Cation–π Interactions in Protein–Protein Interfaces

Peter B. Crowley[1]* and Adel Golovin[2]
[1]*Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal*
[2]*EMBL Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*

ABSTRACT **Arginine is an abundant residue in protein–protein interfaces. The importance of this residue relates to the versatility of its side chain in intermolecular interactions. Different classes of protein–protein interfaces were surveyed for cation–π interactions. Approximately half of the protein complexes and one-third of the homodimers analyzed were found to contain at least one intermolecular cation–π pair. Interactions between arginine and tyrosine were found to be the most abundant. The electrostatic interaction energy was calculated to be ~3 kcal/mol, on average. A distance-based search of guanidinium:aromatic interactions was also performed using the Macromolecular Structure Database (MSD). This search revealed that half of the guanidinium:aromatic pairs pack in a coplanar manner. Furthermore, it was found that the cationic group of the cation–π pair is frequently involved in intermolecular hydrogen bonds. In this manner the arginine side chain can participate in multiple interactions, providing a mechanism for inter-protein specificity. Thus, the cation–π interaction is established as an important contributor to protein–protein interfaces. Proteins 2005;59:231–239.**

Key words: protein complex; homodimer; cation–π interaction; recognition; crystal structure

INTRODUCTION

The analyses of crystal structures of proteins in various oligomeric states have revealed many of the features which determine the nature of the protein–protein interaction.[1–3] Such attributes as the interface size,[4] complementarity[5,6] and amino acid composition (hydrophobicity and polarity) contribute to the free energy change upon protein association. To date, three classes of protein–protein interfaces, homodimers,[7–11] protein complexes (permanent[8,11–14] and transient[15,16]) and crystal-packing interfaces[9,11,17,18] have been analyzed in detail. As might be expected, there is a correlation between the average characteristics of the protein interface and the oligomeric state of a protein. For example, the interface size, geometric complementarity and hydrophobicity increase going from crystal-packing interfaces to protein complexes to homodimers.[15,18]

Clear differences also exist for the interface propensities of amino acids in the various classes of protein interactions. Homodimers are particularly enriched in nonpolar and aromatic side chains while depleted in polar and charged side chains (except arginine).[7,10,11] Similar, though less-pronounced, trends are observed for protein complexes.[13,14] Crystal-packing interfaces on the other hand do not have a strong preference for a particular amino acid composition.[18] Thus, it can be concluded that homodimer interfaces generally resemble the protein interior, while crystal-packing interfaces are similar to the protein surface and protein complex interfaces are somewhere in between. The extent of interfacial hydrogen-bonding is similar in both homodimers and protein complexes, with an average of one hydrogen bond per ~100 Å^2 of interface area.[10,13,15] A slightly lower hydrogen bond frequency is observed in crystal-packing interfaces.[18] It has been shown that, on average, 20% of hydrogen bonds involve main-chain atoms only, while 40% involve main-chain–side-chain interactions.[10,13] Salt bridges are considerably less abundant than hydrogen bonds, occurring on average, with a frequency of about two per 1,000 Å^2 of interface.[11,12]

Surprisingly, arginine, unlike the other charged side chains, is not excluded from protein–protein interfaces. Although homodimer interfaces are similar to the protein interior, arginine is the second largest contributor after leucine.[7,10] In fact, arginine is one of the most abundant interface residues contributing about 10% on average to the interface area, regardless of the type of protein interaction [Fig. 1(A)].[7,10,14,18] The abundance of arginine in protein interfaces relates to the versatility of its side chain as a contributor to intermolecular interactions. The guanidinium group, for instance has a high capacity to donate hydrogen bonds, while the methylene groups can contribute favorably to the hydrophobic effect. Cation–π interactions with aromatic side chains are another potentially beneficial interaction at the protein–protein interface.

Numerous studies have reported the occurrence of cation–π interactions in protein structures[19–25] and in protein–ligand[23,26,27] and protein-DNA[28–30] complexes. These analyses have revealed the preferential (nonrandom) localization of amine groups in the vicinity of aromatic rings. In

---

The Supplementary Materials referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: Marie Curie Foundation, Fifth Framework Programme; Grant number: HPMF-CT-2002-02008.

*Correspondence to: Peter B. Crowley, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. Da República, Apartado 127, Oeiras, Portugal. E-mail: crowley@itqb.unl.pt

Received 23 March 2004; Accepted 9 November 2004

Published online 22 February 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20417

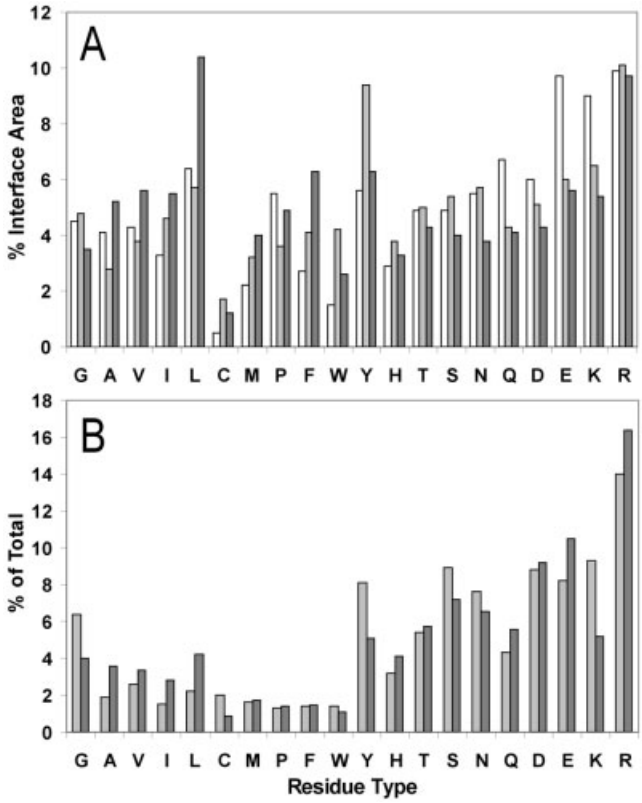

Fig. 1. **A**: The average abundance of amino acids in protein–protein interfaces of (white) large crystal-packing interfaces,[18] (light grey) protein complexes[14] and (dark grey) homodimers.[10] **B**: The average percentage contribution of amino acids to hydrogen bonding in the interfaces found in (light grey) protein complexes (63 interfaces, 588 hydrogen bonds) and in (dark grey) homodimers (115 interfaces, 2,026 hydrogen bonds).

particular, amines bearing positive charge are found to make favorable interactions with the π-electron cloud of aromatic side chains.[19,24] This interaction, electrostatic in nature,[31] is calculated to be more stabilizing than an analogous salt-bridge interaction.[32] Furthermore, the strength of the interaction is not strongly attenuated in water.[32] It is interesting to note that the presence of cation–π interactions in protein structures was (unwittingly) discovered by Warme and Morgan.[33] While performing an analysis of side chain interactions they observed that: "It [Phe] is surprisingly tolerant of -NH_3^+ atoms in its environment as are the other aromatic amino acids."[33]

In an effort to survey protein–protein interfaces for cation–π interactions we have used the methodology developed by Gallivan and Dougherty.[24] Furthermore, we have investigated the interaction geometry between guanidinium groups and aromatic side chains, using search criteria implemented within the Macromolecular Structure Database.[34]

METHODS

Structure Files

The coordinates of the protein crystal structures were retrieved from either the Protein Data Bank[35] or the Protein Quaternary Structure[36] server. The selection of protein complexes,[14,15] homodimers,[10] and crystal-packing interfaces[18] was based on previous analyses documented by Janin and coworkers and Thornton and coworkers. All of the structures had a resolution of ≤ 3 Å. The data sets for the homodimers and crystal-packing interfaces are nonredundant with less than 25% sequence identity between any two members. The list of protein complexes includes representatives of protease–inhibitor, antibody–antigen, enzyme complexes and signaling complexes. In order to extend the analysis of protein complexes a data set of hetero-complexes was extracted from a recently compiled list of structurally nonredundant interfaces.[37] This data set (crystal structures only) contains a diverse range of interfaces, including hetero-oligomers, multi-subunit protein complexes, nucleic acid binding complexes, and protein–peptide complexes (See Supplementary Material).

Polar Interactions

An implementation of the program HBPLUS,[38] available through the Protein–protein Interaction server (http://www.biochem.ucl.ac.uk/bsm/PP/server/), was employed for identifying hydrogen bonds in protein–protein interfaces. The standard geometric criteria were used for defining hydrogen bonds.[38]

Cation–π Interactions

The program CaPTURE[24] (Cation–π Trends Using Realistic Electrostatics) was used to identify interactions between the cationic group of lysine or arginine and the aromatic rings of phenylalanine, tyrosine and tryptophan. This program provides an energetic evaluation of potential cation–π interactions using a variant of the optimized potentials for liquid simulations[39] (OPLS) force field. The methodology has been reported in detail previously[24] and only the selection criteria are described here. In brief, those cation–π pairs within a 10-Å separation cut-off, which cannot accommodate a water molecule at closest contact, are considered "interacting pairs." If the OPLS electrostatic energy (E_{El}) is calculated as ≤ −2.0 kcal/mol the pair is counted as a cation–π interaction. If $E_{El} > -1.0$ kcal/mol, the pair is rejected. When $-2.0 < E_{El} \leq -1.0$ kcal/mol, the cation–π interaction is considered significant only if the van der Waals energy (E_{vdW}) ≤ −1.0 kcal/mol.

Database Surveys

The Macromolecular Structure Database[34] is a relational database for protein structure information and was used to search the PDB for interactions between the arginine guanidinium group and the aromatic rings of phenylalanine, tyrosine, and tryptophan. Interaction pairs were retrieved from protein–protein interfaces in structures of ≤ 3 Å resolution. The search criteria required that the center of the aromatic ring be ≤ 4.5 Å from either the C^{ζ} or the N^{η} atoms of arginine. Plane parameters were calculated using a least squares method, which gives reliable results with respect to the coordinate error in the PDB. Using MSD as an integrated sequence–structure resource,[40] the results were filtered to exclude structures

TABLE I. Classes of Protein–Protein Interfaces Analyzed for Cation–π Interactions

Class[a]	Sample size	Interface size[b] (Å^2)	Normalized areas[c]		Cation–π interactions	
			Arg	Ring	Interface[d]	Total[e]
Crystal-packing[18]	116	1510 (520)	0.4	0.4	14	23
Protein complex[14,15]	81	1910 (760)	0.5	0.9	40	68
Homodimers[10]	120	3880 (2200)	1.0	1.5	42	141
Hetero-complexes[36]	417	—		—	218	482
Total	734	—		—	314	714

[a]See Methods for a brief description of each of the classes.
[b]Average interface size (and standard deviation) for each data set.
[c]Derived from the average area contribution of interface residues, normalized with respect to the arginine content of homodimer interfaces. Ring refers to the combined contributions of tyrosine, phenylalanine, and tryptophan.
[d]The number of interfaces in which at least one cation–π pair was detected.
[e]The total number of cation–π pairs detected in each data set including the duplication arising from twofold symmetry.

in which both chains shared more than 50% sequence identity with another structure in the data set. Statistical analysis of the results was performed using the Oracle functions COUNT, AVG, and STDDEV.

RESULTS AND DISCUSSION

Interface Hydrogen Bonds

Figure 1(B) is a plot of the average percentage contribution of the amino acids to hydrogen bonding in interfaces found in protein complexes and homodimers. Given the larger data set and the (on average) larger interfaces the survey is more informative for homodimers. Nevertheless, it can be seen that the contribution to interface hydrogen bonding is related to the average abundance of the amino acids in interfaces [cf. Fig. 1(A)]. For example, Tyr, which contributes ~9% and ~6% to the interfaces found in protein complexes[14] and homodimers,[10] respectively, is involved in 1.5 × more hydrogen bonds in protein complexes [Fig. 1(B)]. Similarly, glycine and lysine are more important for hydrogen bonding in protein complexes than in homodimers. Leucine, on the contrary, makes twice as many hydrogen bonds in homodimers in agreement with its overall greater abundance (~10% of the interface area in homodimers[10] compared to ~5% in protein complexes[14]). The relatively large contribution made by glycine [Fig. 1(B)] can be rationalized on the basis of its structure. Given the absence of a bulky side chain, this residue is better disposed to participate in main-chain packing and hydrogen bonding. Arginine, which is involved in 14% and 16% of the hydrogen bonds [Fig. 1(B)] found in protein complexes and homodimers, respectively, is the most important amino acid in terms of polar interactions. It would seem therefore that the utility of the guanidinium group in hydrogen bond formation is part of the explanation for the abundance of arginine in protein–protein interfaces. Arginine can also form cation–π interactions the role of which in protein interfaces is explored in the following sections.

Cation–π Interactions

Gallivan and Dougherty have developed a program for identifying cation–π interactions in protein structures.[24] This program was used to detect such interactions in the interfaces of protein complexes, homodimeric proteins, and large crystal-packing interfaces (Table I). Representative examples of interface cation–π interactions are illustrated in Figure 2. It should be noted that cation–π interactions in homodimer interfaces are duplicated as a consequence of the interface symmetry [Fig. 2(A)]. Protein structures, which were found to contain interface cation–π interactions are listed in Table II (See also Supplementary Material).

The occurrence of such interactions in specific complexes is remarkably high. Approximately half of the protein complexes and one-third of the homodimers contain one or more cation–π pairs (Table II). In contrast, cation–π interactions are relatively rare (12% of the data set) in large crystal-packing interfaces (Table II). This is most likely a consequence of the nonspecific nature of the crystal packing interface. It should be noted also that phenylalanine and tryptophan have a low abundance in such interfaces [Fig. 1(A)]. If the total number of observed cation–π pairs is compared with the average interface size (Table I), it can be concluded that such interactions are about four times more common in protein complexes than in crystal-packing interfaces. (This calculation takes into account that the data set of protein complexes is 30% smaller). Such a comparison is less informative in the case of homodimer interfaces (note the large standard deviation, Table I). Nevertheless, taking the on-average larger interfaces into account, there are about three times more cation–π pairs in homodimer interfaces than in crystal-packing interfaces.

Another way of comparing the frequency of cation–π interactions is to look at the normalized interface area contributions of the participating residues. Table I lists the average area contribution of arginine and the aromatic residues to the different interface classes, normalized with respect to the arginine area of homodimer interfaces. In this way it can be seen, for example, that homodimer interfaces have, on average, about twice as much interface area composed of arginine compared to protein complexes and crystal packing interfaces. Similar portions of crystal packing interfaces are derived from arginine and the aromatic residues. Protein complexes, on the other hand, have about twice as much interface area composed of aromatic side chains. Statistically, therefore, twice as

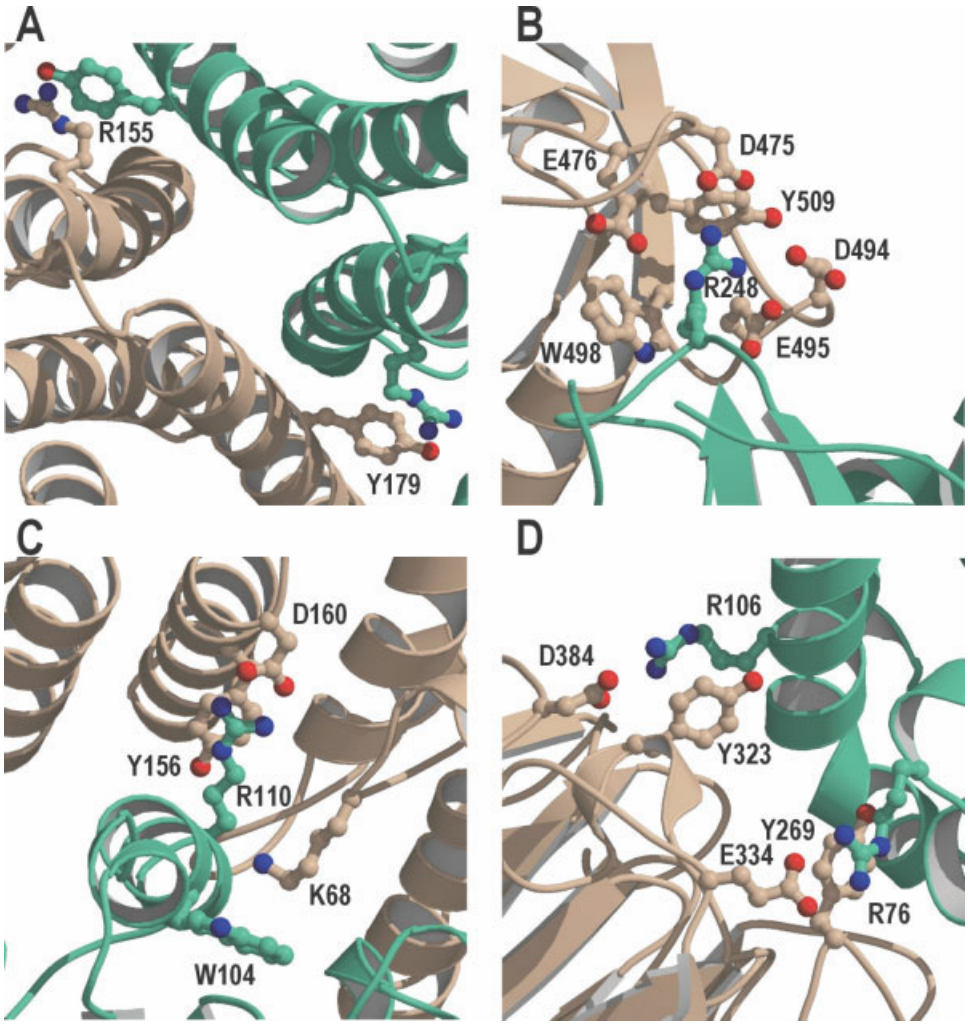

Fig. 2. Selected views of protein interfaces, depicting cation–π interactions. Protein chains are displayed as ribbons, while side chains of interest are represented in ball-and-stick. Side chain oxygen and nitrogen atoms are colored red and blue, respectively. **A**: The homodimer of farnesyl diphosphate synthase[53] (1uby) illustrates the duplication of cation–π pairs as a result of twofold symmetry. **B**: When P53 binds to 53BP2[43] (1ycs) the side chain of R248 is surrounded by a nest of aromatic and acidic side chains. The guanidinium group forms cation–π bonds with W498 and Y509 and a single salt bridge with Asp475. An additional cation–π interaction between R181 and Y424 is not shown. (For purposes of clarity, the loop residues 163–170 of P53 were omitted from the figure.) **C** and **D** correspond to the complexes of Ran-importin beta[54] (1ibr) and Ran-RCC1[55] (1i2m). In both cases arg:tyr cation–π pairs are augmented by salt bridges between the guanidinium group and acidic side chains. The symmetry of the interactions in Ran-RCC1 is particularly striking.

many cation–π interactions might be expected in this category compared to crystal interfaces. There are, however, four times as many such interactions (as calculated above), suggesting a specific role for the cation–π pair in protein complexes. By a similar rationale six times as many cation–π pairs might be expected in homodimer interfaces. There are in fact six times more interactions but the validity of this comparison is questionable considering the large standard deviation of the interface size in homodimers (Table I).

The abundance of particular cation–π pairs is related to the amino acid propensities in the different types of protein–protein interfaces. In agreement with the greater abundance of tryptophan and tyrosine in protein complexes [Fig. 1(A)], arginine:tryptophan and arginine:tyrosine cation–π pairs are more frequent in protein complexes than in homodimers (Table III). Conversely, there are more arginine:phenylalanine pairs in homodimers consistent with the higher abundance of phenylalanine in the homodimer interface.

Bahadur et al. have shown that the binding sites in large homodimer interfaces are generally composed of two or more patches. In fact, seventeen of the 122 homodimer interfaces were found to consist of between three and six patches.[10] Eleven of these interfaces with three-six patches were found to contain cation–π interactions, (Table II), indicating a preference for cation–π interactions in oligomers with large interfaces. This result agrees with a recent analysis of domain-swapped proteins by Dehouck et al. who found cation–π interactions in eight of thirteen investigated

TABLE II. PDB Entries of Protein Interfaces Containing Cation–π Interactions[†]

Large crystal-packing interfaces									
13pk	1b3j	1bea	1c02	1ilr*	1kwa*	1qha	1xgs	1aq0	1fvk
1i4g	1ome	1pva	1sw6						
Protein complexes[a]									
1aip	1ao7	1avw	**1bkd***	1brs	**1cly**	1cbw	1cho	1dvf*	**1efn**
1fbi*	**1fin**	1gg2	**1got**	1gua	**1he1**	1hia	1hwg*	**1i2m***	**1ibr***
1jhl*	1kb5*	1mel*	1mlc*	1nsn*	1osp*	1seb	**1tx4**	1udi	1vfb*
1wq1*	**1ycs***	**1zbd**	2btf	2kai	2ptc*	2sni	3hfl*	3hfm*	3sgb
Homodimers[b]									
1a4i	1ad3*	1ade	1af5	1aor	1b8a*	1b8j	1bd0	1biq	1bis
1bjw	1bmd	1brw*	1bsl	1cdc*	1chm	1csh	1cvu	1daa*	1ebh
1fip	1hxp*	1imb*	1isa	1jhg	1lyn	1mor	1nsy	1oac*	1pgt
1pre*	1r2f	1reg	1trk	1uby	1xso	2arc	2ilk	2lig	2nac*
3dap	8prk								

[†]Protein structures were analysed for cation–π interactions using the program CaPTURE.[24] Interfaces with more than one cation–π pair are marked with an asterisk.
[a]Transient protein complexes[15] are indicated in bold.
[b]Homodimer interfaces, which involve three to six patches[10] are underlined.

TABLE III. Cation–π Interactions in Protein–Protein Interfaces

				Energy (kcal/mol)[c]	
Interface class	Cation–π	Number[a]	H-bonding[b]	E_{El}	E_{vdW}
Crystal-packing interfaces	Arg:Tyr	5	3	−3.4 (1.7)	−2.5 (1.2)
	Arg:Phe	2	1	−2.0	−1.4
	Arg:Trp	1	1	−1.4	−3.2
	Lys:Tyr	4	2	−3.2 (1.0)	−0.8 (0.1)
	Lys:Phe	3	1	−3.2 (0.9)	−0.8 (0.1)
	Lys:Trp	1	—	−4.8	0.2[d]
Protein complexes	Arg:Tyr	30	20	−2.8 (1.0)	−2.4 (0.8)
	Arg:Phe	7	7	−3.1 (1.5)	−2.5 (0.7)
	Arg:Trp	16	8	−4.1 (1.9)	−2.3 (1.1)
	Lys:Tyr	8	3	−3.8 (1.8)	−0.7 (1.0)
	Lys:Phe	1	—	−2.7	−0.8
	Lys:Trp	6	4	−3.7 (1.3)	−0.8 (0.4)
Homodimers	Arg:Tyr	23	10	−2.9 (1.3)	−2.6 (0.9)
	Arg:Phe	11	8	−2.9 (0.9)	−2.3 (0.9)
	Arg:Trp	9	4	−3.9 (1.5)	−2.2 (1.0)
	Lys:Tyr	9	3	−3.3 (1.2)	−0.8 (0.6)
	Lys:Phe	—	—	—	—
	Lys:Trp	1	0	−3.2	−0.8

[a]The number of cation–π pairs in each interface class (excluding duplicates arising from twofold symmetry). It should be noted that some interfaces contain more than one cation–π pair (Table II).
[b]The number of cation–π pairs in which the charged residue is involved in one or more hydrogen bonds.
[c]The electrostatic and van der Waals energies as calculated in CaPTURE,[24] are reported as the average, with standard deviation in parenthesis, of all such cation–π pairs detected in each data set.
[d]The van der Waals energy is calculated to be positive due to steric clashes between the side chains.

structures.[41] Their observations suggest that the cation–π interaction can be important for sealing the edges between intertwined chains. Similarly, and in agreement with previous reports,[20,32] we have noted the tendency for cation–π interactions to occur on the edges of the interface.

Cation–π interactions involving lysine account for some twenty percent of the observed cation–π pairs in both protein complexes and homodimers (Table III). Lysine is less frequent in such interactions in agreement with the strong exclusion of its side chain from the interfaces in protein complexes and homodimers. Janin and coworkers have shown that lysine has a particularly low propensity to be found in the core of interfaces though it is relatively abundant at the rim of interfaces.[10,14] In large crystal-packing interfaces, cation–π pairs involving lysine are as frequent as those involving arginine. Compared with protein complexes and homodimers the interfaces involved in crystal-packing are less stringent in the exclusion of charged side chains.[18] The similar percentage contribution of arginine and lysine in crystal-packing interfaces[18] [Fig. 1(A)] can explain the similar occurrence of these side chains in cation–π interactions (Table III).

Interaction Geometry

The interaction geometries of arginine-mediated cation–π interactions were analyzed. In a previous study of

TABLE IV. Guanidinium-Aromatic Pair Interactions in Protein–Protein Interfaces[†]

	Percent contribution[b]					Average geometries[c]	
Interaction category[a]	Tyr	Phe	Trp^5	Trp^6	Percent total	α	*r*
Planar	22	17	7	5	51	16 (8)	4.1 (0.2)
Oblique	12	11	3	2	28	48 (9)	4.6 (0.2)
Orthogonal	9	8	2	2	21	77 (9)	4.9 (0.2)

[†]Interacting pairs were identified as those where the center of the aromatic ring was ≤4.5 Å from either the C^ζ or the N^η atoms of arginine (See Methods).
[a]Categories are defined as: planar ($0° < \alpha < 30°$), oblique ($30° < \alpha < 60°$), and orthogonal ($60° < \alpha < 90°$).
[b]A total of 910 interactions were observed and the contributions of the three aromatic residues are indicated as percentages of the total. Interactions with the five- and six-membered rings of Trp are distinguished as Trp^5 and Trp^6, respectively.
[c]The average inter-planar angle (α) and the average distance (*r*) from the ring center to the C^ζ atom of arginine. Standard deviations are reported in parentheses.

interplanar residue contacts, Brocchieri and Karlin found it convenient to partition the interplanar angle (α) into three categories, planar ($0° < \alpha < 30°$), oblique ($30° < \alpha < 60°$) and orthogonal ($60° < \alpha < 90°$).[22] When a similar analysis was applied in the present study, it was found that 53% of the cation–π interactions involved planar stacking between the guanidinium group and the aromatic ring. Planar interactions had an average α of 17 (8)° and an average distance (*r*) between the guanidinium C^ζ and center of the aromatic ring of 4.3 (0.6) Å. Twenty-six percent (26%) of the interactions belonged to the oblique category with an average $\alpha = 44$ (9)° and an average $r = 4.8$ (0.5) Å. The remaining 21% of the interactions were of the orthogonal type with an average $\alpha = 73$ (7)° and an average distance $r = 5.1$ (0.5) Å.

Using a relational database query system,[34] the PDB was searched for other examples of interactions between the guanidinium group and aromatic rings in protein interfaces (Table IV). Given that the search was based solely on distance criteria, not all pairs are expected to be energetically favorable cation–π interactions. Arginine:tyrosine pairs were found to be the most abundant, accounting for 43% of the total number of interactions. The inter-planar angle between the guanidinium group and the aromatic ring is plotted per number of interactions in Figure 3. The major peak around 15° indicates that coplanar packing of the guanidinium:aromatic groups is the favored interaction (51% of the data set, Table IV). This observation is in agreement with the results of previous studies.[22,24] In general, the separation distance (*r*) between the guanidinium group and the aromatic ring decreases as the interplanar angle decreases (Table IV). This can be expected since coplanar stacking leads to maximal (shortest distance) contact between the interacting groups.

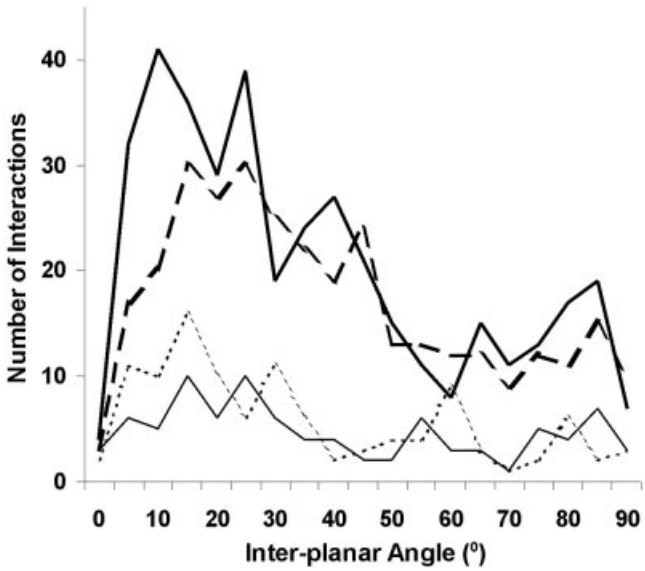

Fig. 3. The inter-planar angle observed between the guanidinium group and the aromatic ring of Tyr (heavy solid line), Phe (dashed line), and Trp (5-membered ring, dotted line, six-membered ring, light solid line) in a total of 910 interaction pairs.

Permanent vs. Transient

The interaction energy of protein complex formation varies in strength in accordance with the particular function of a complex. For example, the assembly of protease-inhibitor complexes is normally associated with a large free energy change. Protein complexes involved in signal transduction, on the other hand, are usually of lower affinity. Identifying the characteristics of the protein interface, which confer the ability to interact weakly or strongly has been a difficult task. In general, there is a trend of higher affinity with larger interfaces.[15] Polar interactions seem to be less indicative since Thornton and coworkers have reported similar values for the hydrogen-bond frequency in both permanent and low affinity complexes.[8,15] The geometric complementarity of the interacting surfaces can, however, provide some indication of the strength of the interaction. Interfaces in weak protein complexes have, on average, lower complementarity than strong protein complexes.[8,15] Crystal-packing interfaces also have poor geometric complementarity in accordance with their inherently low specificity.[14,18] The interfaces in electron transfer complexes, which represent an extreme in terms of low affinity and specificity,[42] have the poorest surface complementarity.[16]

The protein complexes analyzed in this work included twenty-three that are classified as low-affinity.[15] Interest-

ingly, twelve of these contained at least one cation–π interaction (Table I). Previously, it has been noted that some interfaces in electron transfer complexes also utilize cation–π interactions (Crowley and Carrondo[16] and references therein). Therefore, despite the low affinity of transient protein interactions, cation–π pairs are frequently employed at the interface. This might suggest that the cation–π interaction is more important for specificity rather than affinity during protein association. After the encounter complex has formed, the interacting partners must rearrange their relative orientations in search of the optimal binding configuration. Formation of enthalpically favorable cation–π interactions may provide an additional mechanism for selecting the lowest free-energy configuration.

This idea is supported by the fact that the majority of the cation–π pairs are also involved in intermolecular hydrogen bonding. In more than 60% of the interactions involving arginine, the arginine side chain also forms one or more hydrogen bonds with acceptor groups on the partner surface (Table III). Salt bridges were also observed, but less frequently (Fig. 2). However, acidic side chains within a 4.5 A cutoff of the guanidinium group were frequently observed. Sixty percent (60%) of the interfaces with cation–π pairs were found to have one or more cation-anion pairs. In more than half of these the arginine involved in the cation–π pair was also involved in cation–anion interactions. The multiplicity of such interactions is expected to ensure the specificity of binding. An interesting example is the complex of p53 bound to the ankyrin and SH3 domains of 53BP2.[43] Two arginines in this interface form cation–π interactions, one of which is surrounded by a cluster of acidic side chains [Fig. 2(B)]. It is plausible that this acidic patch provides long range attraction for the guanidinium group, which then positions itself to form cation–π interactions with two neighboring aromatic rings [Fig. 2(B)]. The beauty of this example can be appreciated further given that this region of the interface is known to be essential for binding and stability (Gorina and Pavletich[43] and references therein).

This multiplicity of side-chain interactions gives rise to the question of side-chain packing, which has been assessed by various methods.[4–6,13,44] While the packing density in protein interiors[4] and interfaces is high[13] it has been demonstrated by theory[44] and experiment[45] that side-chain interactions are not rigidly defined. Rather than fitting together like pieces in a jigsaw puzzle, the side chains are sufficiently flexible to allow plasticity in the packing.[44] Nevertheless, it has also been shown that amino acids can have well-defined propensities for packing against particular neighbors.[46,47] The cation–π interaction is an example of such a propensity between cationic and aromatic side chains.

Polar Interactions and Energetics

Given the intrinsically low resolution of protein structures, the analysis of polar interactions is limited by inaccuracies in the atomic coordinates.[13] However, the analysis of large data sets allows the overall trends to be identified. Cation–π interactions involving arginine are calculated to contribute, on average, 3.3 (1.4) kcal/mol of electrostatic free energy in favor of binding in protein complexes and homodimers (Table III). In order to extend the analysis, a larger data set of hetero-complexes[37] was investigated (417 structurally nonhomologous, interfaces). More than half the structures were found to contain at least one cation–π pair (Table I, See also Supplementary Material) with an average of 3.0 (1.5) kcal/mol of electrostatic energy for interactions involving arginine.

Although the contribution of intermolecular hydrogen bonds to the binding affinity is debatable,[48] the abundance of such interactions in protein interfaces has been demonstrated.[8,11–15] Salt bridges, however, are relatively rare[11,12] and both theoretical[49] and experimental[50] studies indicate that the cost of desolvation of charged groups outweighs the potential benefit of Coulombic interactions. Cation–π interactions, on the other hand, involve the association of an aromatic ring and a charged group, typically arginine. Gallivan and Dougherty have suggested that desolvation of the former is likely to be beneficial, thus explaining why cation–π interactions are more common than salt bridges in protein structures.[32] However, this treatment neglects the fact that aromatic side chains have the ability to form π-hydrogen bonds with water molecules.[51]

Neutron diffraction studies on the guanidinium ion (structurally homologous to the arginine side chain) provide additional insight into the favorable energetics of cation–π formation. Mason et al. have shown that this cation is poorly hydrated as a consequence of both its low charge density and the incompatibility of the planar, sp^2 NH_2 groups with the local tetrahedral structure of water.[52] Therefore, the burial of an arginine side chain is not expected to be accompanied by a desolvation penalty. On the contrary, it may be favorable since distortions in the local water structure are removed. The accessible surface area (ASA) of the guanidinium group of arginine ($\sim$150 Å^2) was measured in all of the cation–π pairs observed in homodimer interfaces. An average ASA of 27 (21) Å^2 was calculated indicating an $\sim$80% decrease in solvent accessibility with respect to a free arginine side chain. In one-fifth of the cases the guanidinium group was fully buried (< 5 Å^2 ASA). The relative ease of arginine burial not only facilitates cation–π interactions but is a likely explanation for the abundance of arginine-mediated hydrogen bonds in protein–protein interfaces [Fig. 1(B)].

CONCLUSIONS

The predominance of arginine in protein–protein interfaces can be related to its high capacity to donate hydrogen bonds. In addition, it has been shown that cation–π interactions involving arginine are common in the interfaces of protein complexes and homodimers. This interaction can make a significant contribution to the binding energy of protein complex formation and should be included in the list of criteria for characterizing protein interfaces. Furthermore, it has been shown that fifty percent of the guanidinium:aromatic pairs pack together in a coplanar arrangement. The knowledge that cation–π

interactions are important in protein–protein interfaces is potentially useful for protein docking studies. Half of the protein complexes analyzed contained at least one such interaction. Therefore, the presence of cation–π interactions could be used as a means of discriminating chemically relevant docking results from false positives.

ACKNOWLEDGMENTS

P.B.C. is grateful for funding from the Marie Curie Foundation under the Fifth Framework Programme (HPMF-CT-2002-02008). The referees are thanked for their constructive comments. Acknowledged are: O. Keskin for providing a list of protein–protein interfaces, M.A. Carrondo for critically reading the manuscript and J.A.R. Worrall for helpful discussions.

SUPPLEMENTARY MATERIAL

A list of 417 PDB entries of hetero-complexes, which were surveyed for intermolecular cation–π interactions.

REFERENCES

1. Stites WE, Protein-protein interactions: Interface structure, binding thermodynamics, and mutational analysis. Chem Rev 1997;97: 1233–1250.
2. Jones S, Thornton JM. Analysis and classification of protein-protein interactions from a structural perspective. In: Kleanthous C, editor. Protein-protein recognition. New York: Oxford University Press; 2000. p 33–59.
3. Wodak SJ, Janin J. Structural basis of macromolecular recognition. Adv Protein Chem 2003;61:9–73.
4. Richards FM. Areas, volumes, packing and protein structure. Annu Rev Biophys Bio 1977;6:151–176.
5. Lawrence MC, Colman PM. Shape complementarity at protein-protein interfaces. J Mol Biol 1993;234:946–950.
6. Laskowski RA. Surfnet—a program for visualizing molecular surfaces, cavities and intermolecular interactions. J Mol Graph 1995;13:323–330.
7. Janin J, Miller S, Chothia C. Surface, subunit interfaces and interior of oligomeric proteins. J Mol Biol 1988;204:155–164.
8. Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
9. Ponstingl H, Henrick K, Thornton JM. Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins 2000;41:47–57.
10. Bahadur RP, Chakrabarti P, Rodier F, Janin J. Dissecting subunit interfaces in homodimeric proteins. Proteins 2003;53:708–719.
11. Ponstingl H, Kabir T, Gorse D, Thornton JM. Morphological aspects of oligomeric protein structures. Prog Biophys Mol Biol 2004. Forthcoming.
12. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. Protein Sci 1997;6:1426–1437.
13. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–2198.
14. Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. Proteins 2002;47:334–343.
15. Nooren IMA, Thornton JM. Structural characterisation and functional significance of transient protein-protein interactions. J Mol Biol 2003;325:991–1018.
16. Crowley PB, Carrondo MA. The architecture of the binding site in redox protein complexes: implications for fast dissociation. Proteins 2004;55:603–612.
17. Ponstingl H, Kabir T, Thornton JM. Automatic inference of protein quaternary structure from crystals. J Appl Crystallogr 2003;36:1116–1122.
18. Bahadur RP, Chakrabarti P, Rodier F, Janin J. A dissection of specific and non-specific protein-protein interfaces. J Mol Biol 2004;336:943–955.
19. Burley SK, Petsko GA. Amino-aromatic interactions in proteins. FEBS Lett 1986;203:139–143.
20. Flocco MM, Mowbray SL. Planar stacking interactions of arginine and aromatic side chains in proteins. J Mol Biol 1994;235:709–717.
21. Mitchell JBO, Nandi CL, McDonald IK, Thornton JM, Price SL. Amino-aromatic interactions in proteins—is the evidence stacked against hydrogen bonding. J Mol Biol 1994;239:315–331.
22. Brocchieri L, Karlin S. Geometry of interplanar residue contacts in protein structures. Proc Natl Acad Sci USA 1994;91:9297–9301.
23. Ma JC, Dougherty DA. The cation-pi interaction. Chem Rev 1997;97:1303–1324.
24. Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. Proc Natl Acad Sci USA 1999;96:9459–9464.
25. Minoux H, Chipot C. Cation-pi interactions in proteins: can simple models provide an accurate description? J Am Chem Soc 1999;121: 10366–10372.
26. Meyer EA, Castellano RK, Diederich F. Interactions with aromatic rings in chemical and biological recognition. Angew Chem Int Edit 2003;42:1210–1250.
27. Biot C, Buisine E, Rooman M. Free-energy calculations of protein-ligand cation-pi and amino-pi interactions: From vacuum to protein-like environments. J Am Chem Soc 2003;125:13988–13994.
28. 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.
29. Rooman M, Lievin J, Buisine E, Wintjens R. Cation-pi/H-bond stair motifs at protein-DNA interfaces. J Mol Biol 2002;319:67–76.
30. Lamoureux JS, Maynes JT, Glover JNM. Recognition of 5″-YpG-3′ sequences by coupled stacking/hydrogen bonding interactions with amino acid residues. J Mol Biol 2004;335:399–408.
31. Deakyne CA, Meot-Ner M. Unconventional ionic hydrogen-bonds. 2. NH+…pi− complexes of onium ions with olefins and benzene derivatives. J Am Chem Soc 1985;107:474–479.
32. Gallivan JP, Dougherty DA. A computational study of cation-pi interactions versus salt bridges in aqueous media: implications for protein engineering. J Am Chem Soc 2000;122:870–874.
33. Warme PK, Morgan RS. A survey of amino acid side chain interactions in 21 proteins. J Mol Biol 1978;118:289–304.
34. Boutselakis H, Dimitropoulos D, Fillon J, Golovin A, Henrick K, Hussain A, Ionides J, John M, Keller PA, Krissinel E, and others. E-MSD: the European Bioinformatics Institute Macromolecular Structure Database. Nucleic Acids Res 2003;31:458–462.
35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
36. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci 1998;23:358–361.
37. Keskin O, Tsai CJ, Wolfson H, Nussinov R. A new, structurally non-redundant, diverse data set of protein-protein interfaces and its implications. Protein Sci 2004;13:1043–1055.
38. McDonald IK, Thornton JM. Satisfying hydrogen bonding potentials in proteins. J Mol Biol 1994;238:777–793.
39. Jorgensen WL, Maxwell DS, Tirado-Rives 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: 11225–11236.
40. Golovin A, Dimitropoulos D, Oldfield T, Rachedi A, Henrick K. MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites. Proteins 2005;58: 190–199.
41. Dehouck Y, Biot C, Gilis D, Kwasigroch JM, Rooman M. Sequence-structure signals of 3D domain swapping in proteins. J Mol Biol 2003;330:1215–1225.
42. Crowley PB, Ubbink M. Close encounters of the transient kind: redox protein interactions in the photosynthetic redox chain investigated by NMR spectroscopy. Acc Chem Res 2003;36:723–730.
43. Gorina S, Pavletich NP. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. Science 1996;274:1001–1005.
44. Bromberg S, Dill KA. Side chain entropy and packing in proteins. Protein Sci 1994;3:997–1009.
45. Gassner NC, Baase WA, Matthews BW. A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proc Natl Acad Sci USA 1996;93:12155–12158.
46. Singh J, Thornton JM. SIRIUS—an automated method for the

analysis of the preferred packing arrangements between protein groups. J Mol Biol 1990;211:595–615.
47. Bahar I, Jernigan RL. Coordination geometry of non-bonded residues in globular proteins. Fold Des 1996;1:357–370.
48. Baldwin RL. In search of the energetic role of peptide hydrogen bonds. J Biol Chem 2003;278:17581–17588.
49. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science 1995;268:1144–1149.
50. Hendsch ZS, Jonsson T, Sauer RT, Tidor B. Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. Biochemistry 1996;35:7621–7625.
51. Steiner T, Koellner G. Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. J Mol Biol 2001;305:535–557.
52. Mason PE, Neilson GW, Dempsey CE, Barnes AC, Cruickshank JM. The hydration structure of guanidinium and thiocyanate ions: implications for protein stability in aqueous solution. Proc Natl Acad Sci USA 2003;100:4557–4561.
53. Tarshis LC, Proteau PJ, Kellogg BA, Sacchettini JC, Poulter CD. Regulation of product chain length by isoprenyl diphosphate synthases. Proc Natl Acad Sci USA 1996;93:15018–15023.
54. Vetter IR, Arndt A, Kutay U, Gorlich D, Wittinghofer A. Structural view of the Ran-importin beta interaction at 2.3 angstrom resolution. Cell 1999;97:635–646.
55. Renault L, Kuhlmann J, Henkel A, Wittinghofer A. Structural basis for guanine nucleotide exchange on Ran by the regulator of chromosome condensation (RCC1). Cell 2001;105:245–255.