Available online at www.sciencedirect.com
SCIENCE DIRECT®
Polymer 45 (2004) 633–639
polymer
www.elsevier.com/locate/polymer

Influence of cation–π interactions in protein–DNA complexes

M. Michael Gromiha[a,*], C. Santhosh[b], Makiko Suwa[a]

[a]*Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan*

[b]*Centre for Laser Spectroscopy, Science Centre, Manipal Academy of Higher Education, Manipal, India*

Received 21 August 2003; received in revised form 8 October 2003; accepted 8 October 2003

Abstract

Protein–DNA recognition plays a crucial role in gene expression and regulation. In this work, we have analyzed the influence of cation–π interactions to the stability of 62 protein–DNA complexes. A new criterion has been formulated to delineate the cation–π interactions based on (i) the distribution of atoms in the π system (5 and 6-member rings) of DNA bases around the positive charged atoms of Lys and Arg and (ii) the energetic contribution of contacting atoms from electrostatic and van der Waals interactions. Our method shows the presence of cation–π interactions in 92% of the complexes. The side chain of Arg is more likely than that of Lys to be in cation–π interactions. In both Lys and Arg, the cationic groups have stronger cation–π interaction energy than the atoms with effective positive charge. The aromatic chains of purines (A and G) are exhibiting more cation–π interactions than pyrimidines (C and T). The Arg-G pair has the strongest interaction energy of −4.3 kcal/mol among all the possible pairs of amino acids and bases. The interaction energy is always positive for T and we observed few favorable interactions with C. Further, we found that the cation–π interactions due to 5-member rings of A and G are stronger than that with the atoms in 6-member rings. The distribution of base atoms around the charged atoms shows that the N7 in the 5-member ring of G is making significant number of close contacts with NZ of Arg, which is important to establish dominant cation–π interactions.

Keywords: Cation–π interactions; Protein–DNA recognition; Interaction energy

1. Introduction

Protein–DNA recognition plays an important role in all mechanisms of gene expression and control. The structural data obtained from X-ray crystallography and NMR spectroscopy provide valuable information to understand the general features of protein–nucleic acid recognition. Based on the three dimensional structures of protein–DNA complexes, several investigations have been carried out to reveal the mechanism of protein–DNA recognition, such as, the importance of hydrogen bonding and hydrophobic interactions [1], role of CH…O interactions [2], interface surface area between protein and DNA [3], chemical and physical properties of the protein–DNA interface [4], contacts between amino acids and base pairs [5], and geometric features [6]. Recently, the specificity of base-amino acid interactions has been studied by systematic sampling procedure [7,8] and conservation of amino acid residues in protein–DNA complexes [9]. Further, Stawiski et al. [10] proposed a method for identifying DNA-binding proteins from structural and sequence properties of protein–DNA complexes.

The stability and specificity of protein–DNA complexes are determined by several non-covalent interactions, including electrostatic, hydrogen bonding, van der Waals and hydrophobic interactions. In addition, the cation–π interactions are recognized to play an important role to the stability of proteins and protein–DNA complexes [11,12]. Recently, the importance of this interaction has been stressed by several investigators in determining the helicity of α-helical peptides [13], folding of polypeptides [14] and the stability of membrane protein structures [15]. Further, the role of cation–π interactions to the stability of thermophilic proteins has been reported [16,17]. However, the influence of cation–π interactions to the stability of protein–DNA complexes is not yet completely explored.

In this work, we have analyzed the influence of cation–π interactions in 62 protein–DNA complexes. We have focused our study at the protein–DNA interface and hence the cation–π interactions within a protein or DNA are not considered. We have formulated a new criterion to delineate such interactions based on the information about the distribution of atoms in the aromatic rings of DNA bases

* Corresponding author. Tel.: +81-3-3599-8046; fax: +81-3-3599-8081.
E-mail address: michael-gromiha@aist.go.jp (M. Michael Gromiha).

0032-3861/$ - see front matter © 2003 Published by Elsevier Ltd.
doi:10.1016/j.polymer.2003.10.069

around positive charged amino acid residues and the interaction energy between the contacting atoms. We found that 92% of the protein–DNA complexes have cation–π interactions and the pair, Arg-G has the strongest cation–π interaction energy among all possible pairs of bases and amino acid residues. Further, the relative contribution of cation–π interaction energy between (i) Lys and Arg, (ii) cationic group and atoms with substantial positive charge, (iii) purines and pyrimidines, and (iv) 5 and 6-member rings of A and G has been brought out.

2. Materials and methods

2.1. Data set

We have considered a set of 62 non-redundant protein–DNA complexes (sequence homology is less than 25% identity) from the information available in literature [18]. The PDB codes of the proteins used in the present study are, 1a02, 1a74, 1aay, 1azq, 1b3t, 1ber, 1bf5, 1bhm, 1bl0, 1c0w, 1cdw, 1cf7, 1cjg, 1cmcma, 1d02, 1d66, 1dp7, 1ecr, 1fjl, 1gat, 1gcc, 1gdt, 1hcq, 1hcr, 1hdd, 1hlo, 1hry, 1hwt, 1if1, 1ign, 1ihf, 1lmb, 1mdy, 1mey, 1mhd, 1mnm, 1mse, 1oct, 1par, 1pdn, 1per, 1pnr, 1pue, 1pvi, 1pyi, 1rep, 1srs, 1svc, 1tc3, 1tf3, 1tro, 1tsr, 1ubd, 1xbr, 1yrn, 1ysa, 1yui, 2bop, 2drp, 2gli, 2hdc and 3cro. The coordinates of all the protein–DNA complex structures have been taken from the Protein Data Bank [19].

2.2. Definition of cation–π interactions

We have first selected the atoms in the 5 and 6-member rings of DNA bases (A, C, T and G). Among the amino acids, we have chosen the positive charged residues, Lys and Arg, and we considered both (i) the cationic group atoms (NZ in Lys and CZ in Arg) and (ii) the atoms with substantial positive charge (CE in Lys and CD in Arg) to represent positively charged atoms in cation–π interactions [11,15]. If all atoms in the 6- (or 5) member ring of a base are within 10 Å from the positively charged atom (either cationic or with substantial charge) then the π system is considered to have a contact with positive charged atom. For each contacting pairs, we have calculated both the electrostatic and van der Waals energy (see below for details) using AMBER force field. As the positive energies are not favorable for the stability, we have considered the interacting pairs only with negative energies. It has been reported that the strength of electrostatic energy is twice to that of van der Waals interactions in cation–π interactions [11,15]. Hence, we impose the limit of −0.1 kcal/mol for van der Waals energy and −0.2 kcal/mol for electrostatic energy to delineate cation–π interactions. In essence, if the interacting pair (π system-positive charged atom within 10 Å) has $E_{el} < -0.2$ kcal/mol and $E_{vdw} < -0.1$ kcal/mol then the pair is considered to be in cation–π interaction.

2.3. Calculation of amino acid/base composition in protein–DNA complexes

We have computed the number of occurrence of each of the nucleotides (A, C, T or G) in the base (DNA) sequence of all the 62 protein–DNA complexes (n). The composition of each base is computed using the formula,

$$\mathrm{Comp}(i) = n(i)/N$$

where, i stands for the four nucleotides and N is the total number of nucleotides in 62 complexes.

We have followed the same method to compute the composition of amino acids Lys and Arg using protein sequences.

2.4. Computation of cation–π interaction energy

We have used the AMBER 4.1 force field [20] to compute the contribution of cation–π interaction energy. It is the sum of electrostatic (E_{el}) and van der Waals energy (E_{vdw}) terms. The E_{el} is computed using the expression:

$$E_{el} = \sum q_i q_j / \varepsilon r_{ij},$$

where q_i and q_j are, respectively, the charges for the atoms i and j, and r_{ij} is the distance between them. We have used the distant dependent dielectric constant ($\varepsilon = r_{ij}$) to take account of the dielectric damping effect of the Coulomb interactions, as used in other studies on protein–DNA interactions [7].

The van der Waals energy is given by

$$E_{vdw} = 4\varepsilon_{ij}(A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6}),$$

where $A_{ij} = \varepsilon_{ij}^{*}(R_{ij}^{*})^{12}$ and $B_{ij} = 2\varepsilon_{ij}^{*}(R_{ij}^{*})^{6}$; $R_{ij}^{*} = (R_i^{*} + R_j^{*})$ and $\varepsilon_{ij}^{*} = (\varepsilon_i^{*}\varepsilon_j^{*})^{1/2}$; R^{*} and ε^{*} are, respectively, the van der Waals radius and well depth and these parameters are obtained from Cornell et al. [20].

We have applied the AMBER force field in vacuum environment and the solvation effects are not considered in the present work. Since, most of the solvent molecules are excluded in the DNA–protein interface, the intrinsic interactions may contribute directly to the specificity of protein–DNA recognition. On the other hand, when one or more water molecules are occasionally trapped inside the DNA–protein interface, they may affect the electrostatic interactions. The analysis on the structure of protein–DNA complexes in the vicinity of cation–π interaction sites due to the presence of cavities accessible to water molecules will be considered in future.

2.5. Distribution of base atoms towards Lys and Arg

We have further analyzed the factors influencing the cation–π interaction energy. The distance between aromatic ring atoms (of DNA) and positive charged atoms (in amino acid residues) plays a major role and hence we

Table 1
Composition and cation–π interactions exhibited by each amino acid/base in protein–DNA complexes

Amino acid/base	N	Comp	$N_{cat-\pi}$	$\%_{cat-\pi}$
Lys	1014	7.92	148	14.60
Arg	1024	8.00	247	**24.12**
A	678	27.97	140	20.65
G	523	21.58	239	**45.70**
C	532	21.95	16	3.01
T	688	28.38	0	0.00

N : number of amino acid/base; $N_{cat-\pi}$ and $\%_{cat-\pi}$ are, respectively, number and % of cation–π interactions in 62 protein–DNA complexes. The highest percentages of cation–π interactions between positive charged amino acids and among DNA bases are shown in bold.

analyzed the distribution of base atoms around positive charged atoms of Lys and Arg at different distance intervals from 3.5 to 10 Å in steps of 0.5 Å. The relative frequency of occurrence of all atoms and specific atoms (N and C) in 5 and 6-member rings of DNA around the positive charged atoms in Lys and Arg have been discussed.

3. Results and discussions

3.1. Amino acid/base composition and average cation–π interactions in protein–DNA complexes

We have computed the composition of Lys and Arg using protein sequences and that of A, T, C and G with DNA sequences in the 62 protein–DNA complexes. The number and composition of each amino acid/base is presented in Table 1. We found that the amino acid composition is similar for both Lys and Arg. On the other hand, the bases A and T have higher occurrence than C and G.

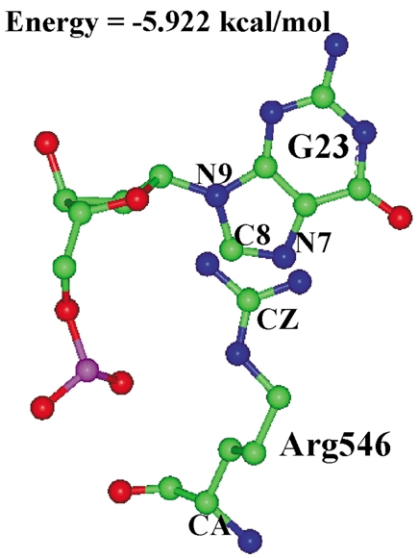

Fig. 1. Cation–π interaction between CZ of Arg546 and 5-member ring of G23 in the protein–DNA complex 1ign.

The number and average cation–π interactions exhibited by each of the amino acid/base are included in Table 1. The contribution of Arg is twice to that of Lys, as observed in globular and transmembrane strand proteins [11,15]. Considering the bases, G has the highest contribution followed by A. The contribution from C is minimal and there is no favorable cation–π interaction from T. This results shows that the pair Arg-G establishes favorable cation–π interaction at the interface of protein–DNA complexes and it is one of the most frequently occurring pairs [5].

3.2. Energetic contribution of cation–π interactions

The energetic contributions produced by all possible combinations of bases and amino acids (A-Lys, A-Arg, G-Lys, G-Arg, C-Lys and C-Arg; T has no favorable cation–π interaction either with Lys or with Arg) have been computed and the results obtained for all the 62 protein–DNA complexes are presented in Table 2. We found that 92% of the complexes (57/62) are exhibited by cation–π interactions whereas only 71% (34 out of 48 complexes) have such interactions based on the criterion of Wintjens et al. [12]. Further analysis indicates that the contribution from Arg (71%) is higher than Lys (53%) in the considered protein–DNA complexes towards cation–π interactions. Among the bases, 84% of the complexes have the contribution from G and 56% from A. Only 12% of the complexes are experienced with cation–π interactions from C. It may be noted that the same complex have the contribution from different amino acids/bases (e.g. A and G in 1a74; Lys and Arg in 1ysa). Moreover, the highest contribution is observed for the combination of Arg and G, as revealed in our previous discussion. This might be due to the fact that Arg-G pair occurs frequently at the protein–DNA interface [5,21].

From Table 2, we observed that the strength of cation–π interaction energy is different in each complex and it varies from −1.26 kcal/mol (1per) to −54.54 kcal/mol (1a74). It is noteworthy that the interaction energy is mainly due to the E_{el} and the contribution from E_{vdw} is very minimal. The bases A, G and C have favorable interactions with Lys/Arg while the energy is always positive for the combinations T-Lys and T-Arg. On an average, G-Arg has the strongest cation–π interaction energy of −4.31 kcal/mol. This trend is similar to the observation of Wintjens et al. [12] that G-Arg system is the most stable one. As an example, the cation–π interaction between G23 and Arg546 in 1ign is displayed in Fig. 1. The CZ of Arg546 is contacting with all the atoms in the 5-membered ring of G23 and the total interaction energy is −5.9 kcal/mol. There is no correlation between the number of amino acids/bases and number of cation–π interactions/cation–π interaction energy. However, we noticed a good correlation between the number of cation–π interactions and cation–π interaction energy (r = 0.90).

Table 2
Energetic contribution due to cation–π interactions

PDB	N_a	N_b	$N_{cat-\pi}$	Cation–π interaction energy (kcal/mol)						
				A-Lys	A-Arg	G-Lys	G-Arg	C-Lys	C-Arg	Total
1a02	385	40	6	0.00	−3.83	0.00	−14.66	0.00	0.00	−18.49
1a74	324	42	25	−7.72	−25.60	−3.29	−17.93	0.00	0.00	−54.54
1aay	85	22	8	0.00	0.00	0.00	−38.87	0.00	0.00	−38.87
1azq	66	16	1	0.00	−1.34	0.00	0.00	0.00	0.00	−1.34
1b3t	294	36	11	−7.08	0.00	−18.29	0.00	0.00	0.00	−25.37
1ber	396	62	5	0.00	0.00	0.00	−17.16	0.00	0.00	−17.16
1bf5	544	36	2	0.00	0.00	−2.11	0.00	−1.51	0.00	−3.62
1bhm	406	23	6	0.00	0.00	0.00	−17.05	0.00	0.00	−17.05
1bl0	116	48	1	0.00	0.00	0.00	−4.76	0.00	0.00	−4.76
1c0w	742	42	4	0.00	0.00	0.00	−5.63	0.00	−2.51	−8.14
1cdw	179	32	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1cf7	149	30	9	0.00	−0.50	0.00	−21.31	0.00	−1.62	−23.43
1cjg	124	44	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1cma	208	19	7	−12.13	0.00	−8.23	0.00	0.00	0.00	−20.36
1d02	397	20	10	0.00	−24.80	0.00	−13.68	0.00	0.00	−38.48
1d66	114	38	6	0.00	0.00	−14.94	0.00	0.00	0.00	−14.94
1dp7	76	16	3	−1.44	0.00	0.00	−8.59	0.00	0.00	−10.03
1ecr	305	30	10	−0.69	−6.58	−2.52	−9.37	0.00	0.00	−19.16
1fjl	182	48	5	0.00	−10.99	0.00	0.00	0.00	0.00	−10.99
1gat	60	16	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1gcc	63	22	8	0.00	0.00	0.00	−28.50	0.00	0.00	−28.50
1gdt	366	69	13	0.00	−19.61	0.00	−13.83	0.00	0.00	−33.44
1hcq	145	72	15	−9.65	0.00	−20.28	−8.97	0.00	0.00	−38.90
1hcr	52	27	3	0.00	−2.69	0.00	−5.52	0.00	0.00	−8.21
1hdd	114	42	3	0.00	0.00	−4.18	0.00	0.00	0.00	−4.18
1hlo	153	22	6	0.00	0.00	0.00	−10.90	0.00	−4.86	−15.75
1hry	73	16	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1hwt	144	80	10	0.00	−3.58	−6.32	−8.09	0.00	0.00	−17.99
1if1	209	52	1	−1.31	0.00	0.00	0.00	0.00	0.00	−1.31
1ign	189	75	12	−2.69	0.00	0.00	−39.85	0.00	0.00	−42.54
1ihf	190	70	3	−1.43	−5.29	0.00	0.00	0.00	0.00	−6.72
1lmb	179	40	8	0.00	0.00	−20.75	0.00	0.00	0.00	−20.75
1mdy	254	56	15	0.00	−5.54	0.00	−21.52	0.00	−7.18	−34.24
1mey	83	52	12	−17.14	−4.54	−8.45	−12.70	0.00	0.00	−42.84
1mhd	246	27	9	−2.87	−4.96	−9.24	−10.81	0.00	0.00	−27.88
1mnm	320	52	14	0.00	−18.48	−14.90	−8.31	0.00	0.00	−41.70
1mse	105	22	5	−3.74	0.00	−3.54	0.00	0.00	−1.56	−8.84
1oct	131	30	2	0.00	−2.96	0.00	−3.12	0.00	0.00	−6.08
1par	208	44	5	0.00	−12.42	0.00	−12.47	0.00	0.00	−24.89
1pdn	123	30	2	0.00	0.00	−6.36	0.00	0.00	0.00	−6.36
1per	126	40	1	0.00	0.00	0.00	0.00	−1.26	0.00	−1.26
1pnr	338	17	2	−4.07	0.00	0.00	−4.45	0.00	0.00	−8.52
1pue	177	64	11	0.00	0.00	0.00	−40.89	0.00	0.00	−40.89
1pvi	312	26	4	0.00	0.00	−5.80	0.00	0.00	0.00	−5.80
1pyi	158	28	6	0.00	0.00	−12.32	0.00	−1.78	0.00	−14.10
1rep	214	42	6	0.00	−0.84	−2.80	−14.62	0.00	0.00	−18.27
1srs	164	38	13	−0.54	−8.11	−11.56	0.00	0.00	0.00	−20.20
1svc	311	19	2	−1.98	0.00	0.00	−2.64	0.00	0.00	−4.62
1tc3	51	41	4	0.00	−0.91	0.00	−13.40	0.00	0.00	−14.31
1tf3	92	30	9	0.00	0.00	−20.81	−11.22	0.00	0.00	−32.03
1tro	405	76	6	0.00	0.00	−4.46	−21.44	0.00	0.00	−25.90
1tsr	585	42	5	0.00	0.00	−7.56	−5.87	0.00	0.00	−13.44
1ubd	114	40	7	0.00	0.00	−8.73	−13.59	0.00	0.00	−22.32
1xbr	367	48	2	0.00	0.00	0.00	−10.59	0.00	0.00	−10.59
1yrn	127	42	7	0.00	−12.12	0.00	−12.94	0.00	0.00	−25.06
1ysa	114	40	6	−2.77	−4.50	−2.94	−6.00	0.00	0.00	−16.20
1yui	54	22	8	−1.31	−11.06	0.00	−9.41	0.00	0.00	−21.78
2bop	85	17	2	0.00	0.00	−4.72	0.00	0.00	0.00	−4.72
2drp	128	76	12	0.00	−12.14	0.00	−39.27	0.00	0.00	−51.40
2gli	155	42	9	0.00	−1.00	−15.62	−6.17	0.00	0.00	−22.78
2hdc	97	34	8	0.00	−8.36	0.00	−7.26	0.00	0.00	−15.61
3cro	130	40	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00

N_a and N_b are, respectively, the number of amino acids and bases.

Table 3
Average cation–π interaction energy (in kcal/mol) for all possible combinations of amino acids and bases

Base	Lys	Arg
A	−2.01	−2.10
G	−2.27	**−4.31**
C	−1.52	−1.36
T	+1.17	+1.55

The highest cation–π interaction energy is bold

3.3. Cation–π interaction energy due to purines and pyrimidines

The relative contribution of cation–π interaction energy due to purine and pyrimidine bases has been analyzed and the summary of results is tabulated in Table 3. We found that the purines (A and G) have the strongest cation–π interaction energy with both Lys and Arg. The average energy due to A and G are, respectively, −2.05 and −3.29 kcal/mol. These energies are higher than that of C and T. The average interaction energy of C with Lys and Arg are, respectively, −1.52 and −1.36 kcal/mol. On the other hand, there is no favorable interaction between T and positively charged amino acids. The average energy due to T is 1.17 kcal/mol with Lys and 1.55 kcal/mol with Arg. This result shows that the purines are more important than pyrimidines to the stability of protein–DNA complexes.

3.4. Contribution of cation–π interaction energy due to cationic groups and atoms with substantial positive charge

We have analyzed the relative contribution of cation–π interactions due to atoms with cationic group (NZ in Lys and CZ in Arg) and substantial positive charge (CE in Lys and CD in Arg). The interaction energy for A, G and C are presented in Table 4 (T has positive energy). We observed that both in Lys and Arg, cationic groups are forming stronger cation–π interaction energy than the atoms with substantial positive charge. This trend is opposite to that reported in thermophilic proteins [17]. Table 4 further indicates that the highest contribution of G-Arg is mainly due to the CZ (cationic group) of Arg, which has the cation–π interaction energy of −5 kcal/mol.

Table 4
Average cation–π interaction energy (in kcal/mol) from cationic group atoms and atoms with substantial positive charge

Base	Lys		Arg	
	NZ	CE	CZ	CD
A	−2.47	−0.95	−2.91	−0.87
G	−3.11	−1.36	**−5.00**	−1.01
C	−1.52	0.35	−1.53	−0.44

The highest cation–π interaction energy is bold.

3.5. Effective cation–π interaction energy due to 6 and 5-member rings of adenine and guanine

The contribution of cation–π interaction energy due to 5 and 6-member rings of A and G has been computed and the results are presented in Table 5. We found that there is no favorable cation–π interaction between the 6-member ring of G and positive charged atoms. On the other hand, we observed very strong cation–π interaction energy with 5-member ring of G (see Table 5). Considering A, there is no significant difference between the number of cation–π interactions with 5 and 6-member rings. However, the 5-member rings have stronger cation–π interaction energy than 6-member rings. This result shows the vital role played by 5-member rings in purines for the contribution to cation–π interactions.

Table 5
Contribution of cation–π interactions from 5 and 6-member rings of A and G

Amino acid-base pair	N_{cont}	$N_{cat-\pi}$	E_{el}	E_{vdw}	$E_{cat-\pi}$
Lys (NZ)-G	*190*	*55*	*−156.209*	*−14.660*	*−170.869*
Lys (CE)-G	*176*	*51*	*−58.683*	*−11.175*	*−69.858*
Arg (CZ)-G	*299*	*110*	*−526.233*	*−23.776*	*−550.009*
Arg (CD)-G	*255*	*23*	*−18.980*	*−4.361*	*−23.341*
Lys (NZ)-A	145	13	−17.315	−3.392	−20.707
	198	*11*	*−35.842*	*−2.845*	*−38.687*
Lys (CE)-A	118	6	−4.282	−1.444	−5.726
	163	*9*	*−11.012*	*−2.429*	*−13.441*
Arg (CZ)-A	302	35	−72.430	−7.075	−79.505
	318	*26*	*−92.452*	*−5.787*	*−98.239*
Arg (CD)-A	209	22	−12.072	−5.464	−17.536
	260	*18*	*−13.950*	*−3.507*	*−17.457*

N_{cont} : number of π systems, which are having contacts with positive charged atoms. $N_{cat-\pi}$: number of cation–π interactions. The contribution due to 5-member rings of A and G are shown in italics.

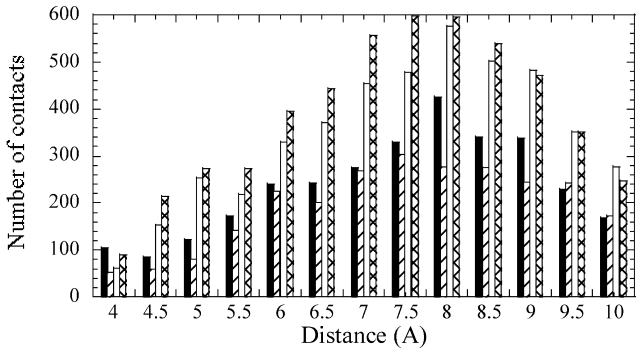

Fig. 2. Total number of contacts between N and C atoms in 5 and 6-member rings of A and G and positively charged atoms at different distances. Each interval has the bin size of 0.5 Å. Hence, the number at the *X*-axis shows the interval, such as, 3.5–4 Å, 4–4.5 Å, 4.5–5 Å etc. Filled column: N in 5-member ring; slant column: N in 6-member ring; empty column: C in 5-member ring and cross column: C in 6-member ring.

3.6. Distribution of CZ in Arg around 5 and 6-member rings of guanine

We have analyzed the distribution of atoms in the 5 and 6-member rings of G and CZ of Arg, which has the strongest cation–π interaction energy. We found that N7 in the 5-member ring of G has close contacts (within 5 Å) with CZ in Arg and this distribution is frequently occurring in protein–DNA complexes as seen in Table 6. On the other hand, there are not many contacts between N3 in the 6-member ring of G and CZ of Arg within the same distance. This might explain the stronger cation–π interaction energy attributed with the 5-member ring of G than the 6-member ring. The distribution of N9 and N1 are similar around CZ of Arg. The carbon atoms C2, C4 and C6 have the distribution in the wide range of 5.5–9.5 Å and few C5 and C8 atoms occur within the range of 5.5 Å.

3.7. Distribution of atoms in 5 and 6-member rings of adenine and guanine around cationic groups and atoms with substantial positive charge in Lys and Arg

We have analyzed the general trend of occurring 5 and 6-member rings of purines around the atoms responsible for cation–π interactions and the results are presented in Table 7. We found that 5-member ring atoms are more frequently occurring around Lys and Arg than 6-member ring atoms. Considering the 5-member ring, the average contacts within 4 Å for N is twice to that of C. At other intervals, the variation is random for the N and C atoms. In the 6-member ring, the difference of average contacts between N and C is not significant.

The preference of N and C atoms in the 5 and 6-member rings shows that the N atoms in 5-member ring atoms are more frequently occurring than 6-member rings around the charged atoms (Fig. 2). On the other hand, C atoms in the 6-member ring are closer to the charged atom than that in 5-member ring.

4. Conclusions

We have systematically analyzed the influence of cation–π interactions to the stability of protein–DNA complexes. A new criterion for identifying cation–π interactions in protein–DNA complexes has been proposed and we found that 92% of the complexes are exhibiting cation–π interactions. Purines play a dominant role in establishing cation–π interactions and the side chain of Arg is more likely to be in cation–π interaction than Lys. We observed the strongest cation–π interaction energy for the G-Arg pair with an average energy of −4.3 kcal/mol and the cationic groups atoms are mainly responsible for

Table 6
Distribution of atoms in the 5 and 6-member rings of G around CZ of Arg

Distance (Å)	5-Member ring					6-Member ring					
	N7	C8	N9	C4	C5	N1	C2	N3	C4	C5	C6
3.5–4.0	**46**	5	0	0	0	0	2	7	0	7	0
4.0–4.5	**17**	14	6	0	27	0	0	9	0	48	26
4.5–5.0	7	38	8	9	29	2	4	5	9	38	20
5.0–5.5	3	14	5	12	16	3	3	2	12	21	15
5.5–6.0	7	9	47	36	10	39	1	11	36	11	21
6.0–6.5	8	9	24	29	9	22	7	12	29	9	9
6.5–7.0	18	13	9	10	14	20	44	42	11	14	10
7.0–7.5	14	19	20	10	17	21	20	43	11	19	12
7.5–8.0	24	47	18	25	14	12	29	24	24	16	15
8.0–8.5	20	17	28	17	18	15	19	13	17	11	19
8.5–9.0	8	11	28	10	9	20	15	15	11	11	12
9.0–9.5	14	5	10	15	13	5	13	14	5	9	6
9.5–10.0	8	4	8	17	13	6	6	8	4	3	3

The highest number of *N* atoms within the distance intervals of 3.5–4.0 and 4.0–4.5 Å is highlighted as bold.

Table 7
Number of 5 and 6-member ring atoms in A and G around the charged atoms of Lys and Arg

Distance (Å)	5-Member ring					6-Member ring					
	N7	C8	N9	C4	C5	N1	C2	N3	C4	C5	C6
3.5–4.0	97	32	7	3	27	9	16	44	3	40	30
4.0–4.5	57	67	29	23	62	15	31	43	22	89	72
4.5–5.0	57	95	65	71	88	37	30	43	70	106	67
5.0–5.5	72	71	101	73	74	88	40	53	72	83	78
5.5–6.0	105	108	136	120	101	138	71	86	120	102	101
6.0–6.5	121	130	121	131	111	121	81	79	133	116	112
6.5–7.0	142	170	133	118	167	140	129	129	120	165	142
7.0–7.5	151	190	178	136	152	129	146	174	140	160	151
7.5–8.0	167	241	258	187	148	145	139	132	171	146	139
8.0–8.5	149	177	192	190	134	144	150	132	163	120	105
8.5–9.0	157	158	181	183	141	99	138	146	124	116	94
9.0–9.5	113	68	117	137	146	81	130	162	70	66	86
9.5–10.0	87	44	82	121	113	78	92	94	45	53	56

cation–π interactions. The comparison of cation–π interaction energy due to 5 and 6-member rings of purines shows that the 5-member rings are forming stronger cation–π interactions than 6-member rings. Further, the distribution of atoms in 5 and 6-member rings towards the charged atoms shows the presence of several close contacts between nitrogen atoms in the 5-member ring and the charged atoms, which are mainly influencing the formation of cation–π interactions. The results obtained in the present study would be helpful to understand the mechanism of protein–DNA recognition.

Acknowledgements

MMG wishes to thank Dr Yutaka Akiyama for encouragement.

References

[1] Mandel-Gutfreund Y, Margalit H. Quantitative parameters for amino acid-base interaction: implications for prediction of protein–DNA binding sites. Nucleic Acids Res 1998;26:2306–12.
[2] Mandel-Gutfreund Y, Margalit H, Jernigan RL, Zhurkin VB. A role for CH…O interactions in protein–DNA recognition. J Mol Biol 1998;277:1129–40.
[3] Nadassy K, Wodak SJ, Janin J. Structural features of protein–nucleic acid recognition sites. Biochemistry 1999;38:1999–2017.
[4] Jones S, van Heyningen P, Berman HM, Thornton JM. Protein–DNA interactions: a structural analysis. J Mol Biol 1999;287:877–96.
[5] Kono H, Sarai A. Structure-based prediction of DNA target sites by regulatory proteins. Proteins 1999;35:114–31.
[6] Pabo CO, Nekludova L. Geometric analysis and comparison of protein–DNA interfaces: why is there no simple code for recognition? J Mol Biol 2000;301:597–624.
[7] Pichierri F, Aida M, Gromiha MM, Sarai A. Free energy maps of base–amino acid interactions for protein–DNA recognition. J Am Chem Soc 1999;121:6152–7.
[8] Sayano K, Kono H, Gromiha MM, Sarai A. Multicanonical monte carlo calculation of the free energy map of the base–amino acid interaction. J Comput Chem 2000;21:954–62.
[9] Luscombe NM, Thornton JM. Protein–DNA interactions: amino acid conservation and the effects of mutations on binding specificity. J Mol Biol 2002;320:991–1009.
[10] Stawiski EW, Gregoret LM, Mandel-Gutfreund Y. Annotating nucleic acid-binding function based on protein structure. J Mol Biol 2003; 326:1065–79.
[11] Gallivan JP, Dougherty DA. Cation–pi interactions in structural biology. Proc Natl Acad Sci USA 1999;96:9459–64.
[12] Wintjens R, Lievin J, Rooman M, Buisine E. Contribution of cation–pi interactions to the stability of protein–DNA complexes. J Mol Biol 2000;302:395–410.
[13] Shi Z, Olson CA, Kallenbach NR. Cation–π interaction in model alpha-helical peptides. J Am Chem Soc 2002;124:3284–91.
[14] Burghardt TP, Juranic N, Macura S, Ajtai K. Cation–π interaction in a folded polypeptide. Biopolymers 2002;63:261–72.
[15] Gromiha MM. Influence of cation–π interactions in different folding types of membrane proteins. Biophys Chem 2003;103:251–8.
[16] Gilis D, Wintjens R, Rooman M. Computer-aided methods for evaluating thermodynamic and thermal stability changes of proteins. Rec Res Dev Protein Engng 2001;1:277–90.
[17] Gromiha MM, Thomas S, Santhosh C. Role of cation–π interaction to the stability of thermophilic proteins. Prep Biochem Biotechnol 2002; 32:355–62.
[18] Selvaraj S, Kono H, Sarai A. Specificity of protein–DNA recognition revealed by structure-based potentials: symmetric/asymmetric and cognate/non-cognate binding. J Mol Biol 2002;322:907–15.
[19] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–42.
[20] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–97.
[21] Pabo CO, Sauer RT. Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem 1992;61: 1053–95.