Available online at www.sciencedirect.com
SCIENCE DIRECT
polymer
Polymer 47 (2006) 709–721
www.elsevier.com/locate/polymer

Analysis of cation–π interactions to the structural stability of RNA binding proteins

S. Chakkaravarthi [a,*], M. Michael Gromiha [b]

[a] School of Biotechnology and Chemical Engineering, Vellore Institute of Technology, Deemed University, Vellore 632 014, Tamilnadu, India
[b] Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), AIST Tokyo Waterfront Bio-IT Research Building, 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan

Received 3 August 2005; received in revised form 17 November 2005; accepted 17 November 2005
Available online 7 December 2005

Abstract

Cation–π interactions play an important role to the stability of protein structures. In this work, we have analyzed the influence of cation–π interactions in RNA binding proteins. We observed cation–π interactions in 32 out of 51 RNA binding proteins and there is a strong correlation between the number of amino acid residues and number of cation–π interactions. The analysis on the influence of short ($< \pm 3$ residues), medium (± 3 or ± 4 residues) and long range contacts ($> \pm 4$ residues) showed that the cation–π interactions are mainly formed by long-range contacts. The cation–π interaction energy for Arg–Trp is found to be the strongest among all interacting pairs. Analysis on the preferred secondary structural conformation of the residues involved in cation–π interaction indicates that the cationic Lys and Arg prefer to be in α-helices and β-strands, respectively, whereas the aromatic residues prefer to be in strand and coil regions. Most of the cation–π interactions forming residues in RNA binding proteins are conserved among homologous sequences. Further, the cation–π interactions have distinct roles to the stability of RNA binding proteins in addition to other conventional non-covalent interactions. The results observed in the present study will be useful in understanding the contribution of cation–π interactions to the stability of RNA binding proteins.

Keywords: Cation–π interactions; RNA binding proteins; Accessible surface area

1. Introduction

Selective binding of proteins to specific sites on nucleic acids has been a challenging and interesting problem since the earliest days of molecular biology. The first protein-nucleic acid recognition problem to be defined was the enzymatic linking of an amino acid with its correct tRNA [1,2], a process whose specificity was seen as crucial for accurate gene expression. Protein recognition of specific RNA sites was also implicit in early studies of ribosome assembly [3,4]. Since then, the participation of specific protein-RNA complexes in a large number of cellular processes has become evident. RNA structures are flexible molecules that display complex secondary and tertiary structures including short lengths of double helices (A-form), hairpin loops, bulges and pseudo-knots. Proteins tend to interact with the complex secondary structure elements such as stem-loops and bulges in RNA [5]. In addition, non-Watson-Crick base pairing can occur in loop regions of RNA structures and such features can also be preferentially identified by proteins [6]. There are several types of interactions, which give an effect to macromolecular structure and interactions. Ion–ion bonds, hydrogen bonds and hydrophobic interactions are often important for both recognition and binding specificity in protein-DNA/RNA interactions. A growing number of experimental and theoretical studies have emphasized the existence of favorable interactions between positively charged groups and π-aromatic systems [7–9]. Both intermolecular and intramolecular cation–π interactions are recognized to play an important role in the stability of protein-DNA complexes [10]. This type of non-covalent binding force is assumed to be significant in protein structure [11] as well as in biomolecular association processes such as antigen–antibody binding [12,13] and receptor–ligand interaction [14,15]. There are reports of this interaction for their role in the enhancement of stability of thermophilic proteins [16,17], folding of polypeptides [18] and the stability of membrane proteins [19,20]. The stability and specificity of

* Corresponding author. Tel.: +91 416 2202616; fax: +91 416 2243092.
E-mail address: chakkaravarthi77@gmail.com (S. Chakkaravarthi).

0032-3861/$ - see front matter
doi:10.1016/j.polymer.2005.11.059

protein-DNA complexes are also reported on the basis of these cation–π interactions [21,22]. Although the structural studies of protein-RNA complexes are mostly focused on discovering the specific mechanisms of protein-RNA interactions by analyzing intra and inter-molecular interactions in diverse aspects, importance of the cation–π interaction in the structural stability of RNA binding proteins has not yet been elucidated.

In this study we have analyzed the cation–π interaction in 51 RNA binding proteins. The energetic contribution due to cation–π interactions have been brought out for each of the 51 proteins and for all six pairs of residues (Arg–Phe, Arg–Tyr, Arg–Trp, Lys–Phe, Lys–Tyr and Lys–Trp) involved in such interactions. The percentage composition of specific amino acid residues contributing to cation–π interactions was calculated. Further, the characteristic features of residues involved in cation–π interactions have been evaluated in terms of secondary structure, solvent accessibility and sequential separation of residues involved in cation–π interactions. We observed that the cation–π interaction energy for the pairs with Arg is stronger than that with Lys. Sequential separation of cation–π interactions in RNA binding proteins shows that most of the interactions are formed due to long range interactions. Cation–π interaction forming residues Lys and Arg prefer to be in α-helices and β-strands, respectively, whereas aromatic residues prefer β-strands and coil regions. Further, most of the residues contributing for cation–π interactions are not involved in binding with RNA.

2. Materials and methods

2.1. Data set

We have considered a set of 51 RNA-binding proteins from the information available in literature [23] for the present study. This set has been obtained with the following conditions: (i) the three dimensional structures of these proteins have been solved with ≤3.0 Å resolution, (ii) the similarity search using PSI-BLAST yielded the *e*-value of less than 0.001 and (iii) the sequence identity is less than 80%. The complexes, whose proteins were homologous but recognized different nucleotide sequences, were included in the data set.

The PDB [24] codes of the proteins are: 1b23, 1b2m, 1b7f, 1c0a, 1c9s, 1cx0, 1dfu, 1di2, 1dk1, 1e7x, 1ec6, 1efw, 1f7u, 1f8v, 1feu, 1ffy, 1fxl, 1g59, 1gax, 1gtf, 1gtn, 1g2e, 1h4q, 1h4s, 1hc8, 1hdw, 1he0, 1he6, 1hq1, 1i6u, 1il2, 1jbr, 1jbs, 1jid, 1k8w, 1knz, 1kq2, 1l9a, 1lng, 1mms, 1qf6, 1qtq, 1ser, 1urn, 1zdh, 1zdi, 2bbv, 2fmt, 5msf, 6msf and 7msf.

2.2. Computation of amino acid composition

The amino acid composition for each amino acid residue that are involved in cation–π interactions (Lys, Arg, Phe, Trp and Tyr) was computed using the standard formula,

$$\mathrm{comp(i)} = \frac{n(\mathrm{i})}{N} \tag{1}$$

where n(i) is the number of amino acids of type i and N is the total number of amino acids in a protein.

2.3. Occurrence and energetic contribution due to cation–π interactions

The number of cation–π interaction in each protein has been calculated using the program CAPTURE [25] available at http://capture.caltech.edu. In the present study only energetically significant interactions ($E_{\mathrm{cat-}\pi} \leq 2$ kcal/mol) were considered. The percentage composition of a specific amino acid residue contributing to cation–π interactions is obtained by the equation,

$$\mathrm{comp}_{\mathrm{cat-}\pi}(\mathrm{i}) = n_{\mathrm{cat-}\pi}(\mathrm{i}) \times \frac{100}{n(\mathrm{i})} \tag{2}$$

where i stands for the five residues, Lys, Arg, Phe, Trp and Tyr, $n_{\mathrm{cat-}\pi}$ is the number of residues involved in cation–π interactions and n(i) is the number of residues of type i in the considered protein structures.

We have computed the energetic contribution of cation–π interactions for each RNA binding protein in the data set and for all possible pairs of positively charged and aromatic amino acids. The total cation–π interaction energy ($E_{\mathrm{cat-}\pi}$) has been divided into electrostatic (E_{es}) and van der Waals energy (E_{vw}) and were computed using the program CAPTURE, which has implemented a subset of OPLS force field [26] to calculate the energies. The electrostatic energy (E_{es}) is calculated using the equation

$$E_{\mathrm{es}} = \frac{q_{\mathrm{i}} q_{\mathrm{j}} e^2}{r_{\mathrm{ij}}} \tag{3}$$

where q_{i} and q_{j} are the charges for the atoms i and j, respectively, and r_{ij} is the distance between them. The van der

Table 1
Composition of aromatic and positively charged residues in RNA binding proteins

Proteins	Lys%	Arg%	Phe%	Tyr%	Trp%
RNA binding protein	7.30±3.70	5.98±2.65	3.79±1.49	3.27±1.55	1.03±0.90
TMH	2.29±1.68	2.97±0.95	7.98±1.67	4.14±0.79	4.19±1.38
TMS	4.73±1.75	3.48±0.85	4.40±1.36	6.56±1.93	1.85±1.33
Globular	5.83	4.74	3.97	3.60	1.48

TMS, transmembrane strand; TMH, transmembrane helical.

Table 2
Cation–π interaction energetic contribution in RNA binding proteins

PDB code	$N_{cat-\pi}$	$-E_{es}$	$-E_{vw}$	$-E_{cat-\pi}$
1b23A	3	6.2	5.31	11.51
1b2mP	1	1.42	1.60	3.02
1b7fA	2	6.96	2.31	9.27
1c0aA	7	19.38	11.93	31.31
1c9sA	0	0.00	0.00	0.00
1cx0A	1	2.71	3.26	5.97
1dfuP	0	0.00	0.00	0.00
1di2A	1	3.83	1.16	4.99
1dk1A	1	4.66	1.18	5.84
1e7xA	0	0.00	0.00	0.00
1ec6A	0	0.00	0.00	0.00
1efwA	9	19.44	15.3	34.74
1f7uA	7	21.98	11.58	33.56
1f8vA	4	7.96	7.35	15.31
1feuA	3	10.35	4.74	15.09
1ffyA	16	53.8	26.24	80.04
1fxlA	0	0.00	0.00	0.00
1g2eA	0	0.00	0.00	0.00
1g59A	7	26.52	10.55	37.07
1gaxA	21	65.93	29.98	95.91
1gtfA	0	0.00	0.00	0.00
1gtnA	0	0.00	0.00	0.00
1h4qA	11	35.64	21.29	56.93
1h4sA	12	36.11	23.46	59.57
1hc8A	0	0.00	0.00	0.00
1hdwA	0	0.00	0.00	0.00
1he0A	0	0.00	0.00	0.00
1he6A	0	0.00	0.00	0.00
1hq1A	0	0.00	0.00	0.00
1i6uA	1	2.43	0.67	3.1
1il2A	8	18.09	13.41	31.5
1jbrA	1	6.19	4.5	10.69
1jbsA	2	6.04	6.82	12.86
1jidA	2	4.77	4.5	9.27
1k8wA	1	1.41	2.49	3.9
1knzA	2	5.27	3.66	8.93
1kq2A	0	0.00	0.00	0.00
1l9aA	3	6.59	3.59	10.18
1lngA	1	2.32	0.76	3.08
1mmsA	1	3.32	1.53	4.85
1qf6A	18	53.75	33.18	86.93
1qtqA	7	30.79	14.54	45.33
1serA	9	25.88	21.89	47.77
1urnA	1	1.34	1.16	2.5
1zdhA	0	0.00	0.00	0.00
1zdiA	0	0.00	0.00	0.00
2bbvA	4	7.18	5.55	12.73
2fmtA	5	15.72	9.39	25.11
5msfA	0	0.00	0.00	0.00
6msfA	0	0.00	0.00	0.00
7msfA	0	0.00	0.00	0.00
Average	3.37±4.98	10.07±15.62	5.98±8.64	16.06±24.08

$N_{cat-\pi}$, number of cation–π interactions in a protein. E_{es}, E_{vw}, $E_{cat-\pi}$ are, respectively, electrostatic, van der Waals and total cation–π interaction energy.

Waals energy is given by

$$E_{vw} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}}\right) - \left(\frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right)\right] \quad (4)$$

where $\sigma_{ij}=(\sigma_{ii}\sigma_{jj})^{1/2}$ and $\varepsilon_{ij}=(\varepsilon_{ii}\varepsilon_{jj})^{1/2}$; σ and ε are the van der Waals radius and well depth, respectively.

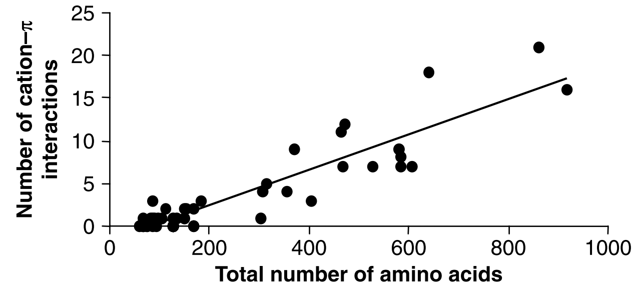

Fig. 1. Relationship between the total number of amino acid residues and number of cation–π interactions in RNA binding proteins (coefficient of correlation = 0.91).

2.4. Location of cation–π interaction forming residues based on secondary structure and solvent accessibility

Secondary structure and solvent accessibility are the two major intermediate steps to understand the structure and function of proteins. We have systematically analyzed the preference for each of the cation–π interaction forming residues based on their location in different secondary structures of RNA binding proteins and their solvent accessibility. We have used the program DSSP to obtain the information about secondary structure and solvent accessibility [27]. The secondary structures have been classified into helix, strand, turn and coil as suggested by Heringa and Argos [28]. Solvent accessibility was divided into three classes viz. 0–20%, 20–50%, and >50% indicating the least, moderate and high accessibility of the amino acid residues, respectively.

2.5. Classification by residue-residue contacts

The amino acid residues involved in the cation–π interactions were classified as short (< ±3 residues), medium (±3 or ±4 residues) and long range (> ±4 residues) based on their location in the amino acid sequence. [29,30]. This classification enabled us to evaluate the contribution of long-range contacts in the formation of cation–π interactions.

2.6. Conservation of amino acid residues

We have evaluated the conservation of residues in each protein with the aid of the Consurf server [31] (http://consurf.tau.ac.il/). This server compares the sequence of a PDB chain with the proteins deposited in Swiss–Prot [32] and identifies the sequences that are homologous to the PDB sequence. These protein sequence alignments were used to classify the residues in each RNA binding protein into nine categories: from highly variable (score = 1) to highly conserved (score = 9).

2.7. Identification of stabilizing residues

We have identified stabilizing residues in each protein using the SRide server, which is available at http://sride.enzim.hu [33,34]. This server computes the different measures of

Table 3
Cation–π interaction energy in RNA binding proteins

PDB code	R-F (−kcal/mole)	R-Y (−kcal/mole)	R-W (−kcal/mole)	K-F (−kcal/mole)	K-Y (−kcal/mole)	K-W (−kcal/mole)
1b23A		R330-Y338 (2.97); R389-Y343 (2.51)	R190-W200 (6.03)			
1b2mP		R77-Y38 (3.02)				
1b7f A					K161-Y160 (6.17); K233-Y234 (3.10)	
1c0aA	R76-F48 (3.40); R208-F157 (4.12)	R2-Y5 (7.40); R245-Y474 (3.02)	R39-W23 (3.21); R245-W429 (6.43)	K412-F340 (3.73)		
1cx0A	R36-F37 (5.97)					
1di2A					K167-Y131 (4.99)	
1dk1A				K183-F114 (5.84)		
1efwA	R214-F163 (3.68); R371-F359 (2.69); R404-F295 (4.49)	R3-Y6 (4.65); R42-Y6 (5.01)	R40-W24 (2.78); R138-W114 (5.05); R353-W351 (3.32)		K129-Y130 (3.07)	
1f7uA		R226-Y206 (3.95); R254-Y251 (4.49); R358-Y118 (3.07); R477-Y471 (5.08); R495-Y565 (7.60)		K102-F113 (5.80)	K156-Y224 (3.57)	
1f8vA	R185-F242 (4.11)	R312-Y169 (4.81)	R160-W252 (5.89)	K61-F69 (3.11)		
1feuA	R72-F89 (3.39); R103-136 (4.32)	R35-Y9 (7.38)				
1ffyA	R112-F120 (3.45); R121-F496 (4.76); R649-F14 (5.45)	R440-Y559 (4.29)	R79-W31 (3.85); R121-W459 (2.90); R399-W398 (10.03); R407-W185 (7.2); R448-W451 (6.02)	K71-F139 (5.93); K81-F50 (5.08); K81-F86 (3.97); K797-F846 (4.24)	K203-Y394 (3.71)	K136-W94 (3.61); K823-W890 (5.55)
1g59A	R137-F106 (6.33)	R45-Y184 (3.36); R147-Y122 (6.77)		K423-F419 (11.84)	K91-Y92 (2.81)	K309-W312 (2.42); K456-W407 (3.54)
1gaxA	R68-F25 (4.48); R102-F110 (4.18); R201-F209 (2.80); R314-F315 (2.88); R318-F315 (2.38)	R168-Y416 (3.85); R635-Y557 (7.74)	R65-W31 (6.63); R102-W459 (4.48); R149-W398 (2.68); R171-W185 (5.44); R448-W415 (3.47); R498-W400 (10.33); R730-W648 (3.86)	K67-F72 (5.58); K118-F143 (5.86); K654-F558 (2.52); K658-F764 (3.64)		K19-W16 (3.21); K130-W138 (6.77); K723-W648 (4.33)
1h4qA	R176-F449 (8.38); R347-F336 (8.63); R470-F425 (5.30)	R301-Y296 (2.32)	K142-W158 (3.90); K247-W127 (10.05)		K122-Y118 (3.73); K222-Y477 (3.70); K243-Y253 (2.08); K342-Y42 (4.84)	K342-W339 (3.22)
1h4sA	R176-F449 (8.60); R347-F336 (7.70); R470-F425 (5.09)	R301-Y296 (3.27)	K34-W143 (3.45); K142-W158 (3.31); K247-W127 (10.07)		K122-Y118 (3.36); K222-Y477 (3.70); K243-Y253 (2.23); K342-Y42 (5.55)	K342-W339 (3.24)
1i6uA				K83-F84 (3.01)		
1il2A	R76-F48 (3.69); R208-F157 (4.39)	R2-Y5 (6.12); R41-Y5 (3.50); R245-Y474 (3.13)	R39-W23 (2.79); R245-W429 (4.66)	K412-F340 (3.22)		
1jbrA			R138-W17 (10.69)			
1jbsA		R120-Y47 (3.81)	R138-W17 (9.05)			
1jidA		R34-Y22 (5.94); R81-Y19 (3.33)				
1k8wA		R141-Y137 (3.90)				
1knzA			R93-W87 (5.13)	K47-F19 (3.80)		
1l92A		R63-Y48 (2.57)	R63-W4 (4.78)		K19-Y7 (2.83)	
1lngA					K19-Y7 (3.08)	
1mmsA	R41-F66 (4.85)					
1qf6A	R191-F192 (3.89); R589-F532 (4.81); R612-F532 (5.63)	R217-Y103 (9.01); R217-Y219 (2.68); R325-Y327 (3.06); R354-Y290 (3.61)	R72-W223 (7.25); R145-W141 (4.70); R207-W206 (4.99); R235-W223 (3.14); R301-W310 (7.42); R423-W434 (3.26); R427-W434 (3.54); R635-W536 (7.85)	K346-F341 (3.15)	K200-Y219 (2.44); K415-Y471 (6.58)	

Table 3 (*continued*)

PDB code	R-F (−kcal/mole)	R-Y (−kcal/mole)	R-W (−kcal/mole)	K-F (−kcal/mole)	K-Y (−kcal/mole)	K-W (−kcal/mole)
1qtqA	R421-F434 (8.58)	R474-Y265 (9.34)	R297-W87 (7.63)	K159-F165 (5.98); K272-F487 (3.16)	K141-Y132 (6.08)	K350-W386 (4.56)
1serA	R209-F205 (6.83); R247-F185 (3.63); R256-F318 (5.21); R329-F295 (3.45)	R314-Y343 (4.55); R358-Y373 (4.26); R363-Y373 (3.33)	R329-W106 (7.68); R359-W355 (8.91)			
1urnA	R36-F37 (2.5)					
2bbvA	R167-F252 (4.85); R300-F112 (2.19)			K68-F76 (3.28)	K91-Y330 (2.41)	
2fmtA		R118-Y203 (2.54)	R116-W117 (4.70); R125-W117 (2.33); R125-W128 (8.49); R213-W237 (7.05)			

The fifth letter of the PDB code indicates the chain. Cation–π interactions are observed in 32 (out of 51) RNA binding proteins.

stability such as surrounding hydrophobicity (H_p), long range order (LRO), stabilization center (SC) and conservation of residues. The stabilization residues in RNA binding proteins have been delineated with certain cutoff values for each term (i.e. the stabilizing residues is the one in which the values for all these four parameters are equal to or greater than the specified cutoff values). In this study, we have used the following conditions to predict the stabilizing residues: (i) $H_p \geq 20$ kcal/mol; (ii) LRO ≥ 0.02; (iii) SC ≥ 1; and (iv) conservation score ≥ 6.

2.8. Identification of binding residues in protein-RNA complexes

We have identified the amino acid residues that are in contact with RNA (backbones and bases) using the information available in amino acid–nucleotide interaction database (AANT) [http://aant.icmb.utexas.edu/global/complexes.html] [35]. AANT uses the program HBPLUS to compute the hydrogen bond interactions between the amino acids and nucleotides, and assigns the interacting residues. We have considered the cation–π interaction forming residues, Lys, Arg, Phe, Trp and Tyr to understand the influence of these residues to form cation–π interactions and binding with RNA.

3. Results and discussion

3.1. Composition of aromatic and positively charged amino acids in RNA binding proteins

The composition of amino acid residues that are involved in cation–π interactions was analysed and the results for RNA binding proteins along with other classes of proteins are presented in Table 1. We observed that in RNA binding proteins, Phe has the highest occurrence among the aromatic residues, which is similar to transmembrane helical (TMH) [20] and globular proteins [25]. Further, the lowest occurrence of Trp is similar to transmembrane strand (TMS) [20] and globular proteins. As observed in globular proteins the number of Lys is higher than Arg in RNA binding proteins [25]. Generally the composition of cation–π interaction forming residues is similar to other globular proteins (Table 1).

3.2. Relationship between number of amino acid residues and number of cation–π interactions

The number of cation–π interactions in each of the RNA binding proteins and their energetic contributions are presented in Table 2. We observed an average of 3.4 cation–π interactions in RNA binding proteins, which is considerably less than that of DNA binding proteins [22]. However, when we considered only the proteins that have cation–π interactions we have noticed an average of 5.4 cation–π interactions. The number of cation–π interactions varies for different proteins; it is zero in the A chain of 1c9s (and 18 other complexes) and 21 in A chain of 1gax. Although the protein length is similar in A chain of 1l9a and 1lng, the number of cation–π interactions varies to 3 and 1, respectively. Further, we observed a strong positive correlation between the number of residues and number of cation–π interactions as shown in Fig. 1, which is

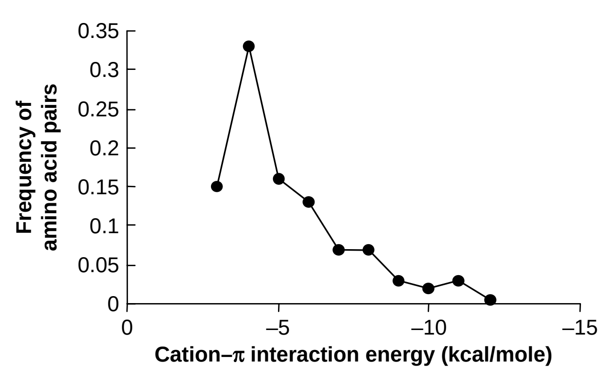

Fig. 2. Frequency of amino acid pairs at different ranges of cation–π interaction energy in RNA binding proteins.

Table 4
Comparison of RNA binding proteins average energy contribution for each amino acid pair involved in cation–π interaction with DNA binding and membrane proteins

Amino acid pair	$-E_{es}$ (kcal/mole)				$-E_{vw}$ (kcal/mole)				$-E_{cat-\pi}$ (kcal/mole)			
	RNABP	DNABP	TMH	TMS	RNABP	DNABP	TMH	TMS	RNABP	DNABP	TMH	TMS
Arg-Phe	6.06±3.26	2.52	3.90	2.70	4.65±3.06	2.47	2.70	2.15	10.71±6.14	4.99	6.60	4.85
Arg-Tyr	4.14±3.33	2.88	3.68	2.56	4.55±3.33	2.47	2.80	2.32	8.69±5.82	5.35	6.48	4.88
Arg-Trp	9.27±7.54	2.63	4.96	5.13	5.81±4.20	2.05	2.59	2.86	15.08±11.40	4.68	7.55	7.99
Lys-Phe	5.68±4.69	3.00	3.59	3.00	1.48±1.09	0.90	2.92	1.15	7.16±5.72	3.09	6.51	4.15
Lys-Tyr	4.63±3.13	3.63	2.32	2.54	1.55±1.37	1.02	1.08	0.91	6.18±4.39	4.65	3.40	3.45
Lys-Trp	5.54±3.52	3.54	2.90	5.39	1.20±0.85	0.76	0.71	0.20	6.74±4.32	4.30	3.61	6.59

RNABP, RNA binding protein; DNABP, DNA binding protein; TMH, transmembrane helix; TMS, transmembrane strand.

similar to transmembrane strand proteins. The correlation coefficient is 0.91.

3.3. Energetic contribution of cation–π interactions in RNA binding proteins

The strength of cation–π interaction energy differs significantly in RNA binding proteins, it is −5.97 kcal/mol for the A chain of 1cx0 and −2.5 kcal/mol for the A chain of 1urn, each having a single cation–π interaction. However, we found positive correlation between the number of cation–π interactions and their energies ($r=0.99$). The composition of cation–π interaction energy into electrostatic and van der Waals energy terms showed that among the 32 out of 51 RNA binding proteins that have cation–π interactions, 28 have stronger electrostatic energy than van der Waals energy and an opposite trend is observed for 4 proteins.

The energetic contribution of each cationic-aromatic pairs of amino acids in RNA binding proteins has been computed and the results are presented in Table 3. The number of residues involved in cation–π interactions is, 51, 121, 58, 59 and 55 for Lys, Arg, Phe, Tyr and Trp, respectively. We found that 62.75% of the RNA binding proteins (32/51) form one or more cation–π interactions and few residues form cation–π interactions with several other residues (e.g. R245 in 1c0a, F315 in 1gax, W117 in 2fmt, etc). The strongest contribution is observed for the interaction between Arg138 and Trp17 in the A chain of 1jbr and the cation–π interaction energy is −10.69 kcal/mol, which is marginally stronger than that observed in DNA binding proteins [22].

Further, in globular and DNA binding proteins, there is an average of one energetically significant cation–π interaction for every 77 and 81 residues, respectively [25,22]. In RNA binding proteins, we have identified 172 cation–π interactions among 12,655 amino acid residues, indicating the presence of one cation–π interaction for every 74 residues. In RNA binding proteins 48% of the cation–π interactions have the energy less than −4 kcal/mol, whereas about 25, 55 and 65% of these interactions have similar energy in globular [25], DNA binding [22] and membrane proteins [20], respectively.

The frequency of cation–π interaction pairs at different intervals of energy is plotted in Fig. 2. We observed that most of the cation–π interactions have the energy in the range of −3 to −4 kcal/mol.

3.4. Average contribution of cation–π interaction energy for different cation–π pairs

We have calculated the average cation–π interaction energy for all the six possible pairs between cationic and aromatic residues in RNA binding proteins and the results are compared with DNA binding and membrane proteins (Table 4). We observed that in RNA binding proteins, Arg–Trp pair has the strongest contribution among all pairs, similar to transmembrane helical and strand proteins [20]. In DNA binding proteins, Arg-Tyr has the strongest cation–π interaction energy. Both DNA and RNA binding proteins have the van der Waals energy more than three times stronger than the electrostatic energy for the interacting pairs containing Lys [21,22]. The comparison on the strength of cation–π interaction energy of each residue pair in different types of proteins showed that the membrane, DNA and RNA binding proteins have stronger cation–π interaction energy for the pairs with Arg than that with Lys. The average cation–π interaction energy of Arg and Lys in RNA binding proteins is −11.49 and −6.69, respectively, which is stronger than that observed in other classes of proteins. This result indicates that the cation–π interaction play an important role to the stability of RNA binding proteins.

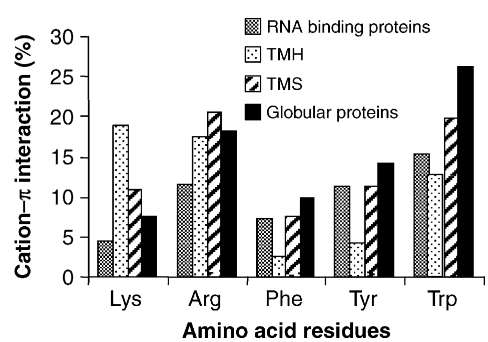

Fig. 3. Percentage of aromatic and positively charged residues contributing towards cation–π interactions in RNA binding, TMH, TMS and globular proteins.

Table 5
Secondary structure, solvent accessibility, conservation score and sequence separation of cation–π interaction forming residues

PDB code	Cation	Residue	Str	ASA	Cons	π	Residue	Str	ASA	Cons	D_{seq}
1b23A	Arg	330	C	39	9	Tyr	338	C	38	9	8
	Arg	389	S	19	9	Tyr	343	S	25	8	46
	Arg	190	T	119	3	Trp	200	H	14	1	10
1b2mP	Arg	77	S	8	9	Tyr	38	S	13	9	39
1b7fA	Lys	161	T	178	1	Tyr	160	T	173	1	1
	Lys	233	C	135	3	Tyr	234	C	64	4	1
1c0aA	Arg	76	S	92	4	Phe	48	S	19	7	28
	Arg	208	S	68	8	Phe	157	C	4	9	51
	Arg	2	C	83	8	Tyr	5	C	66	7	3
	Arg	245	H	29	5	Tyr	474	S	0	9	229
	Arg	39	S	15	9	Trp	23	S	81	9	16
	Arg	245	H	29	5	Trp	429	S	2	9	184
	Lys	412	H	70	8	Phe	340	H	32	7	72
1cx0A	Arg	36	H	165	6	Phe	37	H	25	2	1
1di2A	Lys	167	H	80	9	Tyr	131	S	64	9	36
1dk1A	Lys	183	T	111	3	Phe	114	H	20	4	69
1efwA	Arg	214	S	28	7	Phe	163	C	2	7	51
	Arg	371	H	79	1	Phe	359	S	21	2	12
	Arg	404	H	3	9	Phe	295	C	18	7	109
	Arg	3	C	26	9	Tyr	6	C	44	4	2
	Arg	42	T	23	9	Tyr	6	C	44	4	36
	Arg	40	S	4	9	Trp	24	S	5	9	16
	Arg	138	H	91	3	Trp	114	H	10	1	24
	Arg	353	S	68	8	Trp	351	S	68	5	2
	Lys	129	H	80	7	Tyr	130	T	37	7	1
1f7uA	Arg	226	H	91	5	Tyr	206	H	87	3	20
	Arg	254	H	49	6	Tyr	251	H	41	3	3
	Arg	358	H	9	9	Tyr	188	S	2	9	170
	Arg	477	H	89	6	Tyr	471	S	22	8	6
	Arg	495	H	57	9	Tyr	565	H	80	9	70
	Lys	102	S	72	7	Phe	113	S	15	5	11
	Lys	156	C	46	9	Tyr	244	C	14	7	88
1f8vA	Arg	185	S	19	8	Phe	242	S	2	7	57
	Arg	312	S	79	8	Tyr	169	S	57	9	143
	Arg	252	C	16	7	Trp	252	C	16	7	92
	Lys	61	H	48	9	Phe	69	C	60	8	8
1feuA	Arg	72	S	94	8	Phe	89	S	38	4	17
	Arg	103	S	138	5	Phe	136	C	86	2	33
	Arg	35	S	111	7	Tyr	9	S	89	9	26
1ffyA	Arg	112	C	80	6	Phe	120	H	2	7	8
	Arg	121	H	1	9	Phe	496	H	42	7	375
	Arg	649	H	0	8	Phe	14	C	1	9	634
	Arg	440	H	110	9	Tyr	559	H	5	8	119
	Arg	79	H	2	9	Trp	31	H	16	9	48
	Arg	121	H	1	9	Trp	459	C	2	9	338
	Arg	399	T	178	9	Trp	398	T	72	9	1
	Arg	407	S	34	8	Trp	185	S	27	8	222
	Arg	448	C	37	8	Trp	451	S	11	9	3
	Lys	71	H	2	8	Phe	139	H	1	7	68
	Lys	81	H	21	7	Phe	50	C	17	6	31
	Lys	81	H	21	7	Phe	86	C	44	1	5
	Lys	797	H	48	1	Phe	846	C	17	5	49
	Lys	203	S	28	5	Tyr	394	S	29	9	191
	Lys	136	H	18	5	Trp	94	S	0	8	42
	Lys	823	C	89	6	Trp	890	C	125	7	67
1g59A	Arg	137	H	57	3	Phe	106	C	24	7	31
	Arg	45	T	158	3	Tyr	184	C	59	2	139
	Arg	147	S	18	9	Tyr	122	C	44	9	25
	Lys	423	T	53	1	Phe	419	H	3	1	4
	Lys	91	H	120	1	Tyr	92	H	30	5	1
	Lys	309	H	64	9	Trp	312	H	94	6	3
	Lys	456	H	56	6	Trp	407	C	41	4	49

(*continued on next page*)

Table 5 (*continued*)

PDB code	Cation	Residue	Str	ASA	Cons	π	Residue	Str	ASA	Cons	D_{seq}
1gaxA	Arg	68	H	30	9	Phe	25	C	26	9	43
	Arg	102	C	26	8	Phe	110	H	1	9	8
	Arg	201	S	61	5	Phe	209	C	97	1	8
	Arg	314	C	19	8	Phe	315	C	72	4	1
	Arg	318	H	57	9	Phe	315	C	72	4	3
	Arg	168	S	86	8	Tyr	416	S	40	5	248
	Arg	635	H	74	3	Tyr	557	H	33	6	78
	Arg	65	H	0	9	Trp	20	H	9	9	45
	Arg	102	C	26	8	Trp	407	C	60	6	305
	Arg	149	H	13	6	Trp	462	C	1	7	317
	Arg	171	C	62	7	Trp	360	S	13	8	33
	Arg	448	S	76	8	Trp	415	S	39	8	98
	Arg	498	H	11	9	Trp	400	C	19	9	82
	Arg	730	H	99	9	Trp	648	T	29	7	5
	Lys	67	H	24	6	Phe	72	C	58	3	25
	Lys	118	H	20	8	Phe	143	S	11	7	66
	Lys	654	H	53	9	Phe	588	H	48	7	66
	Lys	658	H	147	2	Phe	764	H	55	1	106
	Lys	19	H	73	1	Trp	16	H	8	1	3
	Lys	130	H	56	5	Trp	138	H	25	7	8
	Lys	723	H	27	3	Trp	648	T	29	7	75
1h4qA	Arg	176	H	98	2	Phe	449	T	45	1	273
	Arg	347	S	20	7	Phe	336	H	38	8	11
	Arg	470	C	19	1	Phe	425	C	14	1	45
	Arg	301	H	88	1	Tyr	296	C	38	6	5
	Arg	142	C	73	9	Trp	158	S	24	8	16
	Arg	247	T	165	7	Trp	127	H	57	7	120
	Lys	122	H	88	6	Tyr	118	H	74	7	4
	Lys	222	C	67	7	Tyr	477	C	34	9	255
	Lys	243	S	85	6	Tyr	253	S	47	3	10
	Lys	342	T	14	6	Tyr	42	H	24	6	300
	Lys	342	T	14	6	Trp	339	H	14	1	3
1h4sA	Arg	176	H	96	2	Phe	449	C	41	1	273
	Arg	347	S	21	7	Phe	336	H	34	8	11
	Arg	470	C	17	1	Phe	425	C	14	1	45
	Arg	301	H	86	1	Tyr	296	C	48	2	5
	Arg	34	T	106	7	Trp	143	C	26	8	109
	Arg	142	C	11	9	Trp	158	S	1	7	16
	Arg	247	T	172	7	Trp	127	H	56	7	120
	Lys	122	H	96	6	Tyr	118	H	65	9	4
	Lys	222	C	59	7	Tyr	477	C	29	3	255
	Lys	243	S	82	6	Tyr	253	S	43	6	10
	Lys	342	S	14	6	Tyr	42	H	23	6	300
	Lys	342	S	14	6	Trp	339	H	16	6	3
1i6uA	Lys	83	T	98	1	Phe	84	T	175	1	1
1il2A	Arg	76	S	78	4	Phe	48	S	20	7	28
	Arg	208	S	64	8	Phe	157	C	3	9	51
	Arg	2	C	36	9	Tyr	5	C	35	7	3
	Arg	41	T	49	8	Tyr	5	C	35	7	36
	Arg	245	H	32	5	Tyr	474	S	0	9	229
	Arg	39	S	22	9	Trp	23	S	4	9	16
	Arg	245	H	32	5	Trp	429	S	2	9	184
	Lys	412	H	54	8	Phe	340	H	37	7	72
1jbrA	Arg	138	S	112	NA	Trp	17	S	103	NA	121
1jbsA	Arg	120	S	6	NA	Tyr	47	C	29	NA	73
	Arg	138	S	125	NA	Trp	17	S	91	NA	121
1jidA	Arg	34	C	99	8	Tyr	22	H	28	9	12
	Arg	81	S	38	9	Tyr	19	C	31	9	62
1k8wA	Arg	141	H	163	9	Tyr	137	H	68	9	4
1knzA	Arg	93	H	22	1	Trp	87	H	80	9	6
	Lys	47	H	68	3	Phe	19	H	7	9	28
1l9aA	Arg	63	C	47	9	Tyr	48	C	93	8	15
	Arg	63	C	47	9	Trp	4	C	28	9	59
	Lys	19	C	84	7	Tyr	7	H	23	8	12
1lngA	Lys	19	C	75	7	Tyr	7	H	21	8	12

Table 5 (*continued*)

PDB code	Cation	Residue	Str	ASA	Cons	π	Residue	Str	ASA	Cons	D_{seq}
1mmsA	Arg	41	H	69	5	Phe	66	S	2	8	25
1qf6A	Arg	191	H	79	4	Phe	192	H	31	2	1
	Arg	589	T	65	7	Phe	532	C	37	4	57
	Arg	612	C	149	2	Phe	532	C	37	4	80
	Arg	217	S	47	9	Tyr	103	S	26	9	114
	Arg	217	S	47	9	Tyr	219	S	62	9	2
	Arg	325	S	117	2	Tyr	327	S	55	5	2
	Arg	354	S	56	9	Tyr	290	C	29	9	64
	Arg	72	H	16	8	Trp	223	C	24	6	151
	Arg	145	H	77	1	Trp	141	H	22	5	4
	Arg	207	H	94	8	Trp	206	H	30	9	1
	Arg	235	H	117	5	Trp	223	C	24	6	12
	Arg	301	T	61	2	Trp	310	H	30	5	9
	Arg	423	C	47	9	Trp	434	H	7	9	11
	Arg	427	C	80	7	Trp	434	H	7	9	7
	Arg	635	T	54	5	Trp	536	T	5	9	99
	Lys	346	S	33	5	Phe	341	H	2	7	5
	Lys	200	S	138	6	Tyr	219	S	62	9	19
	Lys	415	C	97	1	Tyr	471	S	57	2	56
1qtqA	Arg	421	S	113	3	Phe	434	S	24	1	13
	Arg	474	C	90	6	Tyr	265	T	41	4	209
	Arg	297	H	61	9	Trp	87	H	70	7	210
	Lys	159	H	85	6	Phe	165	C	27	3	6
	Lys	272	H	50	9	Phe	487	H	8	6	215
	Lys	141	C	117	3	Tyr	132	H	52	2	9
	Lys	350	S	60	5	Trp	386	S	41	2	36
1serA	Arg	209	H	112	5	Phe	205	T	0	7	4
	Arg	247	S	84	3	Phe	185	C	3	9	62
	Arg	256	C	52	9	Phe	275	S	12	9	19
	Arg	329	S	53	7	Phe	295	H	6	7	34
	Arg	314	S	51	8	Tyr	343	S	25	7	29
	Arg	358	H	62	9	Tyr	373	S	75	4	15
	Arg	363	S	54	9	Tyr	373	S	75	4	10
	Arg	329	S	53	7	Trp	106	C	64	3	223
	Arg	359	H	39	9	Trp	355	T	10	8	4
1urnA	Arg	36	H	81	6	Phe	37	H	14	4	1
2bbvA	Arg	167	S	3	9	Phe	252	C	11	7	85
	Arg	300	S	39	5	Phe	112	S	0	3	188
	Lys	68	H	45	9	Phe	76	C	52	8	8
	Lys	91	S	39	6	Tyr	330	H	13	8	239
2fmtA	Arg	118	S	11	9	Tyr	203	S	124	7	95
	Arg	116	T	86	8	Trp	117	T	21	6	1
	Arg	125	H	7	8	Trp	117	T	21	6	8
	Arg	125	H	7	8	Trp	128	H	59	3	3
	Arg	213	S	36	2	Trp	237	S	38	5	24

Str, secondary structure; H, helix; S, strand; T, turn; C, coil; ASA, accessible surface area or solvent accessibility. The values are in $\AA^2$. Cons, conservation score; D_{seq}, sequence distance of separation between cationic and aromatic residues; NA, not available.

3.5. Relative contribution of amino acids involved in cation–π interactions

We have estimated the percentage of aromatic and positively charged amino acids that are involved in cation–π interactions in RNA binding protein structures. The relative contribution of each of the five amino acid residues in RNA binding proteins along with TMH, TMS and globular proteins is depicted in Fig. 3. We found that the contribution of aromatic residues in RNA binding proteins is (Phe 7.27%, Tyr 11.32% and Trp 15.40%) similar to TMS proteins towards cation–π interactions. Further, the contribution of positively charged residue, Arg is higher (11.57%) than that of Lys (4.48%) in RNA binding proteins, similar to the trend observed in transmembrane strand [20] and globular proteins [25]. In transmembrane helical proteins, both Lys and Arg have approximately equal preference to form cation–π interactions.

3.6. Sequential separation and conservation score

We have calculated the sequential distance between the cationic and aromatic residues for each of the cation–π interactions and the results are presented in Table 5. We found that in RNA binding proteins 9, 8 and 83% of cation–π interactions are influenced by short, medium and long range

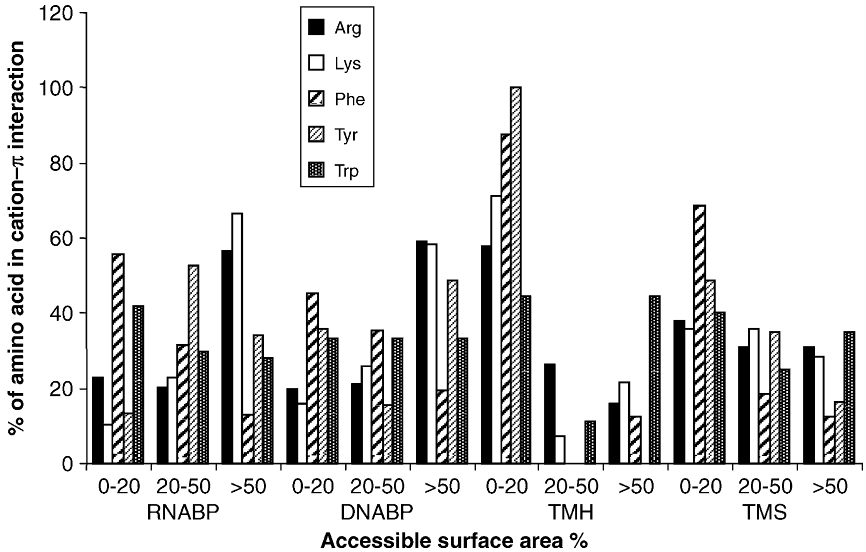

Fig. 4. Comparison of cation–π interaction forming residues in different ranges of ASA for RNA binding (RNABP), DNA binding (DNABP), transmembrane helical (TMH) and strand (TMS) proteins.

interactions. This result revealed that majority of the cation–π interactions in RNA binding proteins are influenced by long range interactions as observed in DNA binding proteins [22]. This result reflects the importance of long range interactions to the stability of all classes of proteins [29].

In Table 5, we have also included the conservation score for all cation–π interactions forming residues in RNA binding proteins. Conservation score calculation needs at least five homologous sequences [31] and hence the conservation score is not available for the proteins 1jbr and 1jbs. Interestingly 27% of the residues have the highest score of 9 and 67% of the residues have the conservation score $\geq$6. On the other hand only 61 and 49% of the cation–π interaction forming residues are conserved in DNA binding [22] and transmembrane strand proteins, [19], respectively. This result revealed that cation–π interaction forming residues in RNA binding proteins are more conserved than that in DNA binding and transmembrane strand proteins.

3.7. Solvent accessibility of cation–π interaction forming residues

We have estimated the solvent accessibility of all residues that are involved in cation–π interaction with the aid of DSSP [27]. We have analyzed the percentage of cation–π interaction forming residues at various range of solvent accessibility, such as: 0–20% (buried), 20–50% (partially buried), and $>$50% (surface exposed) [36–38] and the results are compared with DNA binding and membrane proteins (Fig. 4). The cation–π interaction forming Lys and Arg prefer to be in the surface of DNA and RNA binding proteins whereas these residues prefer to be in the interior of transmembrane helical proteins; there is no preference in transmembrane strand proteins as these residues are widely distributed in all ranges of solvent accessibility. Among the aromatic residues, Phe and Trp prefer to be in the interior of RNA binding proteins whereas Tyr prefers to be partially buried. The trend is different in DNA binding and membrane proteins In DNA binding proteins Tyr prefers to be at the surface and Trp has almost equal preference in all ranges of solvent accessibility. On the other hand most of the cation–π interactions forming aromatic residues in membrane proteins are buried.

3.8. Cation–π interaction forming residues in different secondary structures

We have calculated the occurrence of cation–π interaction forming residues in different secondary structures of RNA

Table 6
Frequency of occurrence of cation–π interaction forming residues in different secondary structures of RNA and DNA binding proteins

Residue	Helix		Coil		Strand		Turn	
	RNABP	DNABP	RNABP	DNABP	RNABP	DNABP	RNABP	DNABP
Lys	52.08 (39.60)	45.2 (46.5)	18.75 (31.59)	12.9 (32.9)	18.75 (14.79)	32.2 (11.1)	10.42 (14.02)	9.7 (9.5)
Arg	36.13 (41.02)	56.7 (53.4)	19.33 (25.47)	14.9 (26.9)	34.45 (20.64)	14.9 (13.3)	10.08 (12.87)	13.4 (6.4)
Phe	33.93 (32.64)	34.3 (34.8)	41.07 (27.13)	12.5 (25.6)	19.64 (31.49)	28.1 (31.3)	5.36 (8.74)	25.0 (8.3)
Trp	33.96 (39.22)	51.9 (49.5)	22.64 (25.53)	25.9 (25.8)	32.08 (28.76)	22.2 (22.5)	11.32 (8.50)	0.0 (2.2)
Tyr	28.57 (34.46)	30.8 (36.8)	28.57 (18.08)	25.6 (27.2)	37.50 (37.85)	33.3 (29.7)	5.36 (9.60)	10.3 (6.3)

The frequency of occurrence of each residue in the whole dataset is shown in parenthesis. RNABP, RNA binding protein; DNABP, DNA binding protein.

binding proteins and the results are presented in Table 6. Further, the data for DNA binding proteins are also included for comparison. We found that in RNA binding proteins the cation–π interaction forming Lys prefers to be in helix while Arg is dominated in β-strands. Most of the cation–π interaction forming aromatic residues are accommodated in β-strands and coil regions. On the other hand, most of the cation–π interactions forming residues in DNA binding proteins prefer to be in helical segments and Tyr prefers to be in β-strands. This observation reveals that cation–π interactions forming cationic and aromatic residues are located in specific secondary structures of RNA binding proteins compared with DNA binding proteins.

3.9. Comparison of cation–π interaction forming residues and stabilizing residues

We have identified 219 stabilizing residues in 39 out of the 51 considered RNA binding proteins (all except 1c0a, 1dk1, 1e7x, 1f7u, 1ffy, 1gax, 1hc6, 1hq1, 1il2, 1knz 1kq2, 1qf6) and the results are presented in Table 7. We observed an average of 2.8% residues as stabilizing ones (219 out of 7884) in RNA binding proteins. Interestingly, only five residues viz. Arg 77, Arg 120, Tyr 7, Arg 314 and Arg 118, respectively, in 1b2m, 1jbs, 1lng, 1ser and 2fmt, identified as stabilizing residues are also involved in cation–π interactions. This result indicates that the cation–π interactions have distinct roles to the stability of RNA binding proteins compared with other conventional non-covalent interactions including hydrophobic, electrostatic, hydrogen bonds, van der Waals etc. as reported for DNA binding proteins [39].

3.10. Role of cation–π interaction forming residues in protein-RNA binding interface

We have identified the binding site residues in all the protein-RNA complexes and result for the cation–π interaction forming residues, Lys, Arg, Phe, Trp and Tyr are presented in Table 8. We observed a significant number of contacts in the interface. However, most of these residues are not involved in cation–π interactions. We observed that just 8% of the binding residues are involved in cation–π interactions. This result indicates that the cation–π interaction forming cationic and aromatic residues play an important role to the stability whereas other residues contribute towards the specificity of protein-RNA complexes.

Table 7
Stabilizing residues in RNA binding proteins

PDB code	Stabilization residues
1b23A	Gly 101, Ala 102, Val 105, Val 132, Phe 134, Arg 244, Gly 245 Leu 304
1b2mP	Pro 39, Glu 58, Pro 60, **Arg 77**, Val 78
1b7fA	Val 129, Tyr 168, Val 171, Leu 196, Val 198, Ala 255, Val 257
1c0aA	No stabilizing residue
1cx0A	Thr 11, Ile 12, Ile 58
1di2A	Cys 149, Ala 171
1dk1A	No stabilizing residue
1efwA	His 522
1f7uA	No stabilizing residue
1f8vA	Pro 78, Phe 159, Ala 162, Val 183, Gly 238, Cys 316, Glu 318, Leu 319
1feuA	Leu 5, Leu 24, Pro 25, Gly 26, Val 37, Val 39, Ile 57, Thr 69, Leu 70, Val 100, Leu 102, Gly 115, Val 126, Ile 137, Ala 173
1ffyA	No stabilizing residue
1g59A	Thr 4, Arg 5, Ile 146
1gaxA	No stabilizing residue
1h4qA	Glu 196, Ile 292, Pro 345
1h4sA	Glu 196, Ile 292, Pro 345
1i6uA	Ile 47, Val 63, Leu 102, Ile 103,Ala 127, Val 129
1il2A	No stabilizing residue
1jbrA	Pro 48, His 49, Leu 94, Glu 95, Pro 97, Arg 120, Val 121, Ile 122, Tyr 123
1jbsA	Pro 48, His 49, Leu 94, Glu 95, Pro 97, **Arg 120**, Val 121, Ile 122, Tyr 123
1jidA	Leu 23, Gly 80, Val 82, Arg 83, Ile 112
1k8wA	Gly 13, Val 14, Leu 15, Leu 17, Met 54, Ile 57, Val 78, Ala 80, Val 155, Ile 172, Val 197, Leu 200, Val 205, Arg 282
1knzA	No stabilizing residue
1l9aA	Asp 9, Gly 62
1lngA	Trp 4, **Tyr 7**, Asp 9, Ala 37, Gly 62, Val 64, Ile 79, Cys 80
1mmsA	Leu 13, Leu 54, Pro 55, Pro 73, Ala 109, Ala 124
1qf6A	No stabilizing residue
1qtqA	Thr 29
1serA	Ala 249, **Arg 314**, Lys 327, Glu 393
1urnA	Thr 11, Ile 12, Ile 58, Gln 85
2bbvA	Pro 85, Asp 86, Val 93, Ala 122, Ala 169, Val 190, Gly 238, Val 298, Ile 299, Lys 314, Ser 336
2fmtA	Ile 6, Ile 7, Phe 8, Ala 9, Gly 10, Thr 11, Gly 31, Val 84, Met 85, Val 86, **Arg 118**, Thr 137, Ile 138

Bolded residues are involved in both stabilization and cation–π interaction.

Table 8
Binding site residues in protein-RNA complexes

PDB code	Contacting residues (Lys, Arg, Phe, Trp and Tyr) with RNA bases
1b23A	Tyr88, **Arg330**, Arg339, Lys 376
1b2mP	Tyr45
1b7fA	Arg155, Arg158, Tyr164, Arg195, Arg197, Arg202, Tyr214, Arg252, Arg287
1c0aA	Arg28, Arg64, Arg222, Arg225, Arg549
1c9sA	Lys37
1cx0A	Lys50, Arg52, Lys80, Arg83, Tyr86, Lys88
1dfuP	Lys14, Arg18, Arg19, Tyr31
1di2A	Nil
1dk1A	Lys107, Arg116, Lys147, Arg164, Tyr168, Arg171
1e7xA	Arg49
1ec6A	Lys43, Arg54, Arg83
1efwA	Arg29, Arg64, Arg78, Lys552
1f7uA	Lys319, Lys340, Tyr347, Lys439, Lys466, Tyr488, Tyr491, **Arg495**, Tyr565
1f8vA	Arg13
1feuA	Arg10, Arg19
1ffyA	**Arg440**, Arg560, Lys595, Arg632, Lys647, Arg653, Lys725, Tyr729, Arg805, **Lys823**, Arg888
1fxlA	Lys69, Lys108, Lys111, Arg116, Arg155, Arg166, Arg172, Lys201
1g2eA	Lys108, Lys111, Arg116, Arg172
1g59A	Arg47, Arg163, Lys243, Arg357, Arg358, Arg417, Arg435
1gaxA	Tyr337, Arg566, Arg570, Arg576, Lys581, Arg587, Trp642, Arg818, Lys831, Arg843
1gtfA	Lys37, Lys56
1gtnA	Lys37, Lys56
1h4qA	**Trp127**, Arg128
1h4sA	Arg125, **Trp127**, Arg128
1hc8A	Lys15, Arg29, Lys47, Arg61, Arg68
1hdwA	Arg49
1he0A	Not available
1he6A	Not available
1hq1A	Arg27, Lys38, Arg53
1i6uA	Lys32, Arg36, Arg78, Lys82, **Lys83**
1il2A	Arg28, Arg222, Arg225, Arg549
1jbrA	Arg65
1jbsA	Not Available
1jidA	Arg14, Phe15, **Tyr19**, **Tyr22**, Arg33, Arg70, Arg101
1k8wA	Lys130, Lys135, **Arg141**, Arg151, Lys176, Tyr179
1knzA	Arg83, **Trp87**, Lys132,
1kq2A	Lys57
1l9aA	**Trp4**, **Tyr7**, Arg14, Arg15, Arg18, **Lys19**, Lys51, Lys52, Arg55, Lys72, Lys77
1lngA	Trp4, **Tyr7**, Arg14, Arg15, Arg18, **Lys19**, Lys51, Lys52, Arg55, Tyr68, Lys69, Lys72, Lys77
1mmsA	Lys10, Lys80, Lys87, Lys93, Arg94, Lys112, Lys133
1qf6A	Tyr205, **Tyr219**, Arg245, Arg375, Tyr462, Lys599, Arg609
1qtqA	Arg133, Arg192, Lys194, Tyr211, Arg238, Lys317, Arg341, Lys401, Arg412, Arg520
1serA	Lys264
1urnA	Lys22, Lys50, Arg52, Lys80, Tyr86, Lys88,
1zdhA	Lys43, Tyr85
1zdiA	Lys61, Tyr85
2bbvA	**Lys68**
2fmtA	Arg42, Lys44, Lys209, Lys246, Lys291, Arg304
5msfA	Lys43, Arg49, Lys61, Tyr85
6msfA	Tyr85
7msfA	Arg49, Tyr85

Bold residues indicate the cation–π interaction forming residues.

4. Conclusions

We have analyzed the influence of cation–π interactions to the stability of RNA binding protein structures. We found that 63% of the considered RNA binding proteins exhibit cation–π interactions and the contribution of Arg is higher than Lys to form cation–π interactions. The cation–π interactions are mainly formed by long range interactions and Arg–Trp has the strongest cation–π interaction energy among all residue pairs. Secondary structure and solvent accessibility of the RNA binding proteins reveals that cation–π interactions forming cationic residues prefer to be in α-helices and β-strands and aromatic residues in β-strands and coil regions. While Arg and Lys prefer the exposed environment, the cation–π interaction forming aromatic amino acids Phe and Trp prefer to be buried. The cation–π interactions have distinct roles to the stability of

RNA binding proteins compared with other conventional non-covalent interactions. Further, the cation–π interaction forming cationic and aromatic residues play an important role to the stability of RNA binding proteins whereas the other residues contribute towards the specificity of protein-RNA complexes. The results obtained in this work will be helpful to understand the contribution of cation–π interactions to the stability and specificity of RNA binding proteins.

Acknowledgements

SC wishes to thank Dr G. Jayaraman for fruitful discussions and the management of Vellore Institute of Technology for providing the infrastructure.

References

[1] Berg P, Ofengand EJ. An enzymatic mechanism for linking amino acids to RNA. Proc Natl Acad Sci 1958;44:78–86.
[2] Hoagland MB. The relationship of nucleic acid and protein synthesis as revealed by studies in cellfree systems. In: Chargaff E, Davidson JN, editors. The nucleic acids. New York: Academic Press; 1960. p. 349–408.
[3] Traub P, Nomura M. Structure and function of *Escherichia coli* ribosomes. VI. Mechanism of assembly of 30 S ribosomes studied in vitro. J Mol Biol 1969;40:391–413.
[4] Nomura M. Assembly of bacterial ribosomes. Science 1973;179:864–73.
[5] Nagai K. RNA-protein complexes. Curr Opin Struct Biol 1996;6:53–61.
[6] Steitz TA. RNA recognition by proteins. In: Gesteland RF, Cech TR, Atkins JF, editors. The RNA world. New York: Cold Spring Harbor Laboratory Press; 1999. p. 427–50.
[7] Scrutton NS, Raine ARC. Cation–π bonding and amino–aromatic interactions in the biomolecular recognition of substituted ammonium ligands. Biochem J 1996;319:1–8.
[8] Jennifer MC, Dougherty DA. The cation–π interaction. Chem Rev 1997; 97:1303–24.
[9] Pletneva EV, Laederach AT, Fulton DB, Kostic NM. The role of cation–π interactions in biomolecular association. Design of peptides favoring interactions between cationic and aromatic amino acid side chains. J Am Chem Soc 2001;123:6232–45.
[10] Wintjens R, Lievin J, Rooman M, Buisine E. Contribution of cation–π interactions to the stability of protein-DNA complexes. J Mol Biol 2000; 302:395–410.
[11] Gallivan JP, Dougherty DA. A computational study of cation–π interactions Vs salt bridges in aqueous media: implications for protein engineering. J Am Chem Soc 2000;122:870–4.
[12] Novotny J, Bruccoleri RE, Saul FA. On the attribution of binding energy in antigen–antibody complexes McPC 603, D1.3, and HyHEL-5. Biochemistry 1989;28:4735–49.
[13] Pellequer JL, Zhao B, Kao HI, Bell CW, Li K, Li QX. Stabilization of bound polycyclic aromatic hydrocarbons by a π–cation interaction. J Mol Biol 2000;302:691–9.
[14] Dougherty DA, Stauffer DA. Acetylcholine binding by a synthetic receptor: implications for biological recognition. Science 1990;250: 1558–60.
[15] Mecozzi S, West AP, Dougherty DA. Cation–π interactions in aromatics of biological medicinal interest: electrostatic potential surfaces as a useful qualitative guide. Proc Natl Acad Sci 1996;93:10566–71.
[16] Chakravarthy S, Varadarajan R. Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. Biochemistry 2000;41:8152–61.
[17] Gromiha MM, Thomas S, Santhosh C. Role of cation–π interaction to the stability of thermophilic proteins. Prep Biochem Biotech 2002;32: 355–62.
[18] Shi Z, Olson CA, Kallenbach NR. Cation–π interaction in model alpha-helical peptides. J Am Chem Soc 2002;124:3284–91.
[19] Gromiha MM, Suwa M. Structural analysis of residues involving cation–π interactions in different folding types of membrane proteins. Int J Biol Macromol 2005;35:55–62.
[20] Gromiha MM. Influence of cation–π interaction in different folding types of membrane proteins. Biophys Chem 2003;103:251–8.
[21] Gromiha MM, Santhosh C, Suwa M. Influence of cation–π interactions in protein-DNA complexes. Polymer 2004;45:633–9.
[22] Gromiha MM, Santhosh C, Ahmad S. Structural analysis of cation–π interactions in DNA binding proteins. Int J Biol Macromol 2004;34: 203–11.
[23] Kima H, Jeonga E, Leeb S-W, Hana K. Computational analysis of hydrogen bonds in protein-RNA complexes for interaction patterns. FEBS Lett 2003;552:231–9.
[24] Berman HM, Westbrook JZ, Feng G, Gilliland TN, Bhat H, Weissig IN, et al. The protein data bank. Nucleic Acids Res 2000;28:235–42.
[25] Gallivan JP, Dougherty DA. Cation–π interactions in structural biology. Proc Natl Acad Sci 1999;96:9459–64.
[26] Jorgensen WL, Maxwell DS, TiradoRives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 1996;118:11225–36.
[27] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–637.
[28] Heringa J, Argos P. Strain in protein structures as viewed through nonrotameric side chains. I. Their position and interaction. Proteins 1989; 37:30–43.
[29] Gromiha MM, Selvaraj S. Influence of medium and long range interactions in different structural classes of globular proteins. J Biol Phys 1997;23:151–62.
[30] Selvaraj S, Gromiha MM. Role of hydrophobic clusters and long-range contact networks in the folding of $(alpha/beta)_8$ barrel proteins. Biophys J 2003;84:1919–25.
[31] Glaser F, Pupko T, Paz I, Bell RE, Bechor D, Martz E, et al. ConSurf: identification of functional regions in proteins by surface mapping of phylogenetic information. Bioinformatics 2003;19:163–4.
[32] Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledge base and its supplement TrEMBL in 2003. Nucleic Acids Res 2003;31:365–70.
[33] Gromiha MM, Selvaraj S. Inter-residue interactions in protein recognition, folding and stability. Prog Biophys Mol Biol 2004;86:235–77.
[34] Magyar C, Gromiha MM, Pujadas G, Tusnady GE, Simon I. Sride: a server for identifying stabilizing residues in proteins. Nucleic Acids Res 2005;33:303–5.
[35] Hoffman MM, Khrapov MA, Cox JC, Yao J, Tong L, Ellington AD. AANT: the amino acid-nucleotide interaction database. Nucleic Acids Res 2004;32:D174–D81.
[36] Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A. Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations. Protein Eng 1999;12:549–55.
[37] Gilis D, Rooman M. Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials. J Mol Biol 1996;257:1112–26.
[38] Gilis D, Rooman M. Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. J Mol Biol 1997;272:276–90.
[39] Gromiha MM. Distinct roles of conventional non-covalent and cation–π interactions in protein stability. Polymer 2005;46:983–90.