their functions, disease pathogenesis, and drug design [4,5]. Currently, it is still challenging to determine the RNA-protein or RNA-ligand complex structures experimentally and computationally. The starting point for RNA complex structure study is to determine the potential binding sites.

Some experimental methods can identify the RNA binding sites, including CLIP-seq [6,7], RIP-seq [8], RNase footprinting [9,10], and SHAPE-MaP [11]. Unfortunately, it is still challenging to precisely determine the binding sites from direct experimentation due to time-consuming and technology limitations. The structure-based computational methods are feasible for binding sites prediction [12,13]. The current available two computational tools are Rsite and Rsite2. These two approaches are based on the tertiary and secondary structure to predict RNA binding sites, respectively [14,15]. Rsite and Rsite2 hypothesize that both the most connected and the most non-connected nucleotides in RNA structure are potential binding sites. These methods first calculate the distances between each nucleotide and all the other nucleotides in RNA structure, and then define the extreme nucleotides in the distance curve as the RNA binding sites. The small benchmark testing shows PPV around 0.5 and sensitivity higher than 0.8. Besides, Rsite needs to be installed locally, while Rsite2 only provides the binding sites but without visualization information. One accurate and user-friendly bioinformatics tool for RNA binding site prediction has been missing.

Previous research showed that the network strategy is effective in extracting structural features. For example, network-based methods have already been used in protein folding, protein–protein, and protein–ligand study [16–19]. The network approach may be a quantitative method for RNA predictions [20]. Therefore, we have proposed a network approach, RBind, to improve the RNA binding site prediction accuracy [21]. 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 the long-range allosteric effect to identify the binding sites. The RNA-protein and RNA-ligand testing datasets show that RBind has a reliable accuracy for RNA binding sites prediction.

RBind has been used by many researchers [4,22]. However, the original RBind algorithm needs to be installed locally. To provide an automatic prediction method, we now offer one easy-to-use web server using only RNA tertiary structures as input information. Compared with current RNA binding site prediction methods,

* Corresponding author.
E-mail address: yjzhaowh@mail.ccnu.edu.cn (Y. Zhao).

https://doi.org/10.1016/j.csbj.2020.10.043

RBinds provides an intuitive user interface, multiple outputs, and visualizations with higher prediction accuracy.

2. Workflow and implementation

RBinds provides a user-friendly online server for RNA binding site prediction. The server contains six modules: Home, Visualization, Links, Output example, Tutorial, and Contacts. The detailed information for each module is as follows. (1) The Home module provides an input box where users can upload an RNA tertiary structure in PDB format. (2) An RNA tertiary structure can be uploaded and visualized in the Visualization module. The RNA structure is shown as a red cartoon representation by default. Users can change the structural color and trace, cartoon, or ball&-stick representations. The key residues can be highlighted in different colors. Users can also scale, rotate the RNA structures, and save the image. (3) The Links module provides RNA tertiary structure prediction, docking, molecular dynamics, and other related useful RNA resources. (4) The Output example module provides one tRNA prediction example. (5) The Tutorial module offers an introduction to the RBinds server. (6) The Contacts module provides our emails for users to comment or ask questions.

The RBinds workflow (Fig. 1) illustrates the computational steps of the webserver: (1) Users only need to upload an RNA tertiary structure in PDB format. (2) RBinds transforms the RNA tertiary structure into the corresponding network. (3) RBinds analyzes the structural network properties (closeness and degree) to predict the binding sites. (4) An output page includes (a) the binding site prediction results; (b) network visualization; (c) the statistical analysis of the closeness and degree results; and (d) the download link of prediction results. (5) Users can get the nucleotide name, nucleotide number, closeness, and degree values if they put the mouse at the node in the network. (6) Users can visualize the RNA tertiary structure in the Visualization module. In contrast to other existing RNA binding site prediction servers, RBinds provides much more computational results with only structure information. For example, RBinds provides the annotated force-directed network for users to understand the network topology. Moreover, RBinds also offers the links of PyMOL, Chimera, and other RNA related software.

2.1. File input

The input of RBinds is RNA tertiary structure in PDB format. The RNA tertiary structure can be downloaded from Protein Data Bank (PDB) database [23], or predicted by RNA tertiary structure prediction methods, such as 3dRNA [24,25], Vfold3D [26], and iFoldRNA [27]. Users can first upload the structure information and then

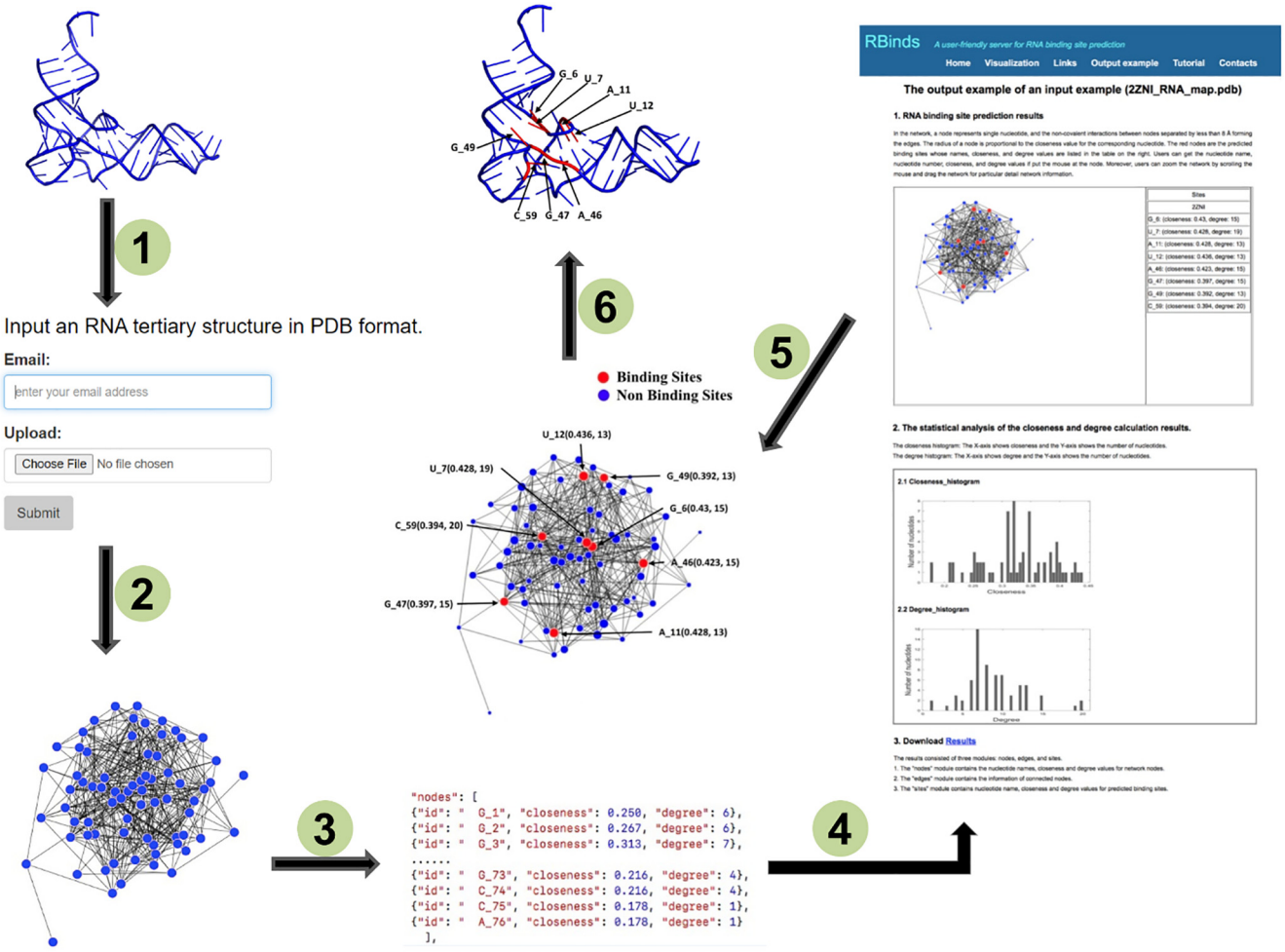

Fig. 1. The workflow of RBinds. (1) Submitting box. RBinds uses the PDB file as input information. (2) Network construction. RBinds transforms the input RNA structure to the network. (3) Network properties calculation. RBinds calculates closeness and degree to predict the binding sites. (4) Output page. The output results include (a) the binding site predictions, (b) network visualization, (c) the statistical analysis of the closeness and degree, and (d) the download link of the prediction results. (5) The annotated force-directed network. (6) RNA structure labeled with binding sites.

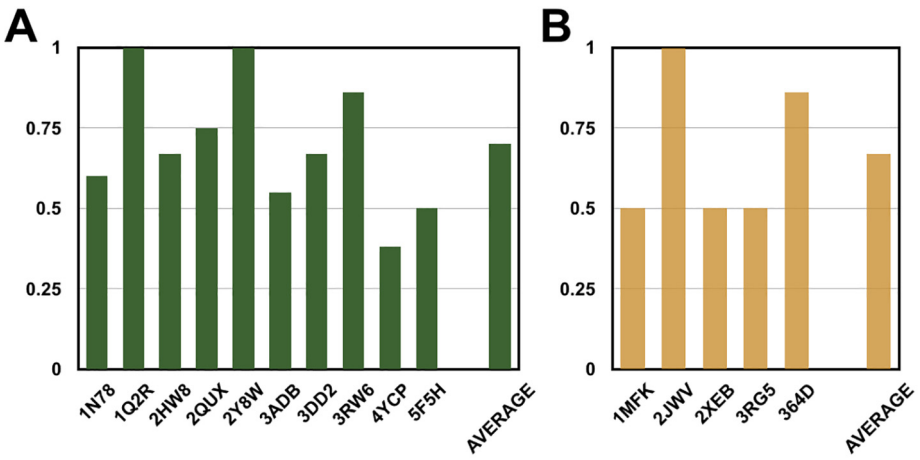

Fig. 2. The prediction performance of the RBinds in the RNA-protein benchmark. (A) The average binding site prediction accuracy is 0.70 in the rigid complex testing set. (B) The average binding site prediction accuracy is 0.67 in the flexible complex testing set.

click "Submit" to run the prediction task in the Home module. If the user submits multiple RNA chains, RBinds will transform the RNAs into one network for binding site prediction. If the user submits one RNA-ligand or RNA-protein complex, RBinds will ignore the ligand or protein information. We also provide one input file example for users to get familiar with the server.

2.2. Network construction

It is generally believed that RNA binding first recognizes the binding pocket and then impacts the global structure for functions. Therefore, both short-range binding pocket and long-range allosteric effects are important toward function. Our previous research suggests that the degree is able to identify the RNA binding pocket, and the closeness cutoff is able to identify critical positions for long-range allosteric effect. Therefore, RBinds uses the short-range degree calculation to identify RNA pockets. Then, it uses the closeness calculation to determine the nucleotides for long-range allosteric effects. The workflow of the algorithm is as follows: (1) The RNA molecule is transformed into a network with nodes and edges (Fig. 1). In the network, a single nucleotide is considered as a node. An edge represents a non-covalent interaction if the shortest distance between the heavy atoms of the two nonconsecutive nucleotides is less than 8 Å. (2) The closeness and degree values of each node are calculated to identify the binding sites. 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:

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

where n is the total number of nucleotides in the network, the $d(x,y)$ is the distance of the shortest path between the node $\times$ and y [13]. The shortest path between two nodes is calculated by the Floyd-Warshall algorithm. The degree of a node is defined as the number of edges connected to the node. (3) RBinds identified the nodes as binding sites when their closeness and degree values are higher than the corresponding average values by a standard deviation, cutoff = average + standard deviation. (Please see the reference [21] for detailed information).

2.3. Backend

The backend of the RBinds uses HTML, JavaScript, PHP, MATLAB, D3.js, and JSmol. The web server includes Home, Visualization, Links, Output example, Tutorial, and Contact modules that are built using HTML scripts. The PHP script will be activated when users upload an RNA tertiary structure in PDB format, and then click the submit button in the "Home" module. The PHP script will invoke an executable program that is generated using a MATLAB script, D3.js library, and JSmol library. The network construction, network property analysis, and binding site prediction are performed using an executable program. The network visualization is implemented by the D3.js library. RNA tertiary structure visualization is implemented by the JSmol library.

3. Output and visualizations

After submitting the prediction task, the server will jump to the output page with the following results. (a) RNA binding site prediction results with the corresponding annotated force-directed network. Users can zoom the network by scrolling the mouse and drag the network for detailed network information. In the network, a node represents a single nucleotide. The radius of a node is proportional to the closeness value for the corresponding nucleotide. Users can get the nucleotide name, number, closeness, and degree values if they put the mouse at the node. The edges represent non-covalent interactions between nucleotides. (b) The statistical analysis of the closeness and degree calculation results can be downloaded separately. In the closeness/degree histogram, the X-axis shows closeness/degree values, and the Y-axis indicates the number of nucleotides. (c) The prediction results can be downloaded with the following information. The "nodes" module contains the nucleotide names, closeness, and degree values. The "edges" module provides the information of connected node pairs. The "sites" module includes nucleotide name, closeness, and degree values for predicted binding sites.

4. Case studies

The previous study has already shown that RBind performs consistently better than existing methods in both RNA-protein and RNA-ligand testing sets [21]. The RNA-protein testing set includes 72 RNA-protein structures, while the RNA-ligand testing set consists of 22 RNA-ligand structures. In the testing, we used the unbound RNA structures to predict the RNA binding sites. The results showed the average accuracy of 0.63 and 0.82 in RNA-protein and RNA–ligand, respectively [21]. To validate the RBinds reliability, we further tested the web server in the RNA-protein benchmark consists of ten rigid docking complexes and five flexible docking complexes [22]. The results show the average accuracy

of 0.70 and 0.67 in rigid and flexible testing sets, respectively (Fig. 2). The RNA-protein benchmark can be downloaded in the Tutorial module. The average running time is less than one minute when users submit an RNA tertiary structure within 100 nucleotides.

In summary, we built a user-friendly web server for predicting RNA binding sites automatically. Users can now use RBinds easily. RBinds can facilitate users for RNA complex study. We will consider the interacting partners for more accurate predictions in the future.

CRediT authorship contribution statement

H.W. prepared the data sets, carried out the the most computational analysis, and wrote the server softwares. Y.Z. designed, supervised the overall study, analyzed the data, and wrote the paper. All authors edited the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the National Natural Science Foundation of China 11704140 and self- determined research funds of CCNU from the colleges' basic research and operation of MOE CCNU20TS004.

References

[1] Panhale A, Richter FM, Ramírez F, Shvedunova M, Manke T, Mittler G, et al. CAPRI enables comparison of evolutionarily conserved RNA interacting regions. Nat Commun 2019;10:1–22.
[2] Sullenger BA, Gilboa E. Emerging clinical applications of RNA. Nature 2002;418:252–8.
[3] Liu Z-P, Wu L-Y, Wang Y, Zhang X-S, Chen L. Prediction of protein–RNA binding sites by a random forest method with combined features. Bioinformatics 2010;26:1616–22.
[4] Chakrabarty B, Naganathan V, Garg K, Agarwal Y, Parekh N. NAPS update: network analysis of molecular dynamics data and protein–nucleic acid complexes. Nucleic acids Res 2019;47:W462–70.
[5] Jian Y, Wang X, Qiu J, Wang H, Liu Z, Zhao Y, et al. DIRECT: RNA contact predictions by integrating structural patterns. BMC Bioinf 2019;20:497.
[6] Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 2010;141:129–41.
[7] Wang T, Xie Y, Xiao G. dCLIP: a computational approach for comparative CLIP-seq analyses. Genome Biol 2014;15:R11.
[8] Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 2010;40:939–53.
[9] Liu G, Mercer TR, Shearwood AMJ, Siira SJ, Hibbs ME, Mattick JS, et al. Mapping of mitochondrial RNA-protein interactions by digital RNase footprinting. Cell Rep 2013;5:839–48.
[10] Peng Y, Soper TJ, Woodson SA. RNase footprinting of protein binding sites on an mRNA target of small RNAs, in Bacterial Regulatory RNA. Bacterial Regulatory RNA. Humana Press. Totowa. NJ. 2012; 213-224
[11] Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat Methods 2014;11:959–65.
[12] De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 2006;34:W362–5.
[13] Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, et al. Network analysis of protein structures identifies functional residues. J Mol Biol 2004;344:1135–46.
[14] Zeng P, Li J, Ma W, Cui Q. Rsite: a computational method to identify the functional sites of noncoding RNAs. Sci Rep 2015;5:9179.
[15] Zeng P, Cui Q. Rsite2: an efficient computational method to predict the functional sites of noncoding *RNAs*. Sci Rep 2016;6:19016.
[16] Bagler G, Sinha S. Assortative mixing in Protein Contact Networks and protein folding kinetics. Bioinformatics 2007;23:1760–7.
[17] Taylor NR. Small world network strategies for studying protein structures and binding. Comput Struct Biotechnol J 2013;5:e201302006.
[18] Wang H, Zhao Y. Methods and applications of RNA contact prediction. Chin Phys B 2020;29:108708.
[19] Zhou T, Wang H, Song L, Zhao Y. Computatioanl study of switching mechanism in add A-riboswitch. J Theor Comput Chem 2020;19:2040001.
[20] Lee C-Y, Lee JC, Gutell RR. Networks of interactions in the secondary and tertiary structure of ribosomal RNA. Physica A 2007;386:564–72.
[21] Wang K, Jian Y, Wang H, Zeng C, Zhao Y. RBind: computational network method to predict RNA binding sites. Bioinformatics 2018;34:3131–6.
[22] Nithin C, Mukherjee S, Bahadur RP. A non-redundant protein-RNA docking benchmark version 2.0. Proteins 2017;85:256–67.
[23] Parasuraman S. Protein data bank. J Pharmacol Pharmacother 2012;3:351–2.
[24] Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y. Automated and fast building of three-dimensional RNA structures. Sci Rep 2012;2:734.
[25] Wang J, Wang J, Huang Y, Xiao Y. 3dRNA v2. 0: An Updated Web Server for RNA 3D Structure Prediction. Int J Mol Sci 2019; 20:4116.
[26] Xu X, Zhao P, Chen S-J. Vfold: a web server for RNA structure and folding thermodynamics prediction. PLoS ONE 2014;9:e107504.
[27] Sharma S, Ding F, Dokholyan NV. iFoldRNA: three-dimensional RNA structure prediction and folding. Bioinformatics 2008;24:1951–2.