The molecular architecture of protein-RNA interfaces are analyzed using a non-redundant dataset of 152 protein-RNA complexes. We find that an average protein-RNA interface is smaller than an average protein-DNA interface but larger than an average protein–protein interface. Among the different classes of protein-RNA complexes, interfaces with tRNA are the largest, while the interfaces with the single-stranded RNA are the smallest. Significantly, RNA contributes more to the interface area than its partner protein. Moreover, unlike protein–protein interfaces where the side chain contributes less to the interface area compared to the main chain, the main chain and side chain contributions flipped in protein-RNA interfaces. We find that the protein surface in contact with the RNA in protein-RNA complexes is better packed than that in contact with the DNA in protein-DNA complexes, but loosely packed than that in contact with the protein in protein–protein complexes. Shape complementarity and electrostatic potential are the two major factors that determine the specificity of the protein-RNA interaction. We find that the H-bond density at the protein-RNA interfaces is similar with that of protein-DNA interfaces but higher than the protein–protein interfaces. Unlike protein-DNA interfaces where the deoxyribose has little role in intermolecular H-bonds, due to the presence of an oxygen atom at the 2′ position, the ribose in RNA plays significant role in protein-RNA H-bonds. We find that besides H-bonds, salt bridges and stacking interactions also play significant role in stabilizing protein-nucleic acids interfaces; however, their contribution at the protein–protein interfaces is insignificant.

**Keywords:** protein-RNA complexes; protein-RNA interfaces; interfaces area; shape complementarity; salt bridges; hydrogen bond

1. **Introduction**

Protein-RNA interactions play vital role in most of the cellular processes. They are important for the assembly and function of ribonucleoprotein particles (RNPs) exemplified by ribosomes and spliceosomes, as well as for the post-transcriptional regulation of gene expression (Cech & Bass, 1986; Draper, 1995; Noller, 1991). Protein-RNA interactions are also essential for chromatin end maintenance, animal development, and virus replication and dissemination (Cusack, 1999; Phipps & Li, 2007). Despite their immense importance, structural basis of protein-RNA interactions are still poorly understood mainly due to the scarcity of three-dimensional atomic structures of such complexes in the Protein Data Bank (PDB) (Berman et al., 2000). Recently, with the advancement in the field of structure determination of protein-RNA complexes, there has been a steady increase in the number of atomic structures of such complexes in the PDB. They can be used for the structure-based analysis to understand the specificity of protein-RNA recognition. The interactions between protein and RNA subunits can be studied in terms of geometric, structural, and physicochemical properties of their interfaces. The knowledge gained from the analysis would provide an insight into the prediction of the RNA-binding sites on the protein surface, docking of protein-RNA complexes, and prediction of binding affinity.

Earlier, few studies have been carried out to understand the structural basis of protein-RNA interactions (Allers & Shamoo, 2001; Jeong, Kim, Lee, & Han, 2003; Jones, Daley, Luscombe, Berman, & Thornton, 2001; Nadassy, Wodak, & Janin, 1999; Treger & Westhof, 2001). These studies have mainly focused on the atomic contacts that mediate the recognition and contribute to the stability of the complexes. However, these studies are based on smaller dataset of protein-RNA complexes. Later, Ellis, Broom, and Jones (2007) and Bahadur, Zacharias, and Janin (2008) have used comparatively larger dataset of protein-RNA complexes. With the exponential increase of structures of protein-RNA complexes in the PDB, it is now necessary to update the dataset and to carry out a statistical analysis on structural and physicochemical characteristics of protein-RNA recognition sites.

*Corresponding author. Email: r.bahadur@hijli.iitkgp.ernet.in

© 2015 Taylor & Francis
In the present study, we have curated a non-redundant dataset of 152 X-ray structures of protein-RNA complexes. The dataset has been classified into four classes based on the type of RNA associated with its partner protein. A number of structural, geometric, and physicochemical parameters are calculated to understand the specificity of the recognition process. The variation of the interface parameters among the different classes of complexes is also studied. We extend our analysis to protein-DNA and protein–protein interfaces; and compared the results to understand the specificity of macromolecular recognition in each type of complexes. This analysis may contribute to the development of novel protein-RNA scoring functions for docking applications, prediction of the RNA-binding sites in proteins as well as prediction of protein-RNA binding affinity. Moreover, our analysis is not only restricted to the binary complexes but can also be extended to the multisubunit assemblies to understand the specificity and affinity of macromolecular interactions that stabilize the assemblies.

2. Materials and methods

2.1. Dataset of the protein-RNA complexes

The atomic structures of protein-RNA complexes with resolution better than 3 Å were curated from the PDB. Permanent multisubunit assemblies such as ribosomes and viral capsids were not included in the present study. All the selected complexes are non-obligate with the possible exception of ribosomal proteins. We kept only those protein-RNA complexes in the dataset where the protein chains contain at least 30 amino acids and RNA chains contain at least five nucleotides (nt). In order to remove the redundancy, we performed pairwise sequence alignment for all the entries in the dataset using BLAST (Altschul, Gish, Miller, Myers, & Lipman 1990). When the protein or the RNA components in two complexes had more than 35% sequence identity, the one with the better resolution was retained. The results were confirmed by performing pairwise sequence alignment using structural superposition in PDBeFold (Krissinel & Henrick, 2004). The sequence identity values obtained using structural superposition is essentially similar to those obtained using BLAST. The final dataset consists of 152 protein-RNA complexes (Table 1). The references and other parameters of the complexes are reported in the Supplementary Table S1. Generally, the modified residues and nucleotide bases are marked with the keyword “HETATM” in the coordinate list in a PDB file. They were substituted by their corresponding amino acids and bases by changing the keyword “HETATM” to “ATOM,” and a list of those is given in the Supplementary Table S2. Datasets of protein-DNA and protein–protein complexes were taken from Setny, Bahadur, and Zacharias (2012), and Hwang, Vreven, Janin, and Weng (2010), respectively.

The complexes were classified into four different classes based on the type of RNA associated with its partner protein (Bahadur et al., 2008). Class A consists of 31 complexes with tRNA. Class B comprises 10 complexes with ribosomal proteins. In a crystallized protein-RNA complex, the RNA interacting with the protein is present in either single-stranded form or in stem-loop form. If the protein mostly interacts with the stem region of the RNA, the complex was assigned under Class C (duplex RNA), and if the protein mostly interacts with the loop region of the RNA, it was categorized under Class D (single-stranded RNA). Class C and Class D consists of 49 and 62 complexes, respectively.

2.2. Interface area, secondary structure, and statistical test

The interface area of a macromolecular complex was estimated as the solvent accessible surface area (SASA) buried in complexation. The interface area or the buried surface area (B) of the protein-RNA complex was calculated by subtracting the SASA of the complex from the sum of the SASAs of the individual subunits (Equation 1).

\[ B = \text{SASA}_{\text{protein}} + \text{SASA}_{\text{RNA}} - \text{SASA}_{\text{complex}} \]  

(1)

The software NACCESS (Hubbard & Thornton, 1993), that implements the Lee and Richards (Lee & Richards, 1971) algorithm, was used for the calculation of SASA. The B of a protein-RNA complex was calculated using the web server PRince (Barik, Mishra, & Bahadur 2012), which uses NACCESS with default parameters. Atoms that loose SASA after complexation were identified as the interface atoms. The program DSSP (Kabsch & Sander, 1983) was used for the secondary structure assignment to protein chains. ANOVA (analysis of variance) test and Student t-test (both one-tailed and two-tailed) at \( \alpha = .05 \) level of significance were performed to find the statistical significance of the calculated parameters.

2.3. Hydrogen bonds, salt bridges, and stacking interactions

Hydrogen bonds (H-bond) at protein-RNA interfaces were calculated using the software HBPLUS (McDonald & Thornton, 1994) with default parameters. The salt bridges at protein-RNA interfaces were calculated when the distance between the side chain nitrogen atoms of positively charged residues and the negatively charged phosphate group of the nucleotides is within 4 Å (Barlow & Thornton, 1983; Xu, Tsai, & Nussinov, 1997).

Stacking interactions at protein-RNA interfaces are usually defined as the \( \pi-\pi \) interactions that can occur
between the side chains of Tyr, Trp, Phe, His and the RNA bases. Moreover, the π–π and π–cation stacking of Arg through its guanidinium moiety onto nucleosides were included in the calculation of stacking interactions (Allers & Shamoo, 2001). We defined the planes at both sides by considering the atoms constituting the aromatic rings, and the center of the plane was calculated as the mid-point of all these atoms. The cut-off distance between the centers of both the planes was kept \( \leq 5 \text{Å} \), and the dihedral angle between the two planes was constrained to \( \leq 30^\circ \) (Allers & Shamoo, 2001). The stagger angle is defined as the angle between the normal to the first plane and the vector joining the centers of the two planes.

2.4. **Shape complementary and Gap volume indices**

The shape correlation index \( (S_c) \) (Lawrence & Colman, 1993) was used to quantify the shape complementarity at protein-RNA interfaces. The \( S_c \) index was calculated using the GPU-SC program (https://www.doe-mbi.ucla.edu/people/luki/sc) with a radii file that was modified to include the nucleotide atoms. Atomic packing at protein-RNA interfaces was also evaluated using the gap volume index (GV) (Jones & Thornton, 1996) given by the following equation

\[
\text{GV Index (Å)} = \frac{\text{Gap volume between molecules}}{B}
\]  

(2)

The gap volume for each complex was calculated using the SURFNET program (Laskowski, 1995).

3. **Results**

3.1. **The protein and the RNA components in protein-RNA complexes**

The dataset of 152 protein-RNA complexes is shown in Table 1. With the possible exception of the ribosomal proteins, all the complexes in our dataset are non-obligate, that is, the protein and the RNA are not permanently associated. The length of the RNA in the complexes varies from 5 to 159 nt with an average of 34 nt. Complexes in Class A have the longest RNA with an average length of 73 nt, while the complexes in Class D have the shortest ones with an average length of only 14 nt. For the protein component, complexes in Class A have the largest polypeptide length with an average of 632 residues, while the complexes in Class B have the smallest polypeptide length with an average of 164 residues (Table 2).

![Table 1. Dataset of protein-RNA complexes.](image-url)

| (A) Complexes with tRNA (31) | | | |
|---|---|---|---|
| 1ASY | 1B23 | 1C0A | 1F7U |
| 1FFY | 1GAX | 1H3E | 1H4S |
| 1J1U | 1N78 | 1QF6 | 2AZX |
| 2CSX | 2DLC | 2DRB | 2DU3 |
| 2FK6 | 2FMT |

| (B) Complexes with ribosomal proteins (10) | | | |
|---|---|---|---|
| 1DFU | 1FEU | 1G1X | 1H6U |
| 1MI | 1MJS | 1MZP | 1S03 |
| 1SDS | 2HW8 |

| (C) Complexes with duplex RNA (49) | | | |
|---|---|---|---|
| 1DI2 | 1H0Q1 | 1MSW | 1N35 |
| 1OOA | 1Q2R | 1R3E | 1R9F |
| 1SI | 1WNE | 1YVP | 2F8S |
| 2GJW | 2GXB | 2OZB | 2PI |
| 2UX | 2R8S |

| (D) Complexes with single-stranded RNA (62) | | | |
|---|---|---|---|
| 1AV6 | 1C9S | 1CVJ | 1G2E |
| 1CJS | 1JBS | 1JID | 1K8W |
| 1KNZ | 1KQ2 | 1LNG | 1M5O |
| 1M8W | 1MV | 1WPU | 1WSU |
| 1ZBH | 1ZH5 | 2BA8 | 2ANR |
| 2ASB | 2B3J |
| 2BH2 | 2B2X | 2DBS | 2G4B |
| 2GIC | 2H2 | 2IX1 | 2J8S |
| 2JEA | 2JLU | 2PY9 |
| 2Q66 | 2QR7 | 2VNU | 2XG |
| 2XRZ | 2XZO | 2AEV | 2BX2 |
| 2D2S | 315X |
| 3IEV | 3K5Q | 3MDG | 3NNR |
| 3O8C | 3PF4 | 3QJ5 | 3R2C |
| 3RC8 | 3T5N | 4E78 |

3.2. **Size of the interface area**

Table 2 shows the size of the protein-RNA interfaces in each class as well as in the entire dataset. Complexes in Class A have the largest average interface area \( (B = 3575 \text{Å}^2) \), while complexes in Class D have the smallest \( (B = 2016 \text{Å}^2) \). ANOVA test confirms the statistical significance of the different values of \( B \) among the four classes \( (p = 5E-7) \). In the entire dataset, the average size of a protein-RNA interface is 2545 \( \text{Å}^2 \). An average protein-RNA interface contains 45 residues and 16 nt. Figure 1 shows the distribution of the size of the
Table 2. Average properties of protein-RNA interfaces.

| Interface parameters | tRNA | Ribosomal Duplex | Single-stranded | All classes | Protein-DNA<sup>a</sup> | Protein–protein<sup>b</sup> |
|----------------------|------|------------------|-----------------|-------------|---------------------|---------------------|
| No of complexes      | 31   | 10   | 49   | 62   | 152                | 115                | 160                |
| Amino acids          |      |      |      |      |                    |                    |                    |
| Range                | 193–1998 | 94–232 | 62–1514 | 66–1370 | 62–1998 | 51–1553 | 44–2000 |
| Mean no per complex  | 632  | 164  | 401   | 356   | 414 ± 334         | 305 ± 230          | 303 ± 199          |
| Nucleotides          |      |      |      |      |                    |                    |                    |
| Range                | 34–148 | 30–84 | 9–159 | 5–97 | 5–159 | 14–70 | –     |
| Mean no per complex  | 73   | 46   | 33    | 14   | 34 ± 30 | 35 ± 12 | –     |
| Average B (Å<sup>2</sup>) | 3575 ± 1260 | 2252 ± 1074 | 2623 ± 1247 | 2016 ± 1155 | 2545 ± 1326 | 3137 ± 1351 | 1919 ± 758 |
| Protein              | 1720 | 1109 | 1255  | 931   | 1208 ± 638        | 1543 ± 670         | –                 |
| Nucleic acid         | 1856 | 1144 | 1368  | 1085  | 1337 ± 691        | 1594 ± 687         | –                 |
| No of interface      |      |      |      |      |                    |                    |                    |
| Amino acids N<sub>aa</sub> | 63   | 35   | 45    | 37   | 45 ± 23 | 52 ± 25 | 57 ± 12 |
| Nucleotides N<sub>nu</sub> | 27   | 22   | 17    | 9    | 16 ± 10 | 23 ± 9 | –     |
| B (Å<sup>2</sup>) per Amino acid | 27   | 32   | 28    | 25   | 27    | 30    | 34    |
| Nucleotide           | 70   | 53   | 79    | 122   | 83    | 69    | –     |
| Fraction of buried atoms (f<sub>bu</sub>) (%) |      |      |      |      |                    |                    |                    |
| Protein              | 24   | 33   | 27    | 33   | 29 ± 9 | 26 ± 8 | 33 ± 9 |
| Nucleic acid         | 24   | 33   | 29    | 32   | 29 ± 8 | 29 ± 8 | –     |
| Packing index LD     |      |      |      |      |                    |                    |                    |
| Protein              | 38   | 40   | 39    | 42   | 40 ± 8 | 43 ± 8 | 42 ± 6 |
| Nucleic acid         | 42   | 43   | 46    | 49   | 46 ± 10 | 51 ± 12 | –     |
| S<sub>g</sub> index<sup>c</sup> | .66  | .65  | .66   | .69  | .67 ± .07 | .65 ± .06 | .7 ± .07 |
| Gv index<sup>d</sup>  | 9.8  | 6.1  | 9.2   | 8.5  | 8.9 ± 6.4 | 6.11 ± 4.4 | 9.67 ± 6.6 |
| H-bonds<sup>e</sup>  |      |      |      |      |                    |                    |                    |
| No per interface     | 25   | 20   | 21    | 15   | 19 ± 10 | 22 ± 11 | 9 ± 5  |
| B per bond (Å<sup>2</sup>) per 1000 Å<sup>2</sup> B | 145  | 113  | 125   | 131  | 134    | 143    | 213    |
| Salt bridges<sup>f</sup> |      |      |      |      |                    |                    |                    |
| No per interface     | 11   | 10   | 9     | 6    | 8 ± 7  | 11 ± 6 | 2 ± 3  |
| B per bridge (Å<sup>2</sup>) per 1000 Å<sup>2</sup> B | 336  | 217  | 276   | 366  | 312    | 293    | 1066   |
| Secondary structures<sup>g</sup> |      |      |      |      |                    |                    |                    |
| No. of residues in helix per interface | 24   | 10   | 18    | 14   | 17 ± 15 | 20 ± 13 | 16 ± 14 |
| per 1000 Å<sup>2</sup> B | 14   | 9    | 14    | 16   | 14    | 13    | 8     |
| No. of residues in strands per interface | 10   | 10   | 9     | 8    | 9 ± 8  | 9 ± 10 | 12 ± 10 |
| per 1000 Å2 B | 6    | 9    | 7     | 8    | 7     | 6     | 6     |
| No. of residues in others per interface | 28   | 15   | 19    | 15   | 19 ± 11 | 24 ± 14 | 29 ± 14 |
| per 1000 Å2 B | 16   | 13   | 15    | 16   | 16    | 15    | 15    |

<sup>a</sup> Calculated on a dataset of 115 protein-DNA complexes taken from (Setny et al., 2012).
<sup>b</sup> Calculated on a dataset of 160 protein–protein complexes taken from (Hwang et al., 2010).
<sup>c</sup> S<sub>g</sub> index was calculated using the GPU-SC program ([https://www.doe-mbi.ucla.edu/people/luki/sc](https://www.doe-mbi.ucla.edu/people/luki/sc)).
<sup>d</sup> Gap index was calculated using SURFNET program (Laskowski, 1995).
<sup>e</sup> H-bonds were calculated using the program HBPLUS (McDonald & Thornton, 1994).
<sup>f</sup> For both protein-nucleic acids and protein–protein interfaces, donor atoms: NZ of Lys, NE, NH1 and NH2 of Arg. For protein-nucleic acids, acceptor atoms: OP1, OP2 while for protein–protein interfaces, acceptor atoms: OD1, OD2 of Asp and OE1, OE2 of Glu were considered for calculating salt bridges.
<sup>g</sup> Secondary structures of proteins were calculated using the program DSSP (Kabsch & Sander, 1983). The number of interface residues found in “helix” or in “strands” or in “others” are given.
interfaces in the entire dataset as well as in each class. The histogram illustrates the wide variation in $B$ with two broad peaks at 2000 Å$^2$ and at 4800 Å$^2$. The former peak is mainly contributed by the Class C and Class D complexes, while the later peak is contributed mainly by the Class A complexes. Except for three complexes (one in Class C and two in Class D), all the complexes in the entire dataset bury more than 800 Å$^2$ (Supplementary Table S1). In Class A, size of the interfaces varies from 1355 to 5769 Å$^2$; in Class B it varies from 1200 to 2949 Å$^2$ with only one exception. The only exception (PDB id: 1G1X; $B = 4945$ Å$^2$) arises due to the interaction between the three protein chains and the two RNA chains. The size of the interfaces in Class C is in the range of 769–7076 Å$^2$. The complex between Z-alpha domain of RNA editing enzyme ADAR1 with double-stranded Z-RNA (PDB id: 2GXB) is the only example in Class C where the interface buries less than 800 Å$^2$. With two exceptions, Class D has interfaces ranging from 814–7843 Å$^2$. In these two exceptions (PDB id: 3D2S, 2A8V), $B$ is less than 800 Å$^2$.

3.3. Asymmetry in protein-RNA interface area

The contribution of the protein and the RNA molecules to $B$ is different: RNA contributes more than its partner protein. This was observed in all the four classes. The asymmetry in the contribution to $B$ can be calculated by the following equation:

$$R = \left( \frac{B_R - B_P}{B} \right)$$

(3)

where $B_R$ is the interface area contributed by the RNA, $B_P$ is the interface area contributed by the protein. In the entire dataset, the average $R$ is 5%, and it varies from 2% in Class B to 8% in Class D. Average $R$ in Class A is 3%, while it is 4% in Class C. Two complexes in Class D (PDB id: 1KNZ and 2A8V) have $R > 20%$. In both the structures the small RNA is completely buried inside the protein. In majority of cases, where $R$ exceeds 10%, the interacting RNA length is less than 15 nt. On the other hand, in 32 complexes where the RNA length exceeds 60 nt, the average $R$ is 4%. This variation is attributed to the convex nature of the RNA surface that binds to the concave protein surface (Bahadur et al., 2008). Since SASA is measured one probe radius away from the molecular surface, the concave surface of the proteins loses less SASA, and thus contributes less to $B$.

3.4. Interface atoms, residues, and nucleotides

Figure 2(A) shows that the number of interface atoms linearly increases with $B$ with a correlation coefficient ($R^2$) of .98 and .97 for the protein and the RNA, respectively. On average, an interface protein atom contributes 8.0 Å$^2$ to $B$, while an interface RNA atom contributes 9.5 Å$^2$. Thus, a protein atom contributes 17% less to $B$ compared to a RNA atom. Similar trend is observed for the interface residues and nucleotides. On the protein side, the number of interface residues linearly increases with $B$ (Figure 2(B), $R^2 = .93$). On average, an interface residue loses 27 Å$^2$ SASA, and this varies from 25 Å$^2$ in Class D to 32 Å$^2$ in Class B (Table 2). On average, an interface nucleotide loses 83 Å$^2$ SASA and this varies from 53 Å$^2$ in Class B to 122 Å$^2$ in Class D. The $B$ per nucleotide depends on the length of the RNA, though the correlation between them is mediocre (Figure 2(B), $R^2 = .67$). In 55 complexes where the RNA length ≤15 nt, an average nucleotide contributes 137 Å$^2$ to $B$. Of these 55 complexes, 48 belong to Class D and 7 belong to Class B. However, for 32 complexes having long...
RNA with 60 nt or more, the average interface nucleotide contributes only 68 Å² to B. About 81% of them belong to Class A. This twofold difference can be explained by the fact that a short RNA usually adopts a single-stranded extended conformation where the nucleotides are more accessible to the protein, compared to RNA with folded tertiary structure.

3.5. **Buried atoms and the local packing density**

During complex formation, an atom may lose its SASA completely, and thus become fully buried in the interface. We calculated the fraction of such buried atoms using the following equation

$$f_{bu} = \frac{\text{Number of fully buried interface atoms}}{\text{Total number of interface atoms}}$$  \hspace{1cm} (4)

On average, $f_{bu}$ on the protein side and the RNA side of the interfaces is similar (29%) in the entire dataset as well as in the different classes ($p = .6$). However, among the different classes, $f_{bu}$ on protein and RNA sides varies. While Class A has the lowest number of 24% (both for protein and RNA), Class B and Class D both have the highest number of about 33% (both for protein and RNA), with $P$ value of 5E-6.

While $f_{bu}$ measures the compactness of the atomic packing at the interface, the local density index (LD) measures the atomic density at each point of the interface (Bahadur, Chakrabarti, Rodier, & Janin, 2004). In brief, LD is defined as the mean number of interface atoms that are within 12 Å of another interface atom. An interface having $N$ atoms, and if $n_i$ atoms are within a distance of 12 Å from a given interface atom $i$, the LD for that subunit is calculated as below

$$\text{LD} = \sum \frac{n_i}{N}$$  \hspace{1cm} (5)

The LD values were calculated for the protein side and the RNA side separately. At the protein side, the density at each atomic position is lower (LD = 40) than at the RNA side (LD = 46) with a $P$ value of 1E-21. This trend is similar in the entire dataset as well as in the different classes.

3.6. **Shape complementarity and gap volume index**

The shape correlation index ($S_c$) was used to quantify the shape complementarity at protein-RNA interfaces (Table 2). $S_c$ index depends on the shape of the interacting surfaces. In protein-RNA complexes, the $S_c$ index varies from .46 to .83 with an average of .67. This average value is essentially similar in different classes ($p = .03$). However, 37% of the protein-RNA complexes resemble protein–protein interfaces with $S_c$ index $\geq .70$. Most of them (55%) belong to Class D with single-stranded RNA. Figure 3 shows the shape complementarity of the protein-RNA interface in four different complexes. In all four cases, convex RNA surface binds to the concave protein surface. The concave nature of the protein surface is an important feature that is often used to predict a possible RNA-binding site (Bahadur et al., 2008).

The GV index of an average protein-RNA interface is 8.9 Å (Table 2). Among the different classes, Class A has the highest GV index (9.8), followed by Class C (9.2), Class D (8.5), and Class B (6.1). The high GV value reflects the poor packing of the interface.
3.7. Chemical composition

Table 3 shows the average chemical composition of the interface and the solvent accessible surface, which are calculated based on the contribution of different types of atoms to B or to SASA. At protein-RNA interfaces, the protein side chain contributes much more to B as well as to SASA than its main chain. The side chain contributes 86% to B, while the main chain contributes only 14%. The contribution of the non-polar (carbon-containing) groups to B as well as to SASA is more at the protein side (55–56%) than at the RNA side (30–33%). At the protein side, the contribution of the neutral polar groups to B and to SASA is almost similar, 21 and 23%, respectively. The positively charged side chains of Arg and Lys contribute more to B than the negatively charged side chains of Asp and Glu. However, the different charged groups contribute equally to SASA. Moreover, positively charged groups are dominant at the interface than at the solvent accessible surface, while the negatively charged groups are dominant at the solvent accessible surface than at the interface.

At the RNA side, contribution of the neutral polar groups to B is the highest (45%), followed by the non-polar groups (33%) and the negatively charged groups (22%). Similar trend is also observed at the solvent accessible surface of the RNA. Both the ribose and the bases contribute equally (38%) to B, while the phosphate contributes little less (25%). At the solvent accessible surface of the RNA, the ribose contributes maximum (38%) followed by the phosphate (34%) and the bases (28%). Between B and SASA, the bases are more abundant at B and the phosphate is more abundant at SASA, while the ribose is equally found at both the surfaces.

3.8. The amino acid and the nucleotide composition

The average composition of the amino acid residues and the nucleotides of the interface and the solvent accessible surface are given in Table 4. The compositions are calculated based on the fraction of surface area contributed by each of the residues or nucleotides to B or to SASA. At the interface, U contributes maximum (30%), while the other three bases contribute almost equally (about 23%). At the solvent accessible surface of the RNA, U contributes maximum (29%) followed by G (26%), C (24%), and A (21%). Between B and SASA, the pyrimidines (U and C) contribute equally at both the surfaces; while A contributes more to B and G contribute more to SASA.
At the protein side, the interface is enriched with the basic residues like Arg and Lys, and depleted in acidic residues like Asp and Glu. Among the aromatic residues, Phe and Tyr are abundant at the interface than Trp. At the solvent exposed surface, charged residues contribute maximum followed by the neutral polar residues. The differences in the amino acid composition at different surfaces can be expressed by the Euclidean distance $\Delta f$:

$$\Delta f^2 = \frac{1}{19} \sum_{i=1}^{20} (f_i - f'_i)^2$$  \hspace{1cm} (6)

where $f_i$ and $f'_i$ are the area fractions contributed by the residue type $i$ to the surface or to the interface (Lo Conte, Chothia, & Janin, 1999), respectively. Figure 4 compares the Euclidean distances among the different surfaces in protein-RNA, protein-DNA, and protein–protein complexes. The amino acid composition of the protein surface in contact with the RNA is quite different than the protein surface that is solvent exposed. The differences in the amino acid composition at different surfaces can be expressed by the Euclidean distance $\Delta f$.

At the solvent exposed surface, charged residues contribute maximum followed by the neutral polar residues. The differences in the amino acid composition at different surfaces can be expressed by the Euclidean distance $\Delta f$:

$$\Delta f^2 = \frac{1}{19} \sum_{i=1}^{20} (f_i - f'_i)^2$$  \hspace{1cm} (6)

where $f_i$ and $f'_i$ are the area fractions contributed by the residue type $i$ to the surface or to the interface (Lo Conte, Chothia, & Janin, 1999), respectively. Figure 4 compares the Euclidean distances among the different surfaces in protein-RNA, protein-DNA, and protein–protein complexes. The amino acid composition of the protein surface in contact with the RNA is quite different than the protein surface that is solvent exposed. Moreover, the protein surface in contact with the RNA is quite similar in terms of amino acid composition with the protein surface in contact with the DNA, but quite different from the protein surface in contact with another protein in protein–protein complexes. Similar trend is observed for the solvent exposed surfaces as well.

### 3.9. Hydrogen bonds

Polar interactions at the protein-RNA interfaces were quantified in terms of H-bonds. We identified total 2938 H-bonds in 152 protein-RNA interfaces. The number of H-bonds varies from 2 to 53 with an average of 19, which is equivalent to the density of one H-bond per 134 Å² of B. The number of H-bonds increase with the size of the interface, although the correlation between them is mediocre ($R^2 = .79$). The H-bond density in protein-RNA interfaces is marginally high compared to protein-DNA interfaces, but significantly higher than the protein–protein interfaces (Table 2). Table 2 also shows the variation of H-bonds in different classes of protein-RNA complexes. Although the average number of H-bond is more in Class A compared to the other classes, the H-bond density is the lowest in this class due to the large size of the interfaces. On the other hand, H-bond density is the highest in Class B complexes.

Table 5 lists the chemical composition of the H-bonds in protein-RNA and protein-DNA interfaces. In both types of complexes, the side chain groups of protein are involved more in H-bond than the main chain groups. The side chain contributes 75% of all protein-RNA H-bonds, while the main chain atoms contribute only 25%. In all H-bonds involving side chains, the charged groups contribute 43%, while the neutral polar groups contribute 32%. Between the main chain nitrogen and oxygen, the former is frequently found than the later in protein-RNA H-bonds.

On the RNA side, the contribution of the phosphate towards the H-bond is maximum (36%), followed by the bases (33%), and the sugar (31%). Among the different bases, G (10%) and U (9%) make more H-bonds than A and C (7% each). In comparison to protein-RNA
interfaces, the contribution of the phosphate to the H-bonds becomes almost double (70%) at the protein-DNA interfaces, while the deoxyribose contributes insignificantly (6%). The bases at the protein-DNA interfaces are less frequently (25%) involved in H-bonds compared to the bases at the protein-RNA interfaces (33%). This is justified with the phenomenon that the bases are already H-bonded in the canonical B-DNA structure compared to single stranded RNA. However, comparing Class C and Class D complexes, the bases in former are found less

Table 4. Amino acid and nucleotide compositions.

| Composition | Protein-RNA | Protein-DNA | Protein–protein |
|-------------|-------------|-------------|-----------------|
|             | Interface   | Surface     | Interface       | Surface |
| Nucleotides |             |             |                 |         |
| A           | 23.3        | 20.6        | 24.1            | 25.7    |
| U/T         | 30.0        | 29.5        | 29.9            | 26.8    |
| G           | 23.5        | 26.1        | 24.3            | 23.9    |
| C           | 23.2        | 23.8        | 21.7            | 23.6    |
| Residues    |             |             |                 |         |
| ALA         | 2.7         | 3.6         | 2.3             | 3.5     |
| ARG         | 20.1        | 11.0        | 23.4            | 11.4    |
| ASN         | 5.3         | 5.1         | 6.8             | 5.0     |
| ASP         | 3.8         | 7.2         | 2.1             | 7.1     |
| CYS         | -0.6        | -0.3        | 5.4             | 7.4     |
| GLN         | 4.2         | 5.7         | 4.8             | 6.1     |
| GLU         | 3.7         | 13.6        | 2.7             | 12.8    |
| GLY         | 4.4         | 3.8         | 4.0             | 3.3     |
| HIS         | 3.4         | 2.5         | 4.2             | 3.3     |
| ILE         | 3.5         | 3.2         | 2.0             | 2.9     |
| LEU         | 4.0         | 5.0         | 3.4             | 5.1     |
| LYS         | 14.9        | 14.8        | 15.0            | 14.5    |
| MET         | 1.8         | 1.8         | 1.5             | 2.2     |
| PHE         | 5.0         | 2.4         | 3.2             | 1.9     |
| PRO         | 3.4         | 4.8         | 2.7             | 4.7     |
| SER         | 5.2         | 4.1         | 6.9             | 4.8     |
| THR         | 4.0         | 4.1         | 5.9             | 4.2     |
| TRP         | 1.4         | 0.8         | 1.3             | 9.1     |
| TYR         | 5.7         | 2.9         | 4.6             | 2.7     |
| VAL         | 2.9         | 3.3         | 2.6             | 3.1     |

*Percent contribution of each type of nucleotide or amino acid residue to B or SASA.

*Calculated on a dataset of 115 protein-DNA complexes taken from (Setny et al., 2012).

*Calculated on a dataset of 160 protein–protein complexes taken from (Hwang et al., 2010).

Figure 4. Euclidean distance (Δf) between amino acid compositions at different surfaces. Values of Δf are calculated by equation 6 using the area based compositions given in Table 4.
the differences in number of stacking interactions made by different residues. The P value (at $\alpha = .05$ level of significance) obtained is .001, rejecting the null hypothesis (mean values are same in all the residues). Figure 5 illustrates the stacking interaction between Tyr615 with U8 and A9 in a Class D complex (PDB id: 3BX2). Here, the aromatic ring of Tyr is stacked between the two bases of RNA. The purines are involved more in stacking interactions (54%) compared to the pyrimidines (46%). Apart from the aromatic residues, Arg contributes significantly (40%) to the stacking interactions. Among the different classes of protein-RNA complexes, the highest number of stacking interactions is found in Class D (58%) interfaces, followed by the interfaces in Class C (27%) and in Class A (15%). Only one stacking interaction is found in Class B interfaces.

3.12. Secondary structural elements at protein-RNA interface

The average number and the density of residues found in different secondary structural elements at the protein-RNA interfaces are shown in Table 2. Here, the category “helices” includes alpha helix, $3_10$-Helix, and pi-helix; the category “strands” includes residues in isolated beta-strand, in extended strands or in beta ladder; and the category “others” includes non-regular secondary structures. An average protein-RNA interface comprises 38% of residues that form helices and 42% of residues that form other structural elements. The frequency of residues at the interface that form strands is minimum, only 20%. The number of residues per unit B (density) in three categories shows that the “others” have the highest density (16 per 1000 Å$^2$ B), the “helices” have the intermediate density (14 per 1000 Å$^2$ B), while the “strands” have the lowest density (7 per 1000 Å$^2$ B). Similar trend is observed in protein-DNA and protein–protein interfaces. Except the interfaces in Class B, the density remains same among the different classes of protein-RNA interfaces. In Class B, “others” are predominant (13 per 1000 Å$^2$ B) and “helices” and “strands” have similar densities (9 per 1000 Å$^2$ B).

4. Discussion

The present study provides a comprehensive analysis of structural and physicochemical features of the protein-RNA interfaces. Moreover, the analysis is also extended to the protein-DNA and the protein–protein interfaces, and the findings in three types of interfaces are compared. Earlier, several structure based analyses were carried out to understand the protein-RNA recognition (Allers & Shamoo, 2001; Bahadur et al., 2008; Ellis et al., 2007; Jeong et al., 2003; Jones et al., 2001; Nadassy et al., 1999; Treger & Westhof, 2001).

### Table 5. Chemical composition of H-bonds.

| H-bonds          | Protein-RNA | Protein-DNA$^a$ |
|------------------|-------------|-----------------|
| Total number     | 2938        | 2585            |
| Number per interface | 19         | 22              |
| Protein chemical group (%)$^b$ |             |                 |
| Main chain O     | 11          | 3               |
| Main chain N     | 14          | 23              |
| Side chain groups|             |                 |
| Charged          | 43          | 37              |
| Neutral         | 32          | 37              |
| Nucleic acid chemical group (%)$^c$ |             |                 |
| Phosphate        | 36          | 70              |
| Sugar$^d$        | 31          | 6               |
| Base             | 33          | 25              |
| Guanine          | 10          | 12              |
| Adenine          | 7           | 3               |
| Cytosine         | 7           | 6               |
| Uracil/Thymine   | 9           | 4               |

$^a$Calculated on a dataset of 115 protein-DNA complexes taken from (Setny et al., 2012).

$^b$Charged side chain includes: N of Arg, Lys and O of Asp, Glu. Neutral side chain includes: N of Asn, Gln, His, Trp; O of Asn, Gln; OH of Ser, Thr, Tyr; and S of Cys, Met.

$^c$Phosphate includes OP1, OP2, and P atoms.

$^d$Oxygen atoms O2′, O3′, O4′ and O5′ are attributed to the sugar moiety.

frequently (32%) involved in H-bonds than in later (49%).

3.10. Salt bridges

On an average, each protein-RNA interface is stabilized by eight salt bridges; however, the range is wide, 0 to 43. The distribution of salt bridges among different classes of protein-RNA interfaces is shown in Table 2. On an average, interfaces in Class A have the highest number of salt bridges (~11 per interface), while interfaces in Class D have the lowest number (~6 per interface). The average numbers in different classes are significantly different with a P value of 6E-4. We find that Arg makes significantly more salt bridges (67%) with the phosphate of RNA as compared to Lys (33%). This trend is similar both in protein–protein and in protein-DNA interfaces.

### Table 6. Stacking interactions between the side chain of aromatic residues and the nucleotide bases.

| Protein-RNA          | 125     |
|----------------------|---------|
| Total number         | 225     |
| Number per interface | 17      |
| Protein chemical group (%) |       |
| Main chain O         | 6       |
| Main chain N         | 8       |
| Side chain groups    |         |
| Charged              | 15      |
| Neutral              | 26      |

The stacking interactions between the side chain of aromatic residues and the nucleotide bases are observed in many protein-RNA complexes. Table 6 lists the stacking interactions in three different types of interfaces. A total of 255 stacking interactions were identified in 152 protein-RNA interfaces. Among the aromatic residues, Tyr makes the highest number of stacking interactions, while Trp makes the lowest number. Both Phe and Tyr prefer to stack with U, each making 25 interactions. ANOVA test was performed to find the statistical significance in
However, all these studies were restricted with relatively small number of protein-RNA complexes available in the PDB during that period. In this study, 152 non-redundant X-ray structures of protein-RNA complexes are used, and the findings are correlated with the results of the previous studies available in the literature.

The average size of a protein-RNA interface is 2545 Å², which is smaller than an average protein-DNA interface (3137 Å²), but larger than an average protein–protein interface (1919 Å²). This result is consistent with the previous studies carried out with relatively smaller dataset (Bahadur et al., 2008; Jones et al., 2001).

However, the average size (3220 Å²) of the protein-RNA interfaces obtained by Ellis et al. (2007) is relatively large. This is mainly due to the presence of multisubunit assemblies like viruses and ribosomes in their dataset. The average number of amino acids and nucleotides at the protein-RNA interface are 45 and 16, respectively. Each amino acid residue contributes 27 Å² to B, while each nucleotide contributes about 83 Å² to B. On an average, contribution of protein to the protein-RNA interface is 1208 Å², while RNA contributes a little more, 1337 Å². This asymmetry can be attributed to the convex shape of the RNA fitting into the concave protein

### Table 6. Number of stacking interactions.

| Aminoacids | Protein-RNA | Protein-DNA | Protein–protein |
|------------|-------------|-------------|-----------------|
|            | Purine | Pyrimidine | Purine | Pyrimidine | Aminoacids |
|            | A     | G     | C     | U     | A     | G     | C     | T     | Hist | Phe | Trp | Tyr | Arg |
| His        | 12    | 4     | 1     | 7     | 6     | 1     | 5     | 2     | 0    | 1   | 0   | 1   | 1   |
| Phe        | 13    | 9     | 6     | 25    | 6     | 1     | 4     | 5     | 0    | 0   | 1   | 4   | 5   |
| Trp        | 6     | 4     | 2     | 6     | 0     | 2     | 1     | 2     | 0    | 0   | 0   | 0   | 2   |
| Tyr        | 19    | 11    | 3     | 25    | 2     | 2     | 2     | 4     | 3    | 0   | 0   | 2   | 6   |
| Arg        | 38    | 21    | 24    | 19    | 6     | 14    | 16    | 33    | 2    | 2   | 3   | 8   | 3   |

*Calculated on a dataset of 115 protein-DNA complexes taken from (Setny et al., 2012).

*Calculated on a dataset of 160 protein–protein complexes taken from (Hwang et al., 2010).
surface (Bahadur et al., 2008). Among the different classes, an average interface in Class A is the largest (3575 Å²), which can be attributed to the large protein and the RNA molecules involved in complexation.

Despite the wide variations in the size of the protein-RNA interfaces, about 61% of them have B in the range between 1700 and 4800 Å². These interfaces are larger than the “standard size” interface defined by Lo Conte et al. (1999). In the present dataset, about 27% protein-RNA complexes have B below 1700 Å². Here, the interacting RNA is very small, ranging from 3 to 19 nt with an average of 8 nt. About 13% of the complexes have B in the range of 4800–7900 Å². Most of these interfaces belong to Class A with large tRNA molecule.

The interaction between the protein and the RNA subunits may result in a flat or twisted interface. The size and shape of the interacting surfaces are very crucial for specific recognition, and has immense importance in structure-based drug design. Protein-RNA recognition is mainly driven by the shape complementarity and electrostatics. Shape complementarity plays an important role in specific recognition (Janin, Rodier, Chakrabarti, & Bahadur, 2006; Lawrence & Colman, 1993), and determines the atomic packing at the interacting surfaces. Atomic packing at the macromolecular interfaces has been studied using different parameters: $S_c$ index (Lawrence & Colman, 1993), Voronoi polyhedral (Cazals, Proust, Bahadur, & Janin, 2006; Lo Conte et al., 1999; Gerstein, Tsai, & Levitt, 1995), gap volume index (GV) (Jones et al., 2001; Laskowski, 1995), and GD (global density) and LD (local density) indices (Bahadur et al., 2008; Bahadur et al., 2004). Janin and his co-workers (Lo Conte et al., 1999; Janin & Chothia, 1976) used the Voronoi volume and showed that protein–protein interfaces are as tightly packed as the protein interiors. An average protein-RNA interface has GV index of 8.9 Å, which is more than an average protein-DNA interface (6.1 Å) but less than an average protein–protein interface (9.7 Å). This suggests that the protein-RNA interfaces are poorly packed compared to the protein-DNA interfaces. Jones et al. (2001) have also used the GV index and concluded that the protein-RNA interfaces are less tightly packed than the protein-DNA interfaces. Among the different classes, the interfaces with tRNA and duplex RNA are poorly packed compared to those with the single-stranded RNA and ribosomal proteins. We find the average $S_c$ index of a protein-RNA complex is .67. This value is slightly more than the average protein-DNA complex (.65) but less than an average protein–protein complex (.70). The $S_c$ index reflects that the protein and RNA surfaces display shape complementarity during the recognition process. Shape complementarity is a primary feature for complex formation and has been used in many docking algorithms (Lawrence & Colman, 1993).

Bahadur et al. (2004) have used the parameters $f_{bu}$, LD, and GD to discriminate the specific and non-specific protein–protein interfaces. Loosely packed interfaces bury fewer atoms in proportion to their size, and hence their $f_{bu}$ is related to the packing of the interfaces. We find that the protein surface in contact with the RNA in protein-RNA complexes is better packed than that in contact with the DNA in protein-DNA complexes, but loosely packed than that in contact with the protein in protein–protein complexes. Complexes in Class C, comprises duplex RNA, have an average $f_{bu}$ close to that of protein-DNA complexes. This illustrates that the interfaces of Class C complexes resemble protein-DNA interfaces in terms of atomic packing. Proteins in both Class C and Class D have high $f_{bu}$ and LD values and thus are closely packed, resembling protein–protein interfaces. Complexes in Class A, comprises tRNAs, have the lowest $f_{bu}$ and LD values, suggesting a poor atomic packing at their interfaces. The DNA surface in contact with the protein is better packed in protein-DNA complexes than the RNA surface in contact with the protein in protein-RNA complexes. The wide range of conformations exhibited by the RNA (Bon, Vernizzi, Orland, & Zee, 2008; Draper, 1999; Jones et al., 2001; Nagai, 1996) may be responsible for the poor atomic packing of the RNA surface at the protein-RNA interfaces.

Previous studies have reported the abundance of positively charged residues at the protein-nucleic acid binding sites (Allers & Shamoo, 2001; Bahadur et al., 2008; Ellis et al., 2007; Jones et al., 2001; Nadassy et al., 1999). We also find that the positively charged residues contribute about 20% to B, while the negatively charged residues contribute only 4%. Besides, aromatic residues also contribute significantly to the protein–RNA interfaces, a phenomenon also observed in the previous studies (Baker & Grant, 2007; Jones et al., 2001; Jeong et al., 2003). These aromatic residues are mainly involved in the stacking interaction with the nucleotide bases as explained below. A comparison between protein-RNA and protein-DNA interfaces shows that the sugar contributes much more to B in the former, while the phosphate contributes much more to B in the later.

The protein-RNA interfaces comprise the non-polar groups involved in van der Waals and hydrophobic interactions, and the polar groups involved in H-bonds. Both contribute to the stability of the complex. The non-polar groups at the protein side contribute more to B than at the RNA side (Table 3). Similar observations were made in the previous studies with a smaller number of protein-RNA complexes (Kim, Yura, & Go, 2006; Morozova, Allers, Myers, & Shamoo, 2006; Treger & Westhof, 2001). Polar interactions can be quantified in terms of H-bonds. On an average, a protein-RNA interface is stabilized by 19 H-bonds, which is translated into H-bond
density of one per 134 Å² of B. Ellis et al. (2007) also reported similar H-Bond density at the protein-RNA interfaces (one per 126 Å² of B). The contribution of the sugar in H-bond formation makes the striking difference between protein-RNA and protein-DNA interfaces. Ribose sugar in RNA is actively involved in H-bond formation due to the presence of the oxygen atom at the 2′ position of the sugar moiety. The 2′OH acts as both donor and acceptor in H-bonds, and is involved in 22% of all direct protein-RNA H-bonds. Treger and Westhof (2001) have also reported similar contribution of the 2′OH in protein-RNA interfaces. The 2′OH plays a significant role in direct as well as water-mediated H-bonds at the protein-RNA interfaces (Barik & Bahadur, 2014).

The major interacting force that contributes to the stability of a protein-RNA interface is the ionic or electrostatic interaction (Allers & Shamoo, 2001; Iwakiri, Tateishi, Chakraborty, Patil, & Kenmochi, 2012). This may be due to the fact that the RNA, being polyanionic, binds the protein surface, which is rich in positively charged amino acids. In protein-RNA interfaces, the oppositely charged residues and nucleotides often interact to form salt bridges. Both H-bonds and salt bridges play a crucial role in macromolecular folding and binding (Costantini, Colonna, & Facchiano, 2008; Xu et al., 1997). A typical salt bridge provides ~40 kJ/mol. We find an average protein-RNA interface is stabilized by eight salt bridges, whereas an average protein-DNA interface is stabilized by 11 salt bridges. These values are significantly larger than the average number of two salt bridges found in a protein–protein interface.

Unlike protein-DNA interfaces, where 40% of the stacking interactions are mediated by the side chains of aromatic residues, in protein-RNA interfaces they contribute to 60%. Besides aromatic residues, Arg plays an important role in stacking interactions. Side chain of Arg is involved in 40% of all stacking interactions found in protein-RNA and protein–protein interfaces. This contribution increases to 60% in protein-DNA interfaces. Allers and Shamoo (2001) have also reported the significant contribution of Arg in stacking interactions. Among the different bases, purines are involved more in stacking interactions (54%) than pyrimidines (46%) in protein-RNA interfaces. This observation is just the reverse in protein-DNA interfaces, where purines contribute only 35% and pyrimidines contribute 65% to all the stacking interactions.

5. Conclusion

The current study aims to decipher the molecular architecture of the protein-RNA interfaces. Several geometric and physicochemical features of protein-RNA interfaces have been derived from a non-redundant dataset of protein-RNA complexes curated from the PDB. This will contribute to our understanding of the molecular and structural basis of protein-RNA recognition. The knowledge gained in this study will be helpful to design knowledge-based scoring function, which can be used in protein-RNA docking applications as well as prediction of RNA-binding sites for a given protein structure. Moreover, protein-RNA complexes comprise large interfaces, and thus are subject to more selection pressure. Understanding how these interfaces cooperate to stabilize large macromolecular assemblies will definitely be one of the important aspects of computational structural biology.

Acknowledgments

R.P.B. acknowledges the support from FAST TRACK funding from SERB, DST, India. A.B. is a recipient of senior research fellowship from CSIR, India. N.C. is a recipient of senior research fellowship from IIT Kharagpur, India. S.P.P. is a recipient of junior research fellowship from DST, India.

Supplementary material

Supplementary materials dealing with ‘Structural parameters of 152 protein-RNA complexes (classified into four different class based on the type of RNA associated with the protein)’ and ‘List of modified amino acid residues and nucleotide bases present in the dataset of 152 protein-RNA complexes’ is available from the authors directly and can be downloaded free of charge from the author’s server at the URL http://www.facweb.iitkgp.ernet.in/~rbahadur/JBSD.html.

References

Allers, J., & Shamoo, Y. (2001). Structure-based analysis of protein-RNA interactions using the program ENTANGLE. Journal of Molecular Biology, 311, 75–86.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215, 403–410.
Bahadur, R. P., Chakrabarti, P., Rodier, F., & Janin, J. (2004). A dissection of specific and non-specific protein–protein interfaces. Journal of Molecular Biology, 336, 943–955.
Bahadur, R. P., Zacharias, M., & Janin, J. (2008). Dissecting protein–RNA recognition sites. Nucleic Acids Research, 36, 2705–2716.
Baker, C. M., & Grant, G. H. (2007). Role of aromatic amino acids in protein–nucleic acid recognition. Biopolymers, 85, 456–470.
Barik, A., & Bahadur, R. P. (2014). Hydration of protein–RNA recognition sites. Nucleic Acids Research, 42, 10148–10160.
Barik, A., Mishra, A., & Bahadur, R. P. (2012). PRince: A web server for structural and physicochemical analysis of Protein-RNA interface. Nucleic Acids Research, 40, W440–W444.
Barlow, D. J., & Thornton, J. (1983). Ion-pairs in proteins. Journal of Molecular Biology, 168, 867–885.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., ... Bourne, P. E. (2000). The protein data bank. Nucleic Acids Research, 28, 235–242.
Bon, M., Vernizzi, G., Orland, H., & Zee, A. (2008). Topological classification of RNA structures. Journal of Molecular Biology, 379, 900–911.
