Preorientation of protein and RNA just before contacting

Dachuan Guo¹, Shiyong Liu¹, Yangyu Huang and Yi Xiao*

Biomolecular Physics and Modeling Group, Department of Physics, Huazhong University of Science and Technology, Wuhan, 430074 Hubei, China

Communicated by Ramaswamy H. Sarma

(Received 15 May 2012; final version received 27 June 2012)

Protein and RNA molecules interact and form complexes in many biological processes. However, it is still unclear how they can find the correct docking direction before forming complex. In this paper, we study preorientation of RNA and protein separated at a distance of 5–7 Å just before they form contacts and interact with each other only through pure electrostatic interaction when neglecting the influence of other molecules and complicated environment. Since geometric complementarity has no meaning at such a distance, this is not a docking problem and so the conventional docking methods, like FTDock, are inapplicable. However, like the usual docking problem, we need to sample all the positions and orientations of RNA surrounding the protein to find the lowest energy orientations between RNA and protein. Therefore, we propose a long-range electrostatic docking-like method using Fast Fourier Transform-based sampling, LEDock, to study this problem. Our results show that the electrostatically induced orientations between RNA and protein at a distance of 5–7 Å are very different from the random ones and are much closer to those in their native complexes. Meanwhile, electrostatic funnels are found around the RNA-binding sites of the proteins in 62 out of 78 bound protein–RNA complexes. We also tried to use LEDock to find RNA-binding residues and it seems to perform slightly better than BindN Server for 23 unbound protein–RNA complexes.

Keywords: electrostatic interaction; RNA-protein interaction; electrostatic funnels; LEDock

Introduction

Protein–RNA interactions play a critical role in many biological processes. However, the mechanism of protein–RNA complex formation is still unclear. As more and more protein–RNA complex structures have been solved, analysis of protein–RNA interactions based on structures has become dominated (Allers & Shamoo, 2001; Ellis, Broom, & Jones, 2007; Gupta & Gribskov, 2011; Jones, Daley, Luscombe, Berman, & Thornton, 2001; Morozova, Allers, Myers, & Shamoo, 2006). These studies enrich our knowledge of protein–RNA recognition. For example, it was found that positively charged residues (arginine, lysine, and histidine) are more favorable in protein–RNA binding interface than negatively charged residues (glutamic acid and aspartic acid). This propensity of charged residues has already been taken as an important component of the scoring function (Perez-Cano, Solernou, Pons, & Fernandez-Recio, 2010) and protein–RNA binding sites prediction methods (Bahadur, Zacharias, & Janin, 2008; Chen & Lim, 2008; Kumar, Gromiha, & Raghava, 2008; Liu, Wu, Wang, Zhang, & Chen, 2010; Zhao, Yang, & Zhou, 2011).

On the other hand, since electrostatic interaction is the only force at large separation (e.g. atom–atom distance > 5 Å) if neglecting the influence of other molecules and ions, it may play a key role in the process of protein–protein/RNA recognition before they become an encounter complex (Drozdov-Tikhomirov et al., 2003; Drozdov-Tikhomirov, Linde, Poroikov, Alexandrov, & Skurida, 2001; Kovalev, Drozdov-Tikhomirov, Poroikov, & Alexandrov, 2004; Qin & Zhou, 2009; Schreiber & Fersht, 1996; Schreiber, Haran, & Zhou, 2009; Schreiber, Shaul, & Gottschalk, 2006; Sinha & Smith-Gill, 2002). For examples, previous studies showed that the measured rates of protein–protein complex formation are much higher than those predicted by Brown collision (Alsallaq & Zhou, 2008; Drozdov-Tikhomirov et al., 2001, 2003; Kovalev et al., 2004; Schreiber et al., 2009; Schreiber & Fersht, 1996). The barnase–barstar complex is a classic example (Camacho, Weng, Vajda, & DeLisi, 1999; Janin, 2008; Liu, Wu, Wang, Zhang, & Chen, 2010; Zhao, Yang, & Zhou, 2011).

*Corresponding author. Email: yxiao@mail.hust.edu.cn
¹These authors contributed equally to this work.

Copyright © 2012 Taylor & Francis
Neumann & Gottschalk, 2009; Schreiber et al., 2009; Spaar, Dammer, Gabdoulline, Wade, & Helms, 2006) and its association rate constant is about $10^3$ times higher than diffusion-controlled association rate constant (Gabdoulline & Wade, 1997). This may be explained by long-range electrostatic interactions. Camacho et al. (1999) calculated the electrostatic energy at 5 Å surface-to-surface separations for 12 protein–protein complexes and found that electrostatic interaction exerts strong effect in preorientation of subunits in the formation of the complexes with strong opposite charges. But their sampling was done only on a particular plane and was insufficient. Kovalev et al. (2004) calculated the electrostatic energy on a homodimer (PDBid:1RGE) at 100 Å separation of two centers of subunits and also found that electrostatic interactions can achieve the effective preorientation of subunits, although they studied only one protein–protein complex. We hope to study this problem for protein–RNA complexes in our paper.

On account of the negative charges from phosphate groups, it is generally believed that electrostatic interactions are much more dominant in protein–RNA recognition before they form contacts with each other than in protein–protein complex formation. Since full spatial sampling of electrostatic interactions at a distance is similar to what docking methods do, it is better to use the efficient Fast Fourier Transform-based (FFT) sampling algorithm, which has been used in protein–protein docking programs, such as GRAMM (Vakser, 1997), FTDock (Gabb et al., 1997), ZDOCK (Chen & Weng, 2002; Chen, Li, & Weng, 2003), ASPDock (Li, Guo, Huang, Liu, & Xiao, 2011), and SDock (Zhang & Lai, 2011). However, these FFT-based sampling methods do not consider a sampling decoy whose subunits are separated 5 Å away, as shown in Figure 1. In fact, some degree of surface overlap is tolerated to account for conformational change when a preequilibrium encounter complex relaxes to the final specific complex. Therefore, in this paper we propose a long-range electrostatic docking-like method (LEDock) based on FFT algorithm to explore long-range recognitions between the subunits of 78 bound protein–RNA complexes and the electrostatic funnel of the protein–RNA interaction. We study the role of electrostatic interaction in pre-orientation of RNA and protein separated at a distance of 5–7 Å, just before they form contacts and interact with each other only through pure electrostatic interaction when neglecting the influence of other molecules and complicated environment.

### Materials and methods

#### Sampling

Since FFT algorithm was first applied in geometric surface recognition method by Katchalski-Katzir et al. (1992), it has been used in many docking protocol (Chen et al., 2003; Chen & Weng, 2002; Gabb et al., 1997; Li et al., 2011; Vakser, 1997; Zhang & Lai, 2011) due to its high efficiency in full space sampling. In our calculations, we also use the FFT-based sampling method to perform the full space sampling. In order to further enhance the efficiency of sampling, we also use the Lattman’s method (Gabb et al., 1997; Lattman, 1972) to remove the redundant angles. The grid step and angle step are set as 2.0 Å and 15°, respectively. Different grid sizes for different molecular sizes are used to make sure that sampling structures can be accommodated. The half of grid size is the sum of protein radius and RNA diameter plus 10 Å. The size of the grids ranged from 100 to 256.

Based on FFT-based sampling algorithm, we construct LEDock algorithm (Figure 2), which can be divided into two parts: geometric constraint (GC) and electrostatic interaction.

#### Geometric constraint

This is to ensure that the distance between the closest atoms of protein and RNA is at 5–7 Å. First, the protein (denoted as P) and the RNA (denoted as R) are mapping onto a three-dimensional grid of $N \times N \times N$ which is assigned as follows:

$$GC_p(l,m,n) = \begin{cases} -1: & \text{surface of } P \\ 100: & \text{core of } P \\ 0: & \text{others} \end{cases}$$

$$GC_R(l,m,n) = \begin{cases} 1: & \text{inside } R \\ 0: & \text{others} \end{cases}$$

Here, $(l,m,n)$ are the coordinates of the grid $(l,m,n = 1, \ldots, N)$. $P$ and $R$ are treated differently. First, any grid node within 7 Å from any atom in $P$ is assigned $-1$. Second, we assign 100 to all grid nodes within 5 Å from
any atom in P, which are considered as the core of P. So, the remaining grid nodes with \(-1\) are considered as surface of P. Any grid node within 1.8 Å from any atom in R is considered as the inside of R. Others are assigned as 0. The value of the core of P is set much larger than others just to make sure that the penalty is large enough when two subunits are too close to each other. We also changed the value of the core of P to 50 and 200, but the results have no significant changes.

The score of GC can be calculated using the correlation function of GC_P and GC_R:

\[
S_{\text{GC}}(a, b, c) = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} GC_P(l, m, n) \\
\times GC_R(l + a, m + b, n + c)
\]

where \(N^3\) is the number of grid points and \(a, b,\) and \(c\) are the translation grid steps of RNA R relative to protein P.

**Electrostatic interactions**

The electrostatic potential between RNA atom \(i\) and protein is given by

\[ V_i = \varphi_i q_i. \]

Here, we adopted the approach by Gabb et al. (1997), except \(q_i\) is the charge on RNA atom \(i\) which is adopted from the partial charge of atom in the AMBER parm03 force field (Meyer et al., 2010) in combination with the pdb2pqr program (Dolinsky, Nielsen, McCammon, & Baker, 2004). And \(\varphi_i\) is the protein electric field strength calculated by the Adaptive Poisson Boltzmann Solver (APBS) (Baker, Sept, Joseph, Holst, & McCammon, 2001). So, the electrostatic energy of each grid node \(S_{\text{LE}}(a, b, c)\) can be expressed as follows:

\[
S_{\text{LE}}(a, b, c) = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \varphi_p(l, m, n) \\
\times q_R(l + a, m + b, n + c)
\]

After each rotation, the electrostatic energy \(S_{\text{LE}}(a, b, c)\) is saved only if the corresponding \(S_{\text{GC}}(a, b, c)\) is less than zero and the \(S_{\text{LE}}(a, b, c)\) is lower than the previous one.

**Evaluation of the closeness between sampling and native orientations**

We evaluate the closeness of the orientations between RNA and protein subunits of sampling structures to that in their native structures as follows: the geometric center of the protein subunit in native protein–RNA structure is considered as the origin of coordinate system. The \(y\)-axis is along the line connecting the geometric centers of the
Table 1. Seventy-eight bound complexes in our protein–RNA data-set.

| Pdb id | Protein chain | RNA chain | N_{\text{aa}}^{\text{a}} | N_{\text{nu}}^{\text{b}} | \alpha^{\text{c}} (°) | RNA type^{d} |
|--------|---------------|-----------|-----------------|-----------------|------------------|--------------|
| 1A34   | A             | BC        | 159             | 20              | 53.2             | Paired RNA   |
| 1AQ3   | A             | R         | 129             | 19              | 22.1             | Paired RNA   |
| 1ASY   | A             | R         | 490             | 75              | 125.6            | tRNA         |
| 1AV6   | A             | B         | 295             | 6               | 140.2            | Unpaired RNA |
| 1B23   | P             | R         | 405             | 74              | 97.7             | tRNA         |
| 1B7F   | A             | P         | 168             | 12              | 19.1             | Unpaired RNA |
| 1CVJ   | A             | M         | 190             | 11              | 56.1             | Unpaired RNA |
| 1E7K   | A             | C         | 128             | 22              | 18.8             | Paired RNA   |
| 1FFY   | A             | T         | 917             | 75              | 84.7             | tRNA         |
| 1G59   | A             | B         | 468             | 75              | 51.6             | tRNA         |
| 1GAX   | A             | C         | 862             | 75              | 20.8             | tRNA         |
| 1H2D   | A             | R         | 182             | 3               | 37.1             | Unpaired RNA |
| 1H3E   | A             | B         | 432             | 86              | 34.3             | tRNA         |
| 1L2    | A             | C         | 590             | 75              | 61.3             | tRNA         |
| 1JBR   | B             | D         | 149             | 31              | 96.8             | Paired RNA   |
| 1JJD   | A             | B         | 128             | 29              | 14.3             | Paired RNA   |
| 1K8W   | A             | B         | 327             | 22              | 7.8              | Paired RNA   |
| 1LNG   | A             | B         | 87              | 97              | 57.0             | Paired RNA   |
| 1M5K   | C             | AB        | 100             | 113             | 33.3             | Paired RNA   |
| 1N1H   | A             | B         | 1267            | 6               | 101.9            | Unpaired RNA |
| 1NB7   | A             | E         | 570             | 4               | 129.6            | Unpaired RNA |
| 1OOA   | A             | C         | 326             | 29              | 7.5              | Paired RNA   |
| 1R9F   | A             | BC        | 136             | 42              | 43.2             | Paired RNA   |
| 1S03   | H             | A         | 129             | 47              | 43.7             | Paired RNA   |
| 1S13   | A             | B         | 149             | 9               | 8.6              | Paired RNA   |
| 1TFW   | A             | JG        | 437             | 26              | 41.2             | Paired RNA   |
| 1U0B   | A             | B         | 461             | 74              | 39.5             | tRNA         |
| 1UV1   | A             | D         | 664             | 6               | 24.3             | Unpaired RNA |
| 1VFG   | A             | C         | 390             | 75              | 136.4            | Paired RNA   |
| 1WMQ   | A             | C         | 147             | 7               | 121.5            | Unpaired RNA |
| 1Y39   | A             | C         | 76              | 58              | 10.2             | Paired RNA   |
| 1YTU   | A             | CD        | 427             | 10              | 33.7             | Paired RNA   |
| 1YVP   | A             | CD        | 538             | 20              | 55.1             | Paired RNA   |
| 1YYW   | A             | EFGH      | 221             | 48              | 1.4              | Paired RNA   |
| 1ZH5   | A             | CD        | 195             | 18              | 8.3              | Paired RNA   |
| 1ZHO   | A             | B         | 228             | 38              | 13.8             | Paired RNA   |
| 2A8V   | B             | E         | 118             | 6               | 70.6             | Unpaired RNA |
| 2ANN   | A             | B         | 178             | 25              | 32.9             | Paired RNA   |
| 2ASB   | A             | B         | 251             | 11              | 56.1             | Unpaired RNA |
| 2AZX   | A             | C         | 477             | 75              | 73.9             | tRNA         |
| 2B3J   | A             | E         | 159             | 16              | 33.3             | Paired RNA   |
| 2BH2   | A             | C         | 433             | 37              | 12.1             | Paired RNA   |
| 2BTE   | A             | B         | 878             | 83              | 12.2             | tRNA         |
| 2BX2   | L             | R         | 517             | 15              | 22.6             | Unpaired RNA |
| 2CSX   | A             | C         | 497             | 75              | 58.3             | tRNA         |
| 2DLC   | X             | Y         | 394             | 76              | 30.7             | tRNA         |
| 2DU3   | A             | D         | 534             | 71              | 46.8             | tRNA         |
| 2E9R   | X             | AB        | 476             | 15              | 55.9             | Paired RNA   |
| 2F8K   | A             | B         | 88              | 16              | 17.6             | Paired RNA   |
| 2FK6   | A             | R         | 320             | 79              | 16.8             | tRNA         |
| 2FM1   | A             | C         | 314             | 77              | 51.4             | tRNA         |
| 2G4B   | A             | B         | 172             | 7               | 92.8             | Unpaired RNA |
| 2GX8   | A             | EF        | 66              | 14              | 12.6             | Paired RNA   |
| 2IPY   | A             | C         | 888             | 30              | 8.5              | Paired RNA   |
| 2IYF1  | A             | B         | 664             | 13              | 65.9             | Unpaired RNA |
| 2JLU   | A             | C         | 451             | 12              | 14.3             | Unpaired RNA |
| 2PJP   | A             | B         | 121             | 23              | 6.7              | Paired RNA   |
| 2PLY   | A             | C         | 258             | 23              | 76.9             | Paired RNA   |
| 2R7R   | A             | X         | 1095            | 7               | 82.6             | Unpaired RNA |
| 2RD2   | A             | B         | 556             | 75              | 40.6             | tRNA         |

(Continued)
protein and its partner RNA in the native structure and z-axis is chosen arbitrarily. \( \theta \) angle is calculated from the projection of the geometric center of the sampling RNA structure with respect to the \( z \)-axis and \( \phi \) is calculated from the projection of the geometric center of the sampling RNA structure onto the \( xy \) plane with respect to the \( y \)-axis. Then, the protein subunit in the sampling protein–RNA structure is overlapped with that in the native structure and thus, the angles \( \theta \) and \( \phi \) of RNA in the sampling structure can be used to calculate the angle \( \alpha \) from the projection of the geometric center of the sampling RNA structure with respect to the \( y \)-axis as follows:

Table 2. RNA-binding residues prediction test set of 23 unbound complexes in our protein–RNA data-set.

| Protein PDB | RNA PDB | Complex PDB |
|-------------|---------|-------------|
| 2V3C A      | M       | 87  96      | 51.2 | Paired RNA |
| 2Z10 A      | CD      | 75  42      | 33.2 | Paired RNA |
| 2ZM5 A      | C       | 316 76      | 51.8 | tRNA      |
| 2Z4M A      | B       | 336 88      | 50.7 | tRNA      |
| 3AEV B      | C       | 219 11      | 19.5 | Unpaired RNA |
| 3BSN A      | PT      | 510 17      | 136.5 | Paired RNA |
| 3D2S A      | E       | 70  6       | 166.9 | Unpaired RNA |
| 3EPH A      | E       | 409 69      | 46.8 | tRNA      |
| 3EQT A      | CD      | 145 16      | 6.7  | Paired RNA |
| 3FTF A      | CD      | 249 44      | 32.5 | Paired RNA |
| 3G0H A      | E       | 424 7       | 6.0  | Unpaired RNA |
| 3G9Y A      | C       | 33  6       | 125.0 | Unpaired RNA |
| 3GPQ A      | E       | 168 3       | 42.5 | Unpaired RNA |
| 3I5X A      | B       | 563 10      | 79.2 | Unpaired RNA |
| 3IEV A      | D       | 308 12      | 47.8 | Unpaired RNA |
| 3K49 A      | B       | 369 10      | 50.5 | Unpaired RNA |
| 3K5Q A      | B       | 412 9       | 124.8 | Unpaired RNA |
| 3KS8 A      | EF      | 184 36      | 20.8 | Paired RNA |

| Protein PDB | RNA PDB | Complex PDB |
|-------------|---------|-------------|
| 2V3C A      | M       | 87  96      | 51.2 | Paired RNA |
| 2Z10 A      | CD      | 75  42      | 33.2 | Paired RNA |
| 2ZM5 A      | C       | 316 76      | 51.8 | tRNA      |
| 2Z4M A      | B       | 336 88      | 50.7 | tRNA      |
| 3AEV B      | C       | 219 11      | 19.5 | Unpaired RNA |
| 3BSN A      | PT      | 510 17      | 136.5 | Paired RNA |
| 3D2S A      | E       | 70  6       | 166.9 | Unpaired RNA |
| 3EPH A      | E       | 409 69      | 46.8 | tRNA      |
| 3EQT A      | CD      | 145 16      | 6.7  | Paired RNA |
| 3FTF A      | CD      | 249 44      | 32.5 | Paired RNA |
| 3G0H A      | E       | 424 7       | 6.0  | Unpaired RNA |
| 3G9Y A      | C       | 33  6       | 125.0 | Unpaired RNA |
| 3GPQ A      | E       | 168 3       | 42.5 | Unpaired RNA |
| 3I5X A      | B       | 563 10      | 79.2 | Unpaired RNA |
| 3IEV A      | D       | 308 12      | 47.8 | Unpaired RNA |
| 3K49 A      | B       | 369 10      | 50.5 | Unpaired RNA |
| 3K5Q A      | B       | 412 9       | 124.8 | Unpaired RNA |
| 3KS8 A      | EF      | 184 36      | 20.8 | Paired RNA |

| Protein PDB | RNA PDB | Complex PDB |
|-------------|---------|-------------|
| 2V3C A      | M       | 87  96      | 51.2 | Paired RNA |
| 2Z10 A      | CD      | 75  42      | 33.2 | Paired RNA |
| 2ZM5 A      | C       | 316 76      | 51.8 | tRNA      |
| 2Z4M A      | B       | 336 88      | 50.7 | tRNA      |
| 3AEV B      | C       | 219 11      | 19.5 | Unpaired RNA |
| 3BSN A      | PT      | 510 17      | 136.5 | Paired RNA |
| 3D2S A      | E       | 70  6       | 166.9 | Unpaired RNA |
| 3EPH A      | E       | 409 69      | 46.8 | tRNA      |
| 3EQT A      | CD      | 145 16      | 6.7  | Paired RNA |
| 3FTF A      | CD      | 249 44      | 32.5 | Paired RNA |
| 3G0H A      | E       | 424 7       | 6.0  | Unpaired RNA |
| 3G9Y A      | C       | 33  6       | 125.0 | Unpaired RNA |
| 3GPQ A      | E       | 168 3       | 42.5 | Unpaired RNA |
| 3I5X A      | B       | 563 10      | 79.2 | Unpaired RNA |
| 3IEV A      | D       | 308 12      | 47.8 | Unpaired RNA |
| 3K49 A      | B       | 369 10      | 50.5 | Unpaired RNA |
| 3K5Q A      | B       | 412 9       | 124.8 | Unpaired RNA |
| 3KS8 A      | EF      | 184 36      | 20.8 | Paired RNA |

*aSequence length of protein chain. 
*bSequence length of RNA chain. 
*cCone angle of the sampling structure with the lowest electrostatic energy. 
*dThree categories of complexes according to the type of RNA.

Table 1. (Continued).

| Pdb id | Protein chain | RNA chain | \(N_{\text{aa}}\) | \(N_{\text{nuc}}\) | \(\alpha\) (°) | RNA type |
|--------|---------------|------------|-------------------|-------------------|----------------|----------|
| 2V3C   | A             | M          | 87                | 96                | 51.2           | Paired RNA |
| 2Z10   | A             | CD         | 75                | 42                | 33.2           | Paired RNA |
| 2ZM5   | A             | C          | 316               | 76                | 51.8           | tRNA      |
| 2Z4M   | A             | B          | 336               | 88                | 50.7           | tRNA      |
| 3AEV   | B             | C          | 219               | 11                | 19.5           | Unpaired RNA |
| 3BSN   | A             | PT         | 510               | 17                | 136.5          | Paired RNA |
| 3D2S   | A             | E          | 70                | 6                 | 166.9          | Unpaired RNA |
| 3EPH   | A             | E          | 409               | 69                | 46.8           | tRNA      |
| 3EQT   | A             | CD         | 145               | 16                | 6.7            | Paired RNA |
| 3FTF   | A             | CD         | 249               | 44                | 32.5           | Paired RNA |
| 3G0H   | A             | E          | 424               | 7                 | 6.0            | Unpaired RNA |
| 3G9Y   | A             | C          | 33                | 6                 | 125.0          | Unpaired RNA |
| 3GPQ   | A             | E          | 168               | 3                 | 42.5           | Unpaired RNA |
| 3I5X   | A             | B          | 563               | 10                | 79.2           | Unpaired RNA |
| 3IEV   | A             | D          | 308               | 12                | 47.8           | Unpaired RNA |
| 3K49   | A             | B          | 369               | 10                | 50.5           | Unpaired RNA |
| 3K5Q   | A             | B          | 412               | 9                 | 124.8          | Unpaired RNA |
| 3KS8   | A             | EF         | 184               | 36                | 20.8           | Paired RNA |

*aSequence length of protein chain. 
*bSequence length of RNA chain. 
\(\alpha\)Cone angle of the sampling structure with the lowest electrostatic energy. 
*dThree categories of complexes according to the type of RNA.
\[ \cos \alpha = \cos(90^\circ - \theta) \cdot \cos \phi. \]

We use this angle \( \alpha \) to indicate the closeness of the orientation between RNA and protein subunits of sampling structures to that in their native structures. Since two monomers of a sampling protein–RNA decoy are separated 5 Å away, normal evaluation criterions such as root-mean-square deviation (RMSD) and fraction of native contacts cannot be used.

To see how many sampling RNA structures have orientations less than a particular value of \( \alpha \) relative to \( y \)-axis, we define the closeness rate of LEDock sampling as follows:

\[
\text{Closeness rate of LEDock}(\alpha) = \frac{N_{\alpha}}{N_{\text{max}}},
\]

where \( N_{\alpha} \) is the number of structures less than a particular value \( \alpha \). \( N_{\text{max}} \) is the number of total structures in data-set and equals to 78 in this study. For comparison, we also calculated the closeness rate of random sampling without electrostatic interactions by the formula given below:

\[
\text{Random closeness rate}(\alpha) = \frac{1 - \cos \alpha}{2}
\]

**Representation of electrostatic energy landscape in spherical coordinates**

For each protein–RNA complex, the sampling structures of the top 5000 lowest electrostatic energies are retained to plot the distribution of their poses by using the similar method in the paper by Fernandez-Recio, Totrov, and Abagyan (2004) as shown in Figures 5 and 6. The angles \( \theta \) and \( \phi \) are used to represent the sampling poses in a two-dimensional sinusoidal projection. The poses are colored from black to yellow according to the electrostatic energies from lower to higher. The 60° and 90° lines of cone angle \( \alpha \) are shown in figures correspondingly. The dot with the lowest electrostatic energy is marked as a yellow cross.

**RNA-binding residues prediction and performance evaluation**

A residue is defined as a RNA-binding residue if the distance between any heavy atom of the protein and any heavy atom of the RNA is less than 4.5 Å. For the prediction of RNA-binding residues, each residue gets a score, which is calculated as follows:

\[
\text{Score}(i) = \sum_{i=1}^{100} \sum_{j=1}^{N} -1/[(\min(d_{ik}^j))^6],
\]

where \( \min(d_{ik}^j) \) is the closest distance between heavy atom \( j \) of residue \( i \) and any heavy atom \( k \) of RNA. \( N \) is the number of heavy atoms in residue \( i \). Here, we used 100 sampling structures of lowest electrostatic energies. And then the scores of residues were normalized as follows:

\[
\text{Score}_{\text{final}}(i) = 1 - (\text{Score}(i) - \text{Score}_{\text{min}})/\text{Score}_{\text{max}} - \text{Score}_{\text{min}}
\]

where \( \text{Score}_{\text{min}} \) and \( \text{Score}_{\text{max}} \) are minimum and maximum values of scores, respectively. Finally, we chose top 30 residues with highest score values as the RNA-binding site residues.

Some widely used measures are adopted in this study, such as sensitivity \( \text{SN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \) and specificity \( \text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}} \). We calculate the Matthews correlation coefficient (MCC) as follows:

\[
\text{MCC} = \frac{(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FN})(\text{TP} + \text{FP})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}.
\]

Here TP, TN, FP, and FN represent the true positives, true negatives, false positives, and false negatives, respectively.

**Data-set**

It is known that the tertiary structures of RNA molecules are the basis to understand their dynamics and functions (Gong & Xiao, 2010; Gong, Zhao, Chen, & Xiao, 2011; Zhang, Bian, Lin, & Wang, 2012; Zhang et al., 2009; Zhang, Zhang, & Wang, 2011; Zhao, Gong, & Xiao, 2011). To better study the association process, the structures of both complex and corresponding unbound subunits should be used, but at present few of them are found simultaneously in Protein Data Bank (PDB) (Bernman et al., 2006). So we use protein–RNA complex structures and their bound subunits here for elaborate analysis and a relatively small unbound data-set for RNA-binding residues prediction. The native protein–RNA complex structures are extracted from the PDB with resolution better than 3.0 Å and sequence identity less than 30%. Missed heavy atoms and hydrogen atoms are added through tleap of AMBER Version 11 program (Case et al., 2010). Finally, we construct a data-set with 78 bound protein–RNA complex structures and they are listed in Table 1. Twenty-three unbound protein–RNA structures listed in Table 2 are selected from data-set.
adopted in the paper by Li, Cao, Su, Yang, and Wang (2012). The selection criteria are the integrality and proper size of the protein and RNA subunits. As we submitted our paper, two data-sets of protein–RNA are published (Barik, Nithin, Manasa, & Bahadur, 2012; Perez-Cano, Jimenez-Garcia, & Fernandez-Recio, 2012) recently. The sizes of the new data-sets are 45 and 71, respectively. And the latter is almost the same size as ours.

Results

Comparing electrostatic and random preorientations

Preorientation of subunits is checked before they become an encounter complex by closeness rates of LEDock and random samplings. Closeness rate relative to the cone angle used here is borrowed from docking problem and just describe how many RNA molecules have their orientation smaller than a particular angle. The higher the closeness rate, the more subunits are closer to their native binding orientations.

As described in Figure 3, the closeness rate of LEDock sampling is much higher than random sampling, especially when cone angle $\alpha$ is smaller (<90°). The largest difference between them is 48.1% when the cone angle $\alpha = 60^\circ$. Preorientation of subunits ($\alpha < 60^\circ$) is found in 57 out of 78 protein–RNA complex subunits (73.1%). This indicates that the orientations of RNA and protein molecules with lowest electrostatic energies are much closer to those in their native complexes than random orientations.

Distributions of docking sampling poses for three types of protein–RNA complexes

All the protein–RNA complexes can be divided into three categories according to the type of RNA: paired RNA type, unpaired RNA type, and tRNA type. The complex consisted of protein and tRNA is considered as tRNA type. If no base pairing exists in RNA subunit, the complex is defined as unpaired RNA type. The others are considered as paired RNA. The closeness rates of these three types of protein–RNA complexes are shown in Figure 4. Protein–paired RNA complexes have a highest closeness rate while protein–unpaired RNA complexes have a lowest closeness rate. Furthermore, for almost all the cone angles, the closeness rates of all three types of protein–RNA complexes are significantly higher than that of random samplings. These suggest that the preorientation of the paired RNA type of complexes may be mainly determined by electrostatic interaction.

Figures 5 and 6 show some typical distributions of the positions of the sampling structures of the first 5000 lowest electrostatic energies for the two types of the protein–RNA complexes. Since tRNA type is just a special case of paired RNA type, their distributions of sampling are similar.

Complexes of paired RNA type

Figure 5 shows the LEDock sampling distributions of two protein–RNA complexes of paired RNA type. Figure 5(B) shows a spherical coordinates projection of the positions of geometric centers of the RNA of the first 5000 lowest electrostatic energies for human LGP2 C-terminal domain and a double-stranded RNA complex (PDB 3EQT) (see ‘Materials and methods’ section for details). The yellow cross represents the position of the geometric center of the RNA with the lowest electrostatic energy. In this case, the angle $\alpha$ of the structure with the lowest electrostatic energy is $6.7^\circ$, which means the orientation of the RNA relative to protein is nearly the
same as that in the native complex structure (Figure 5 (A)).

Figure 5(C) and (D) show the same representation for elongation factor SelB and SECIS RNA complex (PDB 2PLY). The angle $\alpha$ of the structure with the lowest electrostatic energy is 76.9° (>60°), i.e. the orientation of the RNA is very different from that in native complex structure. However, the position of RNA is still at the same side of the binding interface because the angle $\alpha$ is smaller than 90°.

Complex of unpaired RNA type

Figure 6(A) and (B) describes the complex of cap-modifying and 5’-capped mRNA (PDB 1AV6). The cone angle $\alpha$ of the RNA with the lowest electrostatic energy is 140.2°, and the sampling and bound RNA appear having a mirror symmetry as shown in Figure 6(A). Even though most of the sampling structures deviate from the center, the top 5000 sampling structures still include some spots around the center as shown in Figure 6(B).

Figure 6(C)–(F) shows the FBF–2/FBE complex (PDB 3K5Q) and complex of HC-J4 polymerase and short RNA template strand (PDB 1NB7), respectively. In these cases, it seems that pure electrostatic interaction cannot lead to a proper orientation of the protein and RNA subunits on account of interaction between bases of RNA and residues of protein (Figure 6(C) and (D)) and RNA buried in protein (Figure 6(E) and (F)). Stacking interactions between bases and amino acids are common and strong short-range interactions between protein and RNA (Allers & Shamoo, 2001; Koh et al., 2011; Morozova et al., 2006), which cannot be identified using electrostatic interaction. Embedding is also a problem, which is hardly sampled by electrostatic interaction. In most protein–RNA complex of this type, the protein presents in its closed state, in which the binding pocket of protein shrinks to enclose the RNA. Long-range interactions occur when the protein stays in its open state, so it is hard to find the correct orientation of this type of protein–RNA complex using bound structures.
Electrostatic energy funnel

Experimental evidence suggests that the association rate constant is usually higher than diffusion-controlled association when opposite charges are distributed on the binding partners, such as barnase and barstar (Schreiber & Fersht, 1996) and DNase E9 and Im9 (Wallis, Moore, James, & Kleanthous, 1995). Electrostatic energy funnel at the protein surface could more easily capture the partner (RNA), thereby enhancing the binding efficiency.

Here, the distribution of electrostatic energies of the full spatial sampling structures relative to the cone angle $\alpha$ is shown in Figure 7 (those of all 78 protein–RNA complexes are given in Supporting Information). Most of the complexes (62 out of 78 protein–RNA complexes) have electrostatic funnels around the RNA-binding sites of proteins. The cone angle with the minimal average energy is less than 10° in 49 complexes (62.8%) and 60° in 62 complexes (79.5%) out of all 78 protein–RNA complexes.

RNA-binding site prediction

To evaluate the ability of our program in RNA–protein recognition, we also predicted RNA-binding residues using 100 sampling structures of the lowest electrostatic energies (see ‘Materials and methods’ section for details). The performance of the methodology is evaluated on 23 unbound protein–RNA complexes and compared with the BindN server (Wang & Brown, 2006). Figure 8 shows a representative example of the binding site prediction. The predicted binding sites for all cases are listed in Table S1. The ROC curve is shown in Figure 9 and the area under the ROC curve (AUC) is 0.80. We achieved 48% in sensitivity, 88% in specificity, and 0.33 for the MCC value and the BindN server (setting the expected specificity to 80%) achieved 35% in sensitivity, 88% in specificity, and 0.21 for the MCC value. Since our method uses only electrostatic energy without any training method or adjustable parameter, it may be more reliable and physically meaningful than BindN.
Protein–RNA association at different salt concentrations

The results above indicate that electrostatic interaction would be one of the key factors in preorientation of RNA and protein before they form complex structure. To

Figure 7. Distribution of electrostatic energy (in kcal/mol) of sampling structures relative to the cone angle $\alpha$. All sampling structures of each protein–RNA system are shown in corresponding figures. (A)–(D) are four examples of representative. The black line in the middle is the average electrostatic energy of each $3^\circ$ and the dot with the lowest energy is circled by a green ring.

Figure 8. Representative prediction result (bound complex PDBid:2PJP). Red, native binding site; blue, predicted binding site; and yellow, overlapped regions between native and predicted binding sites. RNA molecule is shown in purple.

Figure 9. Roc curve for the prediction of RNA-binding sites including 23 unbound protein–RNA complexes.

Protein–RNA association at different salt concentrations

The results above indicate that electrostatic interaction would be one of the key factors in preorientation of RNA and protein before they form complex structure. To
further show this, we use LEDock to study protein–RNA association in vitro at different salt concentrations. Here, we consider a protein–RNA complex composed of Carnation Italian ringspot virus (CIRV) p19-H and siRNA (PDBid:1RPU) (Vargason, Szittya, Burgyan, & Hall, 2003). The experimental data of the dissociation constants for CIRV p19-H:siRNA binding under different salt concentrations from 0.1 to 1.1 M are shown in Table 3, which are extracted from the reference (Koukiekolo, Sagan, & Pezacki, 2007). We use APBS software to calculate the lowest electrostatic energies of the sampling structure under different ion concentrations and the results are listed in Table 3. We find that the lowest electrostatic energy and the dissociation constant change consistently under different ionic strengths, i.e. the lower the ion concentration, the smaller the dissociation constant and the lower the electrostatic energy. Sampling structures of the lowest electrostatic energy under different ionic strengths are shown in Figure 10. RNA structures colored in red, blue, and yellow represent sampling structures from lower ion concentration to higher one. We find that the lower the ion concentration, the closer to the native structure is the sampling RNA structure with the lowest electrostatic energy. These results indeed show that electrostatic interaction is important to the pre-orientation of CIRV p19-H and siRNA. Under higher ion concentration the strength of electrostatic interaction between RNA and protein is reduced and RNA and protein cannot find their correct orientation. It also shows that LEDock can find proper orientation between the RNA and protein subunits before they form complex.

### Table 3. Dissociation constants and electrostatic energy of the top sampling structure for CIRV p19-H:siRNA binding under a range of ion concentrations.

| pH  | Ionic strength (M) | $K_d$ (nM)$^a$ | Electrostatic energy (kcal/mol)$^b$ |
|-----|-------------------|----------------|-----------------------------------|
| 7.2 | 0.1 (NaCl)        | 513            | −17.0                             |
| 7.2 | 0.2 (NaCl)        | 600            | −9.4                              |
| 7.2 | 0.35 (NaCl)       | 1100           | −5.2                              |
| 7.2 | 0.6 (NaCl)        | ~3100          | −3.0                              |
| 7.2 | 0.85 (NaCl)       | ~5800          | −2.1                              |
| 7.2 | 1.1 (NaCl)        | ~7200          | −1.6                              |

$^aK_d$ is the dissociation constant measured by fluorescence in CIRV-p19 array binding assays. $^b$The electrostatic energy is the lowest electrostatic energy of the sampling structures under each ion concentration.

**Conclusions**

Here, we analyzed the effect of electrostatic interaction on protein–RNA orientation using an FFT-based sampling method. The electrostatic energy algorithm adopted here is Coulomb’s formula and no any other artificial parameter is used, which make the results more reasonable. A large data-set (78 protein–RNA complexes) is also used to get a more convincing conclusion.

The result shows that electrostatic interaction may be one of the important factors in preorientation of protein/RNA subunits at a distance of 5–7 Å. The orientations of RNA and protein molecules with lowest electrostatic energies are very different from the random ones and are much closer to those in their native complexes. Electrostatic funnels are found around the protein at 5–7 Å separation, and most of them are located around RNA-binding sites. These funnels could make proteins to capture RNAs much easily. We also used our method to predict RNA-binding residues and it performs slightly better than BindN (a sequence-based RNA-binding site prediction method).

The role of electrostatic interaction in preorientation is not significant in two classes of protein–RNA complexes. The first class includes single chain RNA without
base pairing. In this case, the stacking interaction but not electrostatic interaction between bases and amino acids is the dominating interaction. The second is when RNA is buried by its receptor protein. In this case, the protein subunits in this class seem to get their closed states through larger conformational change. The electrostatic fields are different between the closed state and its corresponding open state.

Acknowledgments

This work is supported by the NSFC under Grant nos. 31100522 and 11174093 and the National High Technology Research and Development Program of China (2012AA020402) and Specialized Research Fund for the Doctoral Program of Higher Education (20110142120038).

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2012.708604.

References

Allers, J., & Shamoo, Y. (2001). Structure-based analysis of protein–RNA interactions using the program ENTANGLE. *Journal of Molecular Biology, 311*, 75–86.

Alsallag, R., & Zhou, H.X. (2008). Electrostatic rate enhancement and transient complex of protein–protein association. *Proteins, 71*, 320–335.

Bahadur, R.P., Zacharias, M., & Janin, J. (2008). Dissecting protein–RNA recognition sites. *Nucleic Acids Research, 36*, 2705–2716.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., & McCammon, J.A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 10037–10041.

Barik, A., Nithin, C., Manasa, P., & Bahadur, R.P. (2012). A protein–RNA docking benchmark (I): Non-redundant cases. *Proteins, 80*, 1866–1871.

Berman, H.M., Westbrook, J., Feng, Z., Gililland, G., Bhat, T.N., Weissig, H., … Bourne, P.E. (2006). The protein data bank, 1999. *Nucleic Acids Research, 28*, 235–242.

Camacho, C.J., Weng, Z., Vajda, S., & DeLisi, C. (1999). Free energy landscapes of encounter complexes in protein–protein association. *Biophysical Journal, 76*, 1166–1178.

Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C., L., Wang, J., Duke, R.E., … Kollman, P.A. (2010). *AMBER 11*. San Francisco, CA: University of California.

Chen, R., Li, L., & Weng, Z. (2003). ZDOCK: An initial-stage protein-docking algorithm. *Proteins, 52*, 80–87.

Chen, Y.C., & Lim, C. (2008). Predicting RNA-binding sites from the protein structure based on electrostatics, evolution and geometry. *Nucleic Acids Research, 36*, e29.

Chen, R., & Weng, Z. (2002). Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins, 47*, 281–294.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., & Baker, N.A. (2004). PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Research, 32*, W665–W667.

Drozdov-Tikhomirov, L.N., Linde, D.M., Poroikov, V.V., Alexandrov, A.A., & Skurida, G.I. (2001). Molecular mechanisms of protein–protein recognition: Whether the surface placed charged residues determine the recognition process? *Journal of Biomolecular Structure and Dynamics, 19*, 279–284.

Drozdov-Tikhomirov, L.N., Linde, D.M., Poroikov, V.V., Alexandrov, A.A., Skurida, G.I., Kovalev, P.V., & Pota-pov, V.Y. (2003). About factors providing the fast protein–protein recognition in processes of complex formation. *Journal of Biomolecular Structure and Dynamics, 21*, 257–266.

Ellis, J.J., Broom, M., & Jones, S. (2007). Protein–RNA interactions: Structural analysis and functional classes. *Proteins, 66*, 903–911.

Fernandez-Recio, J., Totrov, M., & Abagyan, R. (2004). Identification of protein–protein interaction sites from docking energy landscapes. *Journal of Molecular Biology, 335*, 843–865.

Gabb, H.A., Jackson, R.M., & Sternberg, M.J. (1997). Modeling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology, 272*, 106–120.

Gabdoulline, R.R., & Wade, R.C. (1997). Simulation of the diffusional association of barnase and barstar. *Biophysical Journal, 72*, 1917–1929.

Gong, Z., & Xiao, Y. (2010). RNA stability under different combinations of amber force fields and solvation models. *Journal of Biomolecular Structure and Dynamics, 28*, 431–441.

Gong, Z., Zhao, Y., Chen, C., & Xiao, Y. (2011). Role of ligand binding in structural organization of add A-ribo-switch aptamer: A molecular dynamics simulation. *Journal of Biomolecular Structure and Dynamics, 29*, 403–416.

Gupta, A., & Gribskov, M. (2011). The role of RNA sequence and structure in RNA–protein interactions. *Journal of Molecular Biology, 409*, 574–587.

Janin, J. (1997). The kinetics of protein–protein recognition. *Proteins, 28*, 153–161.

Jones, S., Daley, D.T., Luscombe, N.M., Berman, H.M., & Thornton, J.M. (2001). Protein–RNA interactions: A structural analysis. *Nucleic Acids Research, 29*, 943–954.

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Altman, C., & Vaks, L.A. (1992). Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proceedings of the National Academy of Sciences of the United States of America, 89*, 2195–2199.

Koh, Y.Y., Wang, Y., Qiu, C., Opperman, L., Gross, L., Tanaka, H.T.M., & Wickens, M. (2011). Stacking interactions in PUF-RNA complexes. *RNA, 17*, 718–727.

Koukiekolo, R., Sagan, S.M., & Pezacki, J.P. (2007). Effects of pH and salt concentration on the siRNA binding activity of the RNA silencing suppressor protein p19. *FEBS Letters, 581*, 3051–3056.

Kovalev, P.V., Drozdov-Tikhomirov, L.N., Poroikov, V.V., & Alexandrov, A.A. (2004). Role of the electrostatic interactions in pre-orientation of subunits in the formation of protein–protein complexes. *Journal of Biomolecular Structure and Dynamics, 22*, 111–118.

Kumar, M., Gromiha, M.M., & Raghava, G.P. (2008). Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins, 71*, 189–194.

Lattman, E.E. (1972). Optimal sampling of the rotation function. *Acta Crystallographica Section B, 28*, 1065–1068.
Li, C.H., Cao, L.B., Su, J.G., Yang, Y.X., & Wang, C.X. (2012). A new residue-nucleotide propensity potential with structural information considered for discriminating protein–RNA docking decoys. *Proteins*, 80, 14–24.

Li, L., Guo, D., Huang, Y., Liu, S., & Xiao, Y. (2011). ASPDock: Protein–protein docking algorithm using atomic solvation parameters model. *BMC Bioinformatics*, 12, 36.

Liu, Z.P., Wu, L.Y., Wang, Y., Zhang, X.S., & Chen, L. (2010). Prediction of protein–RNA binding sites by a random forest method with combined features. *Bioinformatics*, 26, 1616–1622.

Meyer, K.A., Ng, K.C., Gu, Z., Pan, Z., Whitten, W.B., & Shaw, R.W. (2010). Combined apertureless near-field optical second-harmonic generation/atomic force microscopy imaging and nanoscale limit of detection. *Applied Spectroscopy*, 64, 1–7.

Morozova, N., Allers, J., Myers, J., & Shamoo, Y. (2006). Protein–RNA interactions: Exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics*, 22, 2746–2752.

Neumann, J., & Gottschalk, K.E. (2009). The effect of different force applications on the protein–protein complex barnase–barstar. *Biophysical Journal*, 97, 1687–1699.

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.

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. *Pacific Symposium on Biocomputing*, 15, 293–301.

Qin, S., & Zhou, H.X. (2009). Dissection of the high rate constant for the binding of a ribotoxin to the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 6974–6979.

Schreiber, G., & Fersht, A.R. (1996). Rapid, electrostatically assisted association of proteins. *Natural Structural Biology*, 3, 427–431.

Schreiber, G., Haran, G., & Zhou, H.X. (2009). Fundamental aspects of protein–protein association kinetics. *Chemical Reviews*, 109, 839–860.

Schreiber, G., Shaul, Y., & Gottschalk, K.E. (2006). Electrostatic design of protein–protein association rates. *Methods in Molecular Biology*, 340, 235–249.

Sinha, N., & Smith-Gill, S.J. (2002). Electrostatics in protein binding and function. *Current Protein and Peptide Science*, 3, 601–614.

Spaa, A., Dammer, C., Gabdoulline, R.R., Wade, R.C., & Helms, V. (2006). Diffusional encounter of barnase and barstar. *Biophysical Journal*, 90, 1913–1924.

Vakser, I.A. (1997). Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins*, 29 (Suppl. 1), 226–230.

Vargason, J.M., Szittya, G., Burgyan, J., & Hall, T.M. (2003). Size selective recognition of siRNA by an RNA silencing suppressor. *Cell*, 115, 799–811.

Wallis, R., Moore, G.R., James, R., & Kleanthous, C. (1995). Protein–protein interactions in colicin E9 DNase-immunity protein complexes. 1. Diffusion-controlled association and femtomolar binding for the cognate complex. *Biochemistry*, 34, 13743–13750.

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

Zhang, J., Bian, Y., Lin, H., & Wang, W. (2012). RNA fragment modeling with a nucleobase discrete-state model. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 85, 021909.

Zhang, J., Dundas, J., Lin, M., Chen, R., Wang, W., & Liang, J. (2009). Prediction of geometrically feasible three-dimensional structures of pseudoknotted RNA through free energy estimation. *RNA*, 15, 2248–2263.

Zhang, C., & Lai, L. (2011). SDOCK: A global protein–protein docking program using stepwise force-field potentials. *Journal of Computational Chemistry*, 32, 2598–2612.

Zhang, Y., Zhang, J., & Wang, W. (2011). Atomistic analysis of pseudoknotted RNA unfolding. *Journal of American Chemical Society*, 133, 6882–6885.

Zhao, Y., Gong, Z., & Xiao, Y. (2011). Improvements of the hierarchical approach for predicting RNA tertiary structure. *Journal of Biomolecular Structure and Dynamics*, 28, 815–826.

Zhao, H., Yang, Y., & Zhou, Y. (2011). Structure-based prediction of RNA-binding domains and RNA-binding sites and application to structural genomics targets. *Nucleic Acids Research*, 39, 3017–3025.