



Using 3dRPC for RNA–protein complex structure prediction

Yangyu Huang¹, Haotian Li¹, Yi Xiao¹✉

¹ Biomolecular Physics and Modeling Group, School of Physics, Huazhong University of Science and Technology, Wuhan 430074, China

Received: 19 July 2016 / Accepted: 5 January 2017 / Published online: 10 February 2017

Abstract 3dRPC is a computational method designed for three-dimensional RNA–protein complex structure prediction. Starting from a protein structure and a RNA structure, 3dRPC first generates presumptive complex structures by RPDock and then evaluates the structures by RPRANK. RPDock is an FFT-based docking algorithm that takes features of RNA–protein interactions into consideration, and RPRANK is a knowledge-based potential using root mean square deviation as a measure. Here we give a detailed description of the usage of 3dRPC. The source code is available at <http://biophy.hust.edu.cn/3dRPC.html>.

Keywords RNA–protein complex, Tertiary structure, Computational prediction, Docking, Scoring function

INTRODUCTION

RNA–protein interactions have drawn much attention recently since they might play important roles in many biological processes (Chen and Varani 2005; Glisovic et al. 2008). It was found that most of the human genome could be transcribed into RNAs but only a small fraction of these RNAs was translated into proteins (Cheng et al. 2005), i.e., most RNAs did not undergo translation. These non-coding RNAs perform their biological functions mostly through RNA–protein interactions and forming RNA–protein complexes. As the protein–protein interactions, the three-dimensional structures of RNA–protein complexes are essential to understand the mechanism of RNA–protein interactions. However, experimental determination of three-dimensional structures of RNA–protein complexes is still difficult and time-consuming at present. To solve this problem, computational methods have been proposed to predict the RNA–protein complex structures.

Most algorithms for predicting complex structure consist of two stages: sampling and scoring. The first stage is sampling conformational space and selecting candidates. Since the conformational space is very large,

a fast and effective sampling method is required. The second stage is evaluation of the candidates using a ranking or scoring function. Compared to the well-developed methods for protein–protein complex structure prediction (Vakser and Aflalo 1994; Gabb et al. 1997; Chen et al. 2003; Dominguez et al. 2003; Kozakov et al. 2006), those for RNA–protein complexes remain to be developed, which mainly focus on the scoring (Chen et al. 2004; Perez-Cano et al. 2010; Tuszynska and Bujnicki 2011; Li et al. 2012; Huang and Zou 2014), while the sampling methods were borrowed from those for protein–protein complex prediction (Vakser and Aflalo 1994; Gabb et al. 1997; Chen et al. 2003). Recently, we proposed a novel protocol for predicting RNA–protein complex structures—3dRPC (Huang et al. 2013). 3dRPC originally consists of a docking procedure RPDock and a scoring function DECK-RP.

RPDock is a docking procedure specific to RNA–protein docking. Based on the fact that the atom packing at the RNA–protein interface is different from that at the protein–protein interface (Jones et al. 1999, 2001; Bahadur et al. 2008), RPDock applies a new set of parameters to calculate the geometric complementarity. Since the electrostatics plays an important role in RNA–protein interaction (Jones et al. 2001; Kim et al. 2006; Terribilini et al. 2006; Bahadur et al. 2008; Kumar et al. 2008; Perez-Cano et al. 2010; Perez-Cano and

✉ Correspondence: yxiao@mail.hust.edu.cn (Y. Xiao)

Fernandez-Recio 2010), RPDock also includes electrostatic effect. RPDock also accounts for the stacking interactions between aromatic side chain and bases. The scoring function DECK-RP has been replaced in the updated 3dRPC by RPRANK, a new knowledge-based potential using Root mean square deviation (RMSD) as a measure. The statistical objects of RPRANK are the conformation differences between residue-base pairs. The residue-base pairs are clustered based on the RMSD between each other. Then the energies of the residue-base pair clusters are decided by statistical method based on the number of pairs in each cluster. Different from other statistical potential, this potential does not use distance to classify the residue-base pairs directly. The RMSD-based potential RPRANK has been tested on Zou's benchmarks (Huang and Zou 2013). The success rate reaches 29.1% for top one and 41.7% for top ten. 3dRPC has been tested on two test sets (Perez-Cano et al. 2012; Huang and Zou 2013) and achieved success rates of 12.1% and 31.9% for top one prediction and 28.8% and 41.7% for top ten, respectively. In the following, we give a detailed description of the usage of 3dRPC.

3dRPC

Stage 1: rigid-body docking by RPDock

RPDock is a FFT-based, rigid-body sampling method. The overall process of RPDock resembles protein-protein docking algorithm FTDOCK (Gabb et al. 1997). First, the protein is discretized into three-dimensional grid and the RNA is rotated by Euler angles and then discretized into three-dimensional grid. Next, a full translation scan is performed. During the translation scan, top three poses are retained according to the RPDock score. Fast Fourier transform is used to accelerate the calculation. The process is repeated until full rotation scan is completed. 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).

Stage 2: scoring by RPRANK

Each presumptive pose generated by RPDock is scored by RPRANK in this stage. RPRANK extracts the residue-base pairs within 10 Å, and then the pairs from decoy complexes are compared with standard pairs that are from native structures. If the RMSD between standard pair and decoy pair is less than 6 Å, the energy of decoy pair will be recorded as same as the standard pair. Finally, the energy of the decoy complex is the sum of the energy of pairs.

PROCEDURE

3dRPC installation

1. To download 3dRPC package, visit the 3dRPC web-page (<http://biophy.hust.edu.cn/3dRPC.html>).
2. Set running environment for 3dRPC. Add the following lines to your "~/.bashrc":

```
export HOME_3dRPC=/home/XXX/3dRPC/,
export X3DNA=${HOME_3dRPC}/ext/X3DNA/,
export PATH=$PATH:${HOME_3dRPC}/ext/fasta/
Type the command in your terminal:
source ~/.bashrc
```

3. Download and install libraries. Three external libraries are required by 3dRPC: FFTW (<http://www.fftw.org/download.html>), BLAS (<http://www.netlib.org/blas/>), and LAPACK (<http://www.netlib.org/lapack/>). The default path of libraries is "\${HOME_3dRPC}/lib/".

[? TROUBLESHOOTING]

4. Install FASTA. FASTA is used for sequence alignment in 3dRPC. The source code of FASTA is located on "\${HOME_3dRPC}/ext/fasta/". Users can execute the following command lines to install FASTA:

```
cd ${HOME_3dRPC}/ext/fasta/,
make
After successful installation, an executable file "fasta35" can be found in "${HOME_3dRPC}/ext/fasta/".
```

5. Install 3dRPC program from the source code. Run the following command lines given below:

```
cd ${HOME_3dRPC}/source,
make
```

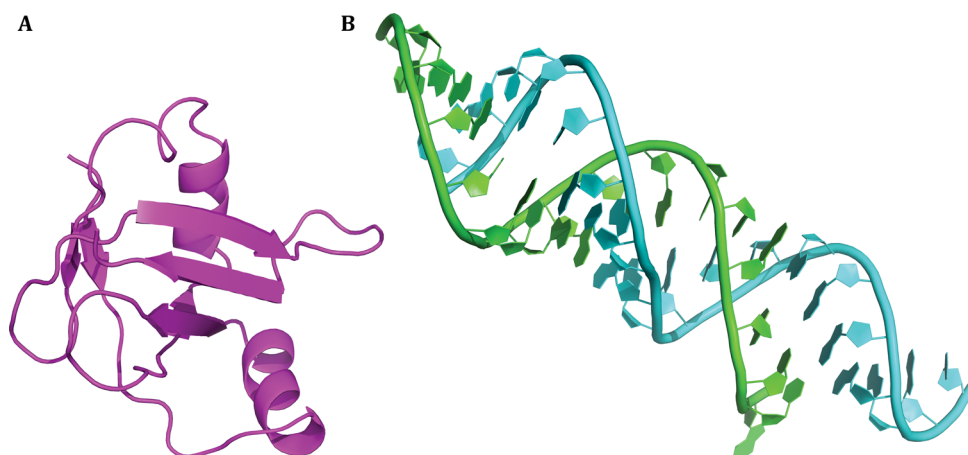
[? TROUBLESHOOTING]

Docking by RPDock

6. Prepare two PDB structures for docking, with one being protein and the other one being RNA. An example is shown in Fig. 1.
7. Prepare the parameter files for RPDock. The parameter files must follow the following formats:

```
RPDock.receptor = 1DFU_r_u.pdb,
RPDock.receptor.chain = V,
RPDock.ligand = 1DFU_l_u.pdb,
RPDock.ligand.chain = CB,
RPDock.outfile = 1DFU.out,
RPDock.grid_step = 1,
```

Fig. 1 An example of docking. The case is obtained from RNA–protein docking benchmark. The PDB code is 1DFU. Unbound protein (A) and unbound RNA (B) are shown in cartoon presentation



RPDock.out_pdb = 10.

The parameter files are further explained in Table 1.

8. Run RPDock by the following command line:

```
"$HOME_3dRPC/source/3dRPC -mode 9 -system 9 -par RPDock.par".
```

"RPDock.par" is the parameter file described previously. After docking is finished, RPDock will generate an output file "1DFU.out" and a number of docked complexes ("complex1.pdb", ..., "complex*.pdb"). An example of the output files is shown below:

G_DATA	13	0	-946.00	13	25	1	3	48.0	0.0	0.0
G_DATA	10	0	-897.00	10	25	5	2	36.0	0.0	0.0
G_DATA	14	0	-858.00	14	25	2	3	48.0	0.0	0.0

Each line represents a docked complex with related information (Table 2). RPDock is a rigid-body docking procedure and the docked complexes depend on the translation vector and the rotation angles (Fig. 2).

9. Generate complexes by the following command line:

```
"$HOME_3dRPC/source/3dRPC -mode 9 -system 8 -par RPDock.par".
```

"RPDock.par" is the same parameter file that is used for docking. Users can change the number of complexes generated.

Scoring with RPRANK

10. Prepare a list of complex structures to be scored by the following format:

complex1.pdb	V	CB
complex2.pdb	V	CB

The first column is the file name of the complex structures, the second column is the chain ID of protein and the last column is the chain ID of RNA.

11. Prepare the parameter file "scoring.par" for scoring:

```
list = list,
out = RMSD.score.
```

12. Run the command to score the complexes in the list:

```
"${HOME_3dRPC}/source/3dRPC -mode 8 -system 9 -par scoring.par".
```

According to the parameter, the output of scoring is saved in the file "RMSD.score". An example of the output is shown below:

complex1.pdb	-93.2882
complex2.pdb	-145.628

The first column is the name of the complex and the second column is the corresponding energy given by RMSD-based score.

Table 1 Explanation of parameter files for RPDock—"RPDock.par"

RPDock.receptor	File name of protein structure
RPDock.receptor.chain	Chain ID of protein
RPDock.ligand	File name of RNA structure
RPDock.ligand.chain	Chain ID of RNA
RPDock.outfile	Output file name of RPDock
RPDock.grid_step	Grid step of RPDock, 1 is recommended
RPDock.out_pdb	Number of complexes generated

Table 2 Explanation of information contained in the output files of RPDock

Column 4	RPDOCK score
Column 6–8	Translation vector
Column 9–11	Rotation angles

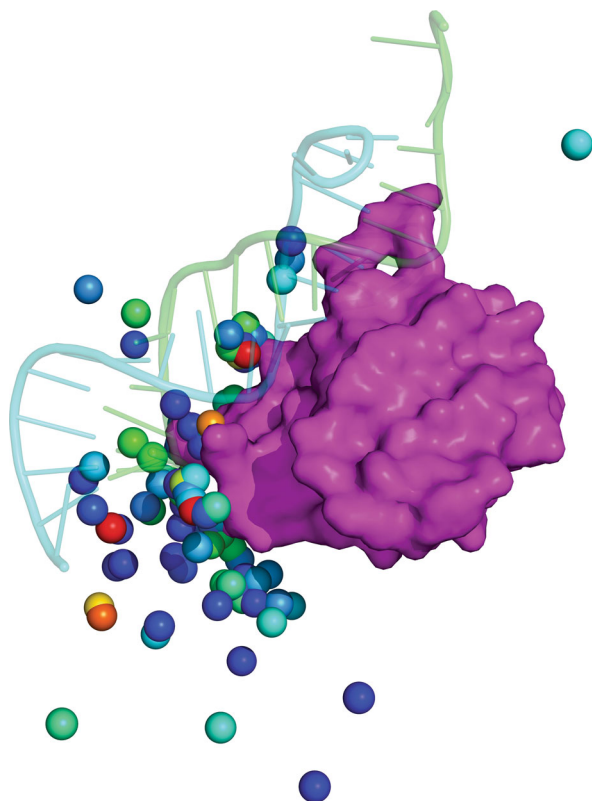


Fig. 2 An example of docking. The native complex (1DFU) is shown in cartoon. The centroids of top 100 poses according to RPDock score are shown in sphere with rainbow color representing RPDock score. The red color represents high RPDock score

Result analysis of RPDock decoy

13. Prepare the parameter file for analysis:

```
RPDock.resfile = 1DFU.out,
RPDock.max_matches = 10,
native.receptor_pdb_filename = 1DFU_r_b.pdb,
native.ligand_pdb_filename = 1DFU_l_b.pdb,
native.receptor.chainid = P,
native.ligand.chainid = MN,
decoy.receptor_pdb_filename = 1DFU_r_u.pdb,
decoy.ligand_pdb_filename = 1DFU_l_u.pdb,
decoy.receptor.chainid = V,
decoy.ligand.chainid = CB,
rmsd.output = 1DFU.rmsd.dat (Table 3).
```

Table 3 Explanation of the parameter files

RPDock.resfile	Output of RPDock
RPDock.max_matches	Number of complexes
native.receptor_pdb_filename	Native protein structure
native.ligand_pdb_filename	Native RNA structure
native.receptor.chainid	Chain ID of native protein
native.ligand.chainid	Chain ID of native RNA
decoy.receptor_pdb_filename	Protein structure used for docking
decoy.ligand_pdb_filename	RNA structure used for docking
decoy.receptor.chainid	Chain ID
decoy.ligand.chainid	Chain ID
rmsd.output	Output file of result analysis

14. Run the following command:

```
"${HOME_3dRPC}/source/3dRPC -mode 2 -system 0 -par rmsd.par".
```

The “rmsd.par” is the parameter file described in step 15. After the calculation is finished, an outfile, named as “1DFU.rmsd.dat” according to the parameter, will be generated. The output files are formatted as following:

#Decoy	R_rmsd	L_rmsd	I_rms	fnat	fnon
1	0.744382	34.1629	14.6322	0	1
2	0.744382	32.8772	14.5631	0.0178571	0.964286

Further explanation of the files is shown in Table 4.

Table 4 Explanation of output files

#Decoy	Decoy number
R_rmsd	RMSD of receptor (protein)
L_rmsd	RMSD of ligand (RNA)
I_rms	Interface RMSD
fnat	Native contact fraction
fnon	Non-native contact fraction

[? TROUBLESHOOTING]

Step 3: How to install BLAS and LAPACK in Mac?

Open the file “BLAS/make.inc” or “LAPACK/make.inc”, find the line that says: “PLAT = _LINUX” and change it to “PLAT = _MACOS”. Type “make” in your terminal to install BLAS and LAPACK.

Step 5: What can I do if I get error while installing 3dRPC?

Make sure that BLAS, LAPACK and FFTW libraries are successfully installed in your system. Open the file “\${HOME_3dRPC}/source/Makefile”, find the line starting

with "LAPACK_LIBS" and "BLAS_LIBS", make sure that the paths of the libraries are correctly assigned.

Acknowledgements This work is supported by the National Natural Science Foundation of China (31570722, 11374113) and the National High Technology Research and Development Program of China (2012AA020402).

Compliance with Ethical Standards

Conflict of interest Yangyu Huang, Haotian Li, and Yi Xiao declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bahadur RP, Zacharias M, Janin J (2008) Dissecting protein-RNA recognition sites. *Nucleic Acids Res* 36:2705–2716
- Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688
- Chen Y, Varani G (2005) Protein families and RNA recognition. *FEBS J* 272:2088–2097
- Chen R, Li L, Weng Z (2003) ZDOCK: an initial-stage protein-docking algorithm. *Proteins* 52:80–87
- Chen Y, Kortemme T, Robertson T, Baker D, Varani G (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
- Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308:1149–1154
- Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125:1731–1737
- Gabb HA, Jackson RM, Sternberg MJ (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 272:106–120
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* 582:1977–1986
- Huang SY, Zou X (2013) A nonredundant structure dataset for benchmarking protein-RNA computational docking. *J Comput Chem* 34:311–318
- Huang SY, 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, Liu S, Guo D, Li L, Xiao Y (2013) A novel protocol for three-dimensional structure prediction of RNA-protein complexes. *Sci Rep* 3:1887
- Jones S, van Heyningen P, Berman HM, Thornton JM (1999) Protein-DNA interactions: a structural analysis. *J Mol Biol* 287:877–896
- Jones S, Daley DTA, Luscombe NM, Berman HM, Thornton JM (2001) Protein-RNA interactions: a structural analysis. *Nucleic Acids Res* 29:943–954
- Kim OTP, Yura K, Go N (2006) Amino acid residue doublet propensity in the protein-RNA interface and its application to RNA interface prediction. *Nucleic Acids Res* 34:6450–6460
- Kozakov D, Brenke R, Comeau SR, Vajda S (2006) PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* 65:392–406
- Kumar M, Gromiha AM, Raghava GPS (2008) Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins* 71:189–194
- Li CH, Cao LB, Su JG, Yang YX, Wang CX (2012) A new residue-nucleotide propensity potential with structural information considered for discriminating protein-RNA docking decoys. *Proteins* 80:14–24
- Perez-Cano L, Fernandez-Recio J (2010) Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. *Proteins* 78:25–35
- Perez-Cano L, Solernou A, Pons C, Fernandez-Recio J (2010) Structural prediction of protein-RNA interaction by computational docking with propensity-based statistical potentials. *Pac Symp Biocomput* 2010:293–301
- Perez-Cano L, Jimenez-Garcia B, Fernandez-Recio J (2012) A protein-RNA docking benchmark (II): extended set from experimental and homology modeling data. *Proteins* 80:1872–1882
- Terribilini M, Lee JH, Yan CH, Jernigan RL, Honavar V, Dobbs D (2006) Prediction of RNA binding sites in proteins from amino acid sequence. *RNA* 12:1450–1462
- Tuszynska I, Bujnicki JM (2011) DARS-RNP and QUASI-RNP: new statistical potentials for protein-RNA docking. *BMC Bioinform* 12:348
- Vakser IA, Aflalo C (1994) Hydrophobic docking: a proposed enhancement to molecular recognition techniques. *Proteins* 20:320–329