Bioinformatics, 34(7), 2018, 1238–1240
doi: 10.1093/bioinformatics/btx742
Advance Access Publication Date: 27 November 2017
Applications Note

OXFORD

Structural bioinformatics

3dRPC: a web server for 3D RNA–protein structure prediction

Yangyu Huang, Haotian Li and Yi Xiao*

Institute of Biophysics, School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

*To whom correspondence should be addressed.
Associate Editor: Ivo Hofacker

Received on July 20, 2017; revised on November 10, 2017; editorial decision on November 11, 2017; accepted on November 24, 2017

Abstract

RNA–protein interactions occur in many biological processes. To understand the mechanism of these interactions one needs to know three-dimensional (3D) structures of RNA–protein complexes. 3dRPC is an algorithm for prediction of 3D RNA–protein complex structures and consists of a docking algorithm RPDOCK and a scoring function 3dRPC-Score. RPDOCK is used to sample possible complex conformations of an RNA and a protein by calculating the geometric and electrostatic complementarities and stacking interactions at the RNA–protein interface according to the features of atom packing of the interface. 3dRPC-Score is a knowledge-based potential that uses the conformations of nucleotide-amino-acid pairs as statistical variables and that is used to choose the near-native complex-conformations obtained from the docking method above. Recently, we built a web server for 3dRPC. The users can easily use 3dRPC without installing it locally. RNA and protein structures in PDB (Protein Data Bank) format are the only needed input files. It can also incorporate the information of interface residues or residue-pairs obtained from experiments or theoretical predictions to improve the prediction.

Availability and implementation: The address of 3dRPC web server is http://biophy.hust.edu.cn/3dRPC.

Contact: yxiao@hust.edu.cn

RNA–protein interactions play important roles in many biological processes. The detailed information of these interactions can be obtained from 3D structures of RNA–protein complexes. But at present the number of experimental 3D RNA–protein structures is very limited. Many computational algorithms for predicting RNA–protein complex structures have been proposed (Chen, 2004; Li *et al.*, 2012; Perez-Cano *et al.*, 2010; Tuszynska and Bujnicki, 2011; Zhang *et al.*, 2017), including our protocol 3dRPC (Huang *et al.*, 2013).

3dRPC originally consisted of a docking procedure RPDOCK and a scoring function DECK-RP (Huang *et al.*, 2013). RPDOCK is a docking procedure specific to RNA–protein complexes to sample the conformational space of a RNA–protein complex. It used a different set of parameters from those for protein–protein complexes to calculate the geometric complementarity since the atom packing at the RNA–protein interface is different from that at the protein–protein interface. It also includes electrostatic effect and stacking interactions between aromatic side chains and bases. DECK-RP is a distance- and environment-dependent, coarse-grained and knowledge-based potential for RNA–protein complexes that uses an improved reference state. In the current version 3dRPC, DECK-RP has been replaced by a new knowledge-based potential 3dRPC-Score (Li *et al.*, 2017). It uses the conformations of nucleotide-amino-acid pairs as statistical variables and takes into account both distance and orientation between the amino acid and nucleotide of a pair instead of the residue-residue distances since the energy of a nucleotide-amino-acid pair depends on its conformations. 3dRPC-Score was benchmarked on different unbound docking decoy sets (Li *et al.*, 2017), and compared with DECK-RP and ITScore-PR (Huang and Zou, 2014). The results showed that 3dRPC-Score performs better than DECK-RP. Furthermore, 3dRPC-Score is good for the decoys sets with near-native structures of lower quality while ITScore-PR is good for the decoys sets with near-native structures of

higher quality. In general, 3dRPC-Score performs consistently well for different test sets.

3dRPC has been used by many researchers (Cruz *et al.*, 2015; Rabal *et al.*, 2016). But the original 3dRPC program was not a web server and needed to be installed locally by users and so was not so easy to use, although we have given the detailed steps of installing 3dRPC locally (Huang *et al.*, 2016). Now we have built a web server of 3dRPC and the users just need to input the 3D structures of the partners of RNA–protein complexes and click the 'submit' button after which they can obtain the predicted complex structures automatically.

In the following we give a brief description of the algorithm and web server of 3dRPC. The algorithm of 3dRPC is divided into two steps:

(1) Rigid-body docking by RPDOCK (Huang *et al.*, 2013). RPDOCK is an FFT-based rigid-body sampling method, which is similar to the protein–protein docking algorithm FTDOCK (Gabb *et al.*, 1997). First, the protein is discretized into a three-dimensional grid and the RNA is rotated in Euler angles and then discretized into the three-dimensional grid too. Next, a full translation scan is performed. After the translation scan, top three poses are retained according to the RPDOCK score. FFT is used to accelerate the calculation and the structures of the complexes are sampled in Fourier phase space. The process is repeated until full rotation scan is completed. The RPDOCK score is composed of two items: geometric complementarity (GC) and electrostatics (ELEC). The electrostatics is calculated by Coulomb's formula with a distance-dependent dielectric and the charge is extracted from AMBER force field (Case *et al.*, 2005). In order to consider the effect of stacking interactions between aromatic side chains and bases, they are assigned different weights. In the current web-server version RPDOCK can also incorporate information of interface nucleotides/amino-acids predefined or obtained from experiments or theoretical predictions [e.g. SRCpred (Fernandez *et al.*, 2011)] to improve the prediction by assigning a larger confidence factor to the restrained nucleotides/amino-acids during the FFT, which uses the same method that we used for protein–protein docking (Li *et al.*, 2013).

(2) Scoring by 3dRPC-Score (Li *et al.*, 2017). In this step the models generated by RPDOCK are scored by 3dRPC-Score. For each model, all the nucleotide-amino-acid pairs are extracted from the complex and their all-atom RMSDs in relative to the standard pairs of 3dRPC-Score are calculated, respectively. The score of a pair is set to that of the standard pair that has the minimum RMSD value in relative to it. The score of the model is the sum of the scores of all the pairs. Then, the generated models are ranked according to their scores. Similarly, in the current web-server version 3dRPC-Score can use the distance restraints of interface nucleotide-amino-acid pairs predefined or obtained from experiments or theoretical predictions to filter the generated models first (Fig. 1).

The 3dRPC webserver has a mainpage that contains a new task area, a submitted tasks area, a task query area and an area with references, from the top down. Users can directly use 3dRPC and need not to register. In the new task area there are five input fields (email, protein structure, RNA structure, number of predictions and scoring functions). In this area there also includes an 'Advanced settings' in which the users can input the information of interface residues or residue-pair distances obtained from experiments or theoretical predictions, which can be used to improve the prediction accuracy. The main steps of predicting 3D RNA–protein complex structures by using 3dRPC webserver is as follows:

(1) Firstly, load PDB-formated structure files of the protein and RNA partners in the textboxes 'Protein structure' and 'RNA Structure', respectively; Secondly, input how many models you want to predict in the textbox 'Number of Predictions' and the default value is 10. Thirdly, you can select scoring functions to use. At present there are two scoring functions: 3dRPC-Score and RPDECK. The default is 3dRPC-Score. Fourthly, if you have information of interface residues or nucleotide-amino-acid pair and their distances, you can input the restraints in the 'Advanced Settings'. For the interface residues you can set the confidence factor (between 0 and 1) to the restrained nucleotides or amino acids for the FFT. The choice of this factor value depends on the reliability of the information of the interface residues obtained from experiments or predictions. The factor value can be set to 1 if the information is completely accurate. Finally, click the button 'Submit' to submit the task. The users can input their email address, and then the server would send an email automatically to remind the user after finishing the task.

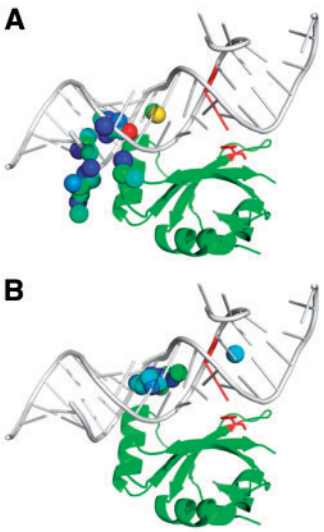

Fig. 1. The top 32 prediction results of an RNA–protein complex (PDB ID: 1DFU) not using (**A**) and using (**B**) the distance restraints between amino acid and nucleotide (V31Y B73A 7 Å, red) as filter. The native structures of the RNA and protein partners are in light and dark colors, respectively. The spheres represent predicted 32 positions of the centers of mass of the RNA molecule in relative to the protein. In (B) 30 of 32 predictions are near-native structures (RMSD < 10 Å) (Color version of this figure is available at *Bioinformatics online*.)

(2) After the submission of the task the user is presented a waiting page. Once the task is finished, the waiting page would refresh automatically to turn to the results page. The result page contains a list of predicted structures and their respective scores and links to download them or view them with the web browser. Clicking 'view' can view a predicted 3D structure directly in the web browser. JSmol is used for 3D structure visualization (Hanson and Lu, 2017). Clicking 'download' can download the PDB-formated file of a predicted 3D structure. Clicking 'download all' would download all the predicted structures. In the results page there are the information of the used scoring function and number of predictions. There you can also download the file of used parameters, which can be directly used as input file in the offline version of 3dRPC.

3dRNA relies on the methodology of rigid body docking and so its prediction accuracy of a RNA–protein complex may be lower when there are larger changes between the unbound and bound conformations of the partners. It mainly applies to the cases where the structures of both RNA and protein are rather well established and are not expected to change substantially due to complex formation. However, RPDOCK program considered the fact that the atom packing of RNA–protein interface was looser than that of the protein–protein

interface and it can account for small amounts of flexibility between unbound and bound conformations of the partners of an RNA–protein complex by not penalizing minor steric clashes in the interface region (Huang *et al*., 2013; Katchalski-Katzir *et al*., 1992).

It is also noted that only heavy atoms are considered during docking and scoring in 3dRPC because standard structures from PDB are without hydrogens. The docking program allows existence of a thick surface of the partners of the complex and some clashes between them in the complex structures as mentioned above. The geometric effect of hydrogen atoms can be accounted by them. Furthermore, there are no special requirements for the input structures, which are just the standard PDB structure files. If the input structures contain nucleotides other than A, C, G, U or have missing amino acids or nucleotides, 3dRPC will neglect them automatically during calculations.

In summary, we built a web server for predicting 3D RNA–protein complex structures automatically. The users now can use 3dRPC easily.

Funding

This work was supported by the National Natural Science Foundation of China under Grant No. 11374113 and 31570722 and the National High Technology Research and Development Program of China under Grant No. 2012AA020402.

Conflict of Interest: none declared.

References

Case,D.A. *et al*. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem*., **26**, 1668–1688.

Chen,Y. (2004) A new hydrogen-bonding potential for the design of protein–RNA interactions predicts specific contacts and discriminates decoys. *Nucleic Acids Res*., **32**, 5147–5162.

Cruz,J.W. *et al*. (2015) Growth-regulating Mycobacterium tuberculosis VapC-mt4 toxin is an isoacceptor-specific tRNase. *Nat. Commun*., **6**, 7480.

Fernandez,M. *et al*. (2011) Prediction of dinucleotide-specific RNA-binding sites in proteins. *BMC Bioinformatics*, **12**, S5.

Gabb,H.A. *et al*. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol*., **272**, 106–120.

Hanson,R.M. and Lu,X.J. (2017) DSSR-enhanced visualization of nucleic acid structures in Jmol. *Nucleic Acids Res*, **45**, W528–W533.

Huang,S.Y. and Zou,X. (2014) A knowledge-based scoring function for protein–RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res*., **42**, e55.

Huang,Y. *et al*. (2016) Using 3dRPC for RNA–protein complex structure prediction. *Biophys. Rep*., **2**, 95–99.

Huang,Y. *et al*. (2013) A novel protocol for three-dimensional structure prediction of RNA–protein complexes. *Sci Rep*., **3**, 1887.

Katchalski-Katzir,E. *et al*. (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA*, **89**, 2195–2199.

Li,C.H. *et al*. (2012) A new residue-nucleotide propensity potential with structural information considered for discriminating protein–RNA docking decoys. *Proteins*, **80**, 14–24.

Li,H. *et al*. (2017) A pair-conformation-dependent scoring function for evaluating 3D RNA–protein complex structures. *PLoS One*, **12**, e0174662.

Li,L. *et al*. (2013) How to use not-always-reliable binding site information in protein–protein docking prediction. *Plos One*, **8**, e75936.

Perez-Cano,L. *et al*. (2010) Structural prediction of protein–RNA interaction by computational docking with propensity-based statistical potentials. *Pac Symp. Biocomput*., 293–301.

Rabal,O. *et al*. (2016) In silico aptamer docking studies: from a retrospective validation to a prospective case study-TIM3 aptamers binding. *Mol. Ther. Nucleic Acids*, **5**, e376.

Tuszynska,I. and Bujnicki,J.M. (2011) DARS-RNP and QUASI-RNP: new statistical potentials for protein–RNA docking. *BMC Bioinformatics*, **12**, 348.

Zhang,Z. *et al*. (2017) A combinatorial scoring function for protein–RNA docking. *Proteins*, **85**, 741–752.