PROTEINS: Structure, Function, and Bioinformatics 67:1179–1184 (2007)

APPLICATION NOTE

Cation–Aromatic Database

A. Srinivas Reddy, G. Madhavi Sastry, and G. Narahari Sastry*
Molecular Modeling Group, Organic Chemical Sciences, Indian Institute of Chemical Technology, Hyderabad 500007, India

ABSTRACT **Cation–aromatic database (CAD) is a publicly available web-based database that aims to provide further understanding of interaction between a cation and the π-moiety of the aromatic residues in protein data bank (PDB). All the metal ions and basic amino acid residues are considered as cations. The motifs are defined on the basis of distance between the cation and the centroid of the aromatic ring. This generalized way of defining the motifs enables a thorough inspection and analysis of the type of interactions in proteins. Of these pairs, those having the cation within a cylinder with the arene ring as a base are identified. This geometric criterion enables the identification of strong cation–π interactions. A tool to identify the interactions in a user-given protein is also added to the database. CAD is freely accessible via the Internet at http://203.199.182.73/gnsmmg/databases/cad/ Proteins 2007;67:1179–1184.**

Key words: aromatic amino acids; cation–π interactions; statistical analysis; metal interactions

INTRODUCTION

The regulation of metal ions plays a major role in enzymes to catalyze a range of biological reactions involved in signaling pathways, maintaining the electrochemical gradient for ion transport, and translocation of proteins.[1–5] Identification and characterization of the metal ion binding sites and their selectivity have received immense attention in the past few decades.[6–9] It is evident from the earlier studies that metal ions can bind to aromatic groups in a covalent as well as noncovalent fashion. The noncovalent interactions between the metal ions and an aromatic ring, which are considered as a strong cation–aromatic interactions, are increasingly recognized as an important binding force relevant to structural biology.[10–12]

Number of computational studies and carefully designed experimental studies were performed to understand the strengths and structural features of the cation–aromatic complexes.[13–16] To assess the nature of these interactions, Gallivan and Dougherty used an energy-based criterion to identify favorable interactions in proteins.[10] On the basis of this, Crowley and Golovin found that at least half of the protein complexes and one third of homodimers contained cation–π pairs.[17] Biot et al. defined cation–π interactions in protein–ligand complexes using a distance and an angle criterion.[18] The interaction energies were evaluated using ab initio quantum chemistry calculations by reducing the cation–π pairs to simplified systems.[19,20] Theoretical and experimental studies have shown that these interactions can be quite strong, both in the gas phase and aqueous media.[21–26]

However, in most of these studies, the cation is the side chain protonated nitrogens of the basic amino acid residues, but understanding of the metal ion interactions is rather scarce. The binding of metal ion has an option to complex σ or π fashion, especially in the case of heteroaromatics. Therefore, understanding the nature of M-aromatic interactions is key to realize and model various processes in metalloproteins. The purpose of the present study is to develop and present an exhaustive database of M-aromatic motifs found in proteins present in the protein data bank (PDB) based on relevant geometrical criteria and obtain meaningful statistics. M-aromatic motifs include both σ as well as π interaction geometries in the database. Such a database helps in screening and developing new methods for identifying and ranking these interactions. Statistical analysis has been carried out on the cation–aromatic pairs, to assess the frequency of occurrence of various metal ions in cation–aromatic interactions.

METHODS

The PDB database is searched for the instances wherein the cation is located within 6.0 Å from the cent-

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

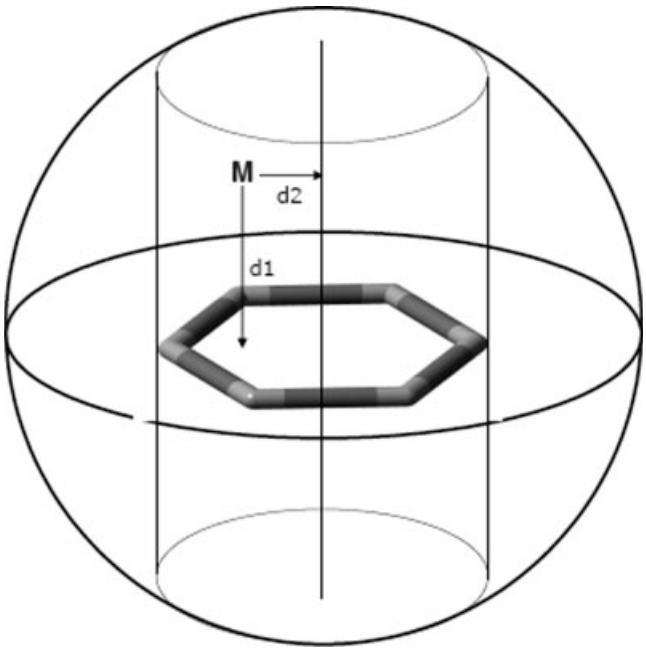
Grant sponsor: CSIR; IICT communication no. 060326.

G. Madhavi Sastry's current address is Informatics division, GVK Biosciences Pvt Limited, Hyderabad, India.

*Correspondence to: G. Narahari Sastry, Molecular Modeling Group, Organic Chemical Sciences, Indian Institute of Chemical Technology, Hyderbad - 500007, India. E-mail: gnsastry@iict.res.in or gnsastry@yahoo.com

Received 2 May 2006; Revised 26 July 2006; Accepted 8 August 2006

Published online 21 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21202



Scheme 1.

roid of the aromatic ring. This, we call it as a cation–aromatic motif, and since they all occur in a sphere of radius 6.0 Å from the centroid, we name them as the motifs present inside the sphere. A lower cut-off distance of 2.0 Å is taken to avoid the covalent bonds. The aromatic systems include the aromatic side chains of the residues tryptophan (Trp), tyrosine (Tyr), phenylalanine (Phe), and histidine (His). All the metal ions present in the PDB are considered as cations apart from the protonated basic amino acid residues lysine (Lys) and arginine (Arg). However, His can act either as cation or as an aromatic moiety depending on its protonation state, and in our study, both the possibilities are considered.

To depict the cation–π motifs, a virtual cylinder is considered below and above the plane of the aromatic ring (Scheme 1). The height of the cylinder is equal to the radius of the sphere (6.0 Å). The base of the cylinder is taken as a circle having a radius equivalent to the distance between the centroid and any atom of the aromatic ring. Obviously, this is also the radius of the cylinder. In case of an indole ring, the two 5- and 6-membered rings are treated individually as a π-system. This includes cation–π motifs, which are obtained by treating indole as a single entity without loss of generality. Moreover, defining the base of the cylinder by considering the entire indole would give certain irrelevant motifs. Thus, in the construction, the two rings in the indole have been considered independent of each other. If the cation is present inside the cylinder, it is considered as a strong cation–π interaction, and we call it as a motif inside the cylinder.

If the motif is inside the cylinder, we set a flag CYL_CHK to "YES" in the database, otherwise flag is set to "NO." However, it should be noted that the cation–π interactions wherein the cation is located outside the cylinder may also be significant.[10] Thus, while the cylindrical approach gives purely cation–π motifs, the sphere approach also include the σ-type interaction of the cation with the aromatic ring.

The position of the cation is specified with two distances $d1$ and $d2$ considering the centroid of the aromatic ring as the origin. $d1$: Perpendicular distance from the cation to the plane of the aromatic ring; $d2$: perpendicular distance from the cation to the principle axis of the aromatic ring (Scheme 1).

A total of 34,917 protein structures from PDB (updated till 1st February 2006) are considered and with a FORTRAN program, we have evaluated all the possible cation– aromatic interactions. The database is organized in two tables, where one table contains the details of the proteins viz. pdb-id, class of the protein, date of deposition, resolution. The second table consists of the information about the cation–π motifs present in each of the protein. For each of these motifs, the details like name of the cation and aromatic ring involved and their corresponding residue-id and chain-id along with the distances from cation to each of the aromatic ring atoms, centroid, and the perpendicular distances to the aromatic plane ($d1$) and the principle axis ($d2$) are provided.

The database is implemented in MYSQL (version 3.23.54) on RedHat 9.0 operating system. HTML and Java servlets with TOMCAT 5.4 are used in the front-end design and development, respectively. The front-end provides the user to find all the cation–aromatic interactions present in specified protein (Fig. 1). Many filters like resolution, class, aromatic, cation, cation–centroid distance ($d1$ and $d2$) have been provided to trim the output with according to requirement. An online tool is also made available to calculate the cation–aromatic interactions present in the protein inputted by the user.

RESULTS AND DISCUSSIONS

There are about 6,124,083 cation–aromatic motifs present in the database. Among them, about 6% (369,493) of the motifs are present inside the cylinder. Percentages of total motifs involving the aromatic side chain of Phe, Tyr, and Trp are 22.63, 22.37, and 19.47, respectively, whereas for His it is 35.51. In case of the motifs inside the cylinder, these side chains of the four aromatic amino acids are almost equal ranging from 24 to 26% of the total motifs. The analysis carried out by Gallivan and Dougherty shows that over one-fourth of the typtophans in the PDB experience energetically significant cation–π interactions.[10] Similarly, Ruan and Rodger using guided ion beam tandem mass spectrometry observed that the binding of Na^+ or K^+ with aromatic residues is more favored for Trp, followed by Tyr and Phe.[16] The database analysis of cylindrical motifs shows that Trp involved are more in number relative to Phe or Tyr till the cation–centroid distance of almost 4 Å, beyond which the Phe also increases. However, in the case of spherical motifs, the domination of Trp residues is not seen. This implies that Trp is the pref-

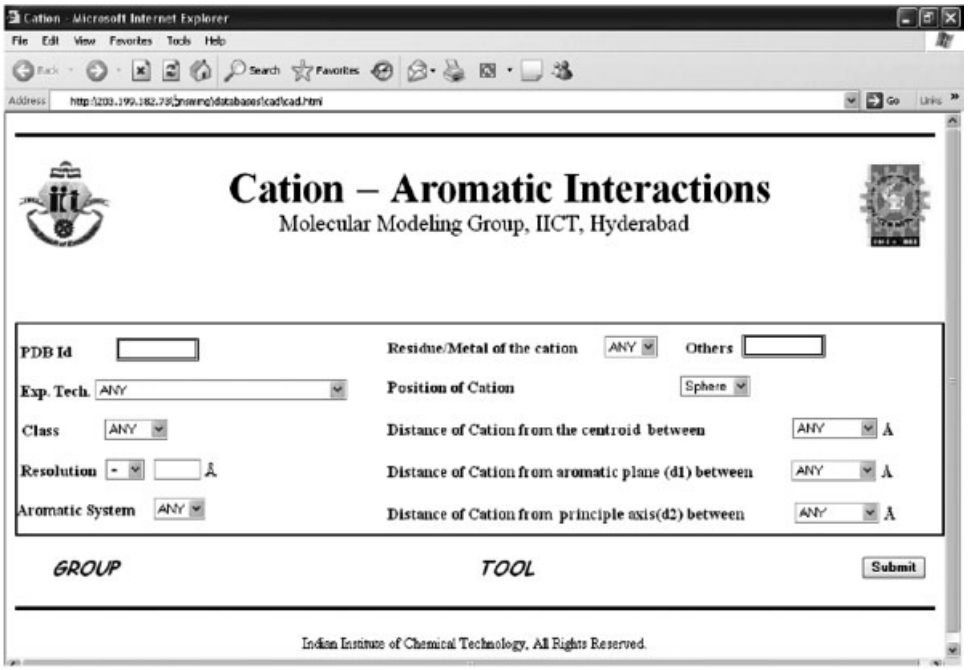


Fig. 1. The snapshot of the web interface to the database.

erential residue for cation–π interactions and also Trp prefers a π interaction to σ and distorted σ interaction. Quantum mechanical studies also illustrate that the indole ring, an aromatic moiety of Trp residue, has high interaction energy with metal ions.[19] In case of motifs of various metal cations, a total of 37 different metal cations are involved in 446,253 cation–aromatic interactions in all the proteins of PDB. Among them, the cations of 18 different metals viz. lithium (Li), potassium (K), rubedium (Rb), beryllium (Be), aluminum (Al), cerium (Ce), lead (Pb), antimony (Sb), barium (Ba), ruthenium (Ru), yttrium (Y), gadolinium (Gd), uranium (U), indium (In), iridium (Ir), palladium (Pd), samarium (Sm), and silver (Ag) do not have any single motif inside the cylinder. The cations of cobalt (Co), tungsten (W), nickel (Ni), aurum (Au), tellurium (Te), chromium (Cr), and strontium (Sr) have less than 5 motifs in the cylinder. Zinc (Zn) cation has the highest 61.8% proportion of motifs, followed by iron (Fe) 19.58%, copper (Cu) 6.40%, selenium (Se) 4.33%, and calcium (Ca) 2.93% among the total spherical motifs, involving metal cations. The motifs of magnesium (Mg), mercury (Hg), sodium (Na), arsenic (As), manganese (Mn), and cadmium (Cd) are present in smaller proportions. Lanthanum (La) even though has 250 spherical motifs, they are from only three proteins (pdb ids: 1DJG, 1KQV, and 1NKF) that to 240 are only present in 1NKF which is an NMR structure with 30 models, and each model has a lanthanum (III) cation, which has interaction with the tyrosine residues present in the same model and other models. The cations of the basic residues Lys, Arg, and His are involved in 92.7% of the total cation–aromatic motifs. Among the motifs present in cylinder, about 98.5% involve the Lys, Arg, and His cations. Among these three cations, His contributed in higher percentage (35.1%), followed by Lys (31.6%) and Arg (25.89%). In the work presented by Crowley and Golovin, it was found that cation–π interactions involving Arg are common in the interfaces of protein complexes.[17] Similarly, Biot et al. analyzed proteins bound to ligand molecules containing a nucleic acid base as the π system indicated the dominance of Arg as the cation counterpart.[18] In our studies, we find that the number of cylindrical motifs with Arg is more within the cation–centroid distance range between 2.5 and 4 Å. However, Lys predominates beyond 4 Å. Moreover, the total number of motifs containing Lys within 6 Å is larger than those containing Arg. This implies that the marked preference for Arg at the protein interfaces decreases when considering the whole protein. However, the preference for Arg still exists.

The number of cation–aromatic motifs increases as the distance between the cation and the centroid of the aromatic ring increases (Fig. 2). If we consider $d1$ distance ranges, the number of motifs are picking up between the 3 and 3.5 Å distance ranges and decreasing gradually with the increase of cation–aromatic distance. In cylinder, the motifs are increasing with the cation–centroid or cation–aromatic distance and are high in the distance range of 3.5–4.0 Å, except for the motifs involving the imidazole moiety of the histidine as the aromatic. The number of motifs involving histine as aromatic is high at shorter distance ranges and decreases with the increase of cation–centroid or cation–aromatic distance. This is due to the preference of imidazole ring to form σ motifs with the cations, as electron rich nitrogen atom is present in its ring. Earlier, quantum mechanical calculations indicate the preference of histidine to involve in σ-type interac-

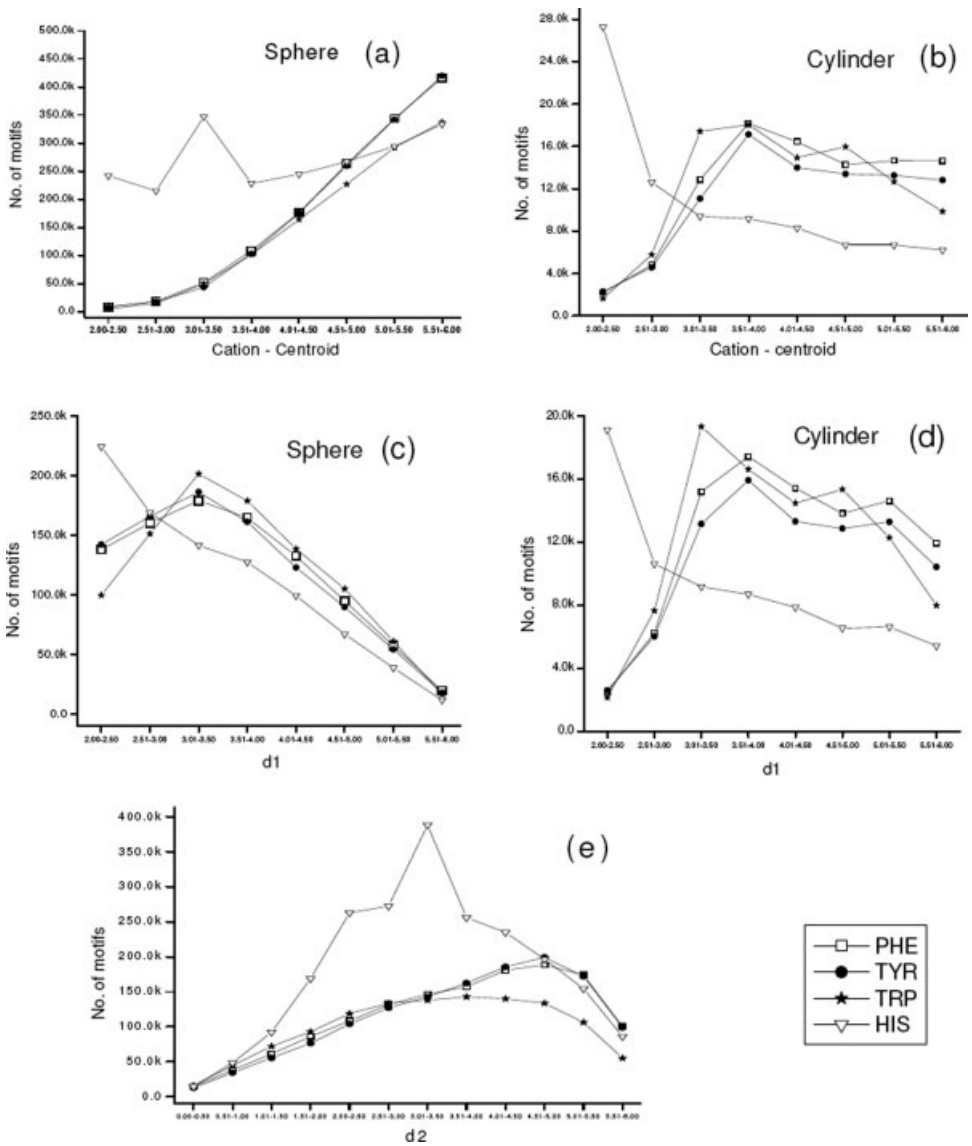


Fig. 2. Statistics of the motifs of the considered aromatics in the cation distance ranges from the centroid (**a** and **b**), aromatic plane (**c** and **d**) and principle axis (**e**) of the aromatic ring. (a) and (c) correspond to the motifs in sphere. (b) and (d) correspond to the motifs in cylinder.

tion.[27,28] Database analyses indicate a similar trend. The difference in the number of spherical motifs and cylindrical motifs is largest for His as an aromatic indicating its marked preference for the σ-type interaction. Moreover, in 57% of its total motifs, the cation is histidine itself in its protonated form. If we consider the variation of number of motifs with distance $d2$ for various amino acid residues, it increases gradually with the maximum between 3.0 and 3.5 Å, followed by a steady decrease. Individually for each aromatic, the peak is observed at 4.5–5.0 Å for Phe, Tyr and 3.5–4.0 Å for Trp and 3.0–3.5 Å for His.

The number of motifs involving the metal cations is observed to vary with the cation–centroid, $d1$ and $d2$ distances. (Fig. 3) With the increase in the cation–centroid distance, the number of motifs of the Zn, Fe, Cu metal ions gradually reach a maximum (∼45–55%) at 3.0–3.5 Å, and for other metal cations the peak is observed at 5.0–5.5 Å. The motifs inside the cylinder have their maximum number inbetween 3.5 and 5.5 Å. If we consider the motifs according to the distance $d1$ for metal ions, the maximum number of motifs for Zn, Fe, Cu, Ca, Mg, and Na is in 2.0–2.5 Å range if consider the spherical motifs, whereas it is maximum at 4.0–5.5 Å in case of cylindrical motifs. This indicates the preference of these metal cations to form σ-type of motifs over the π-type. For the Hg, As, Mn, and Cd, the motifs are high in the $d1$ range of 3–4.5 Å. The number of motifs increases with distance $d2$ and maximum between 3.0 and 3.5 Å, except for Se, Ca, Na, and Mn. For these metal ions, the maximum number of motifs is at a $d2$ range of 5.0–6.0 Å.

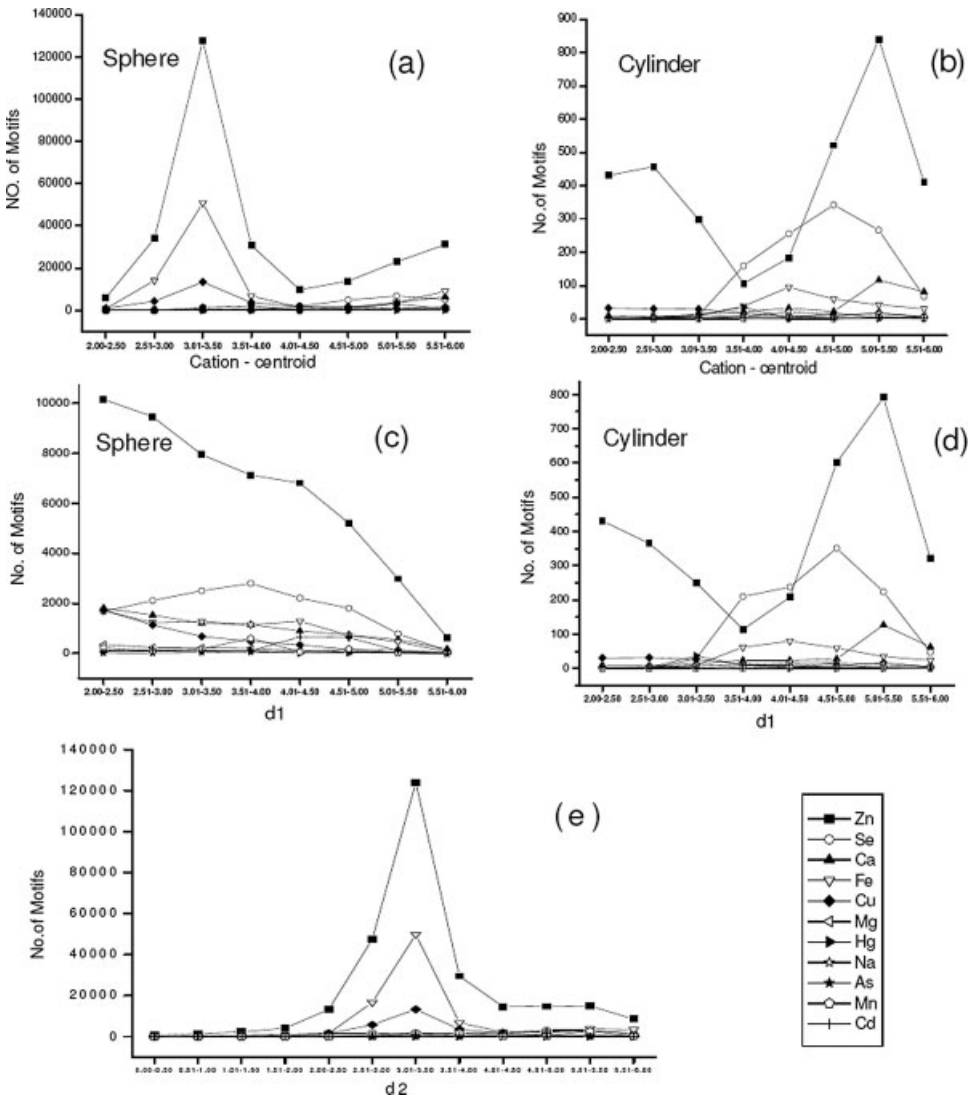


Fig. 3. Statistics of the motifs of the various cations of the metals in the cation distance ranges from the centroid (**a** and **b**), plane (**c** and **d**) and principle axis (**e**) of the aromatic ring. (a) and (c) correspond to the motifs in sphere, (b) and (d) correspond to the motifs in cylinder. A zoomed picture of Figure 3 is available in supporting information.

CONCLUSIONS

The comprehensive database presents exhaustive information on the cation– aromatic interactions in the proteins and also furnishes the frequency of their occurrence around the aromatic motif at a given distance. The statistical analysis reveals that the aromatic side of the histidine moiety prefers to bind in a σ fashion, whereas the rest shows propensity to bind in π-fashion. The predominance of σ-type interaction in His moiety may be traced to the presence of electron deficient nitrogen atom. Trp is dominating in the cation–π distance ranges, followed by Phe and Tyr. Even though most of the cation–aromatic interactions are contributed by basic amino acid residue cations Lys, Arg, and His, metal ions too have significant number of cation–aromatic interactions (cation–σ and cation–π). In the cations of basic amino acid residues, Arg cation has in the cation–π interaction distance ranges. Among all the metal cations. Zn followed by Fc has more cation–aromatic interactions. An on-line tool has been incorporated in the database, which furnishes all the cation–π interactions for any new protein. It is planned that the freely accessible public domain database will be maintained by our group and will be updated periodically.

REFERENCES

1. Suelter CH. Enzymes activated by monovalent cations: patterns and predictions for the enzyme-catalysed reactions are explored. Science 1970;168:789–795.
2. McRee DE. Living with metal ions. Nature Struct Biol 1998;5: 8–10.

3. Agranoff DD, Krishna S. Metal ion homeostasis and intracellular parasitism. Mol Microbiol 1998;28:403–412.
4. Pyle AM. Metal ions in the structure and function of RNA. J Biol Inorg Chem 2002;7:679–690.
5. Chaturvedi UC, Shrivastava R. Interaction of viral proteins with metal ions: role in maintaining the structure and functions of viruses. FEMS Immunol Med Microbiol 2005;43:105–114.
6. Ma JC, Dougherty DA. The cation–π interaction. Chem Rev 1997;97:1303–1324.
7. Wouters J. Cation–π (Na^+-Trp) interactions in the crystal structure of tetragonal lysozyme. Prot Sci 1998;7:2472–2475.
8. Bindu PH, Sastry GM, Sastry GN. Characterization of calcium and magnesium binding domains of human 5-lipoxygenase. Biochem Biophys Res Commun 2004;320:461–467.
9. Chourasia M, Sastry GM, Sastry GN. Proton binding sites and conformational analysis of H^+K^+-ATPase. Biochem Biophys Res Commun 2005;336:961–966.
10. Gallivan JP, Dougherty DA. Cation–pi interactions in structural biology. Proc Natl Acad Sci 1999;96:9459–9464.
11. Meyar EA, Castellano RK, Diederich F. Interactions with aromatic rings in chemical and biological recognition. Angew Chem Int Ed 2003;42:1210–1250.
12. Paulini R, Müller K, Diederich F. Orthogonal multipolar interactions in structural chemistry and biology. Angew Chem Int Ed 2005;44:1788–1805.
13. Wall SLD, Meadows ES, Barbour LJ, Gokel GW. Synthetic receptors as models for alkali metal cation–π binding sites in proteins. Proc Natl Acad Sci 2000;97:6271–6276.
14. Hu J, Barbour LJ, Gokel GW. Probing alkali metal–π interactions with the side chain residue of tryptophan. Proc Natl Acad Sci 2002;99:5121–5126.
15. Masson E, Schlosser M. *p*-Arene/Metal Binding: an issue not only of structure but also of reactivity. Org Lett 2005;7:1923–1925.
16. Ruan C, Rodgers MT. Cation–π Interaction: structures and energetics of complexation of Na^+ and K^+ with the aromatic amino acids, phenylalanine, ayrosine, and tryptophan. J Am Chem Soc 2004;126:14600–14610.
17. Crowley PB, Golovin A. Cation–π interactions in protein–proteininterfaces. Proteins: Struct Funct Bioinform 2005;59:231–239.
18. Biot C, Buisine E, Kwasigroch J-M, Wintjens R, Rooman M. Probing the energetic and structural role of amino acid/nucleobase Cation–π interactions in protein–ligand complexes. J Biol Chem 2002;227:40816–40822.
19. Reddy AS, Sastry GN. Cation [M=H^+, Li^+, Na^+, K^+, Ca^{2+}, Mg^{2+}, NH_4^+ and NMe_4^+] interactions with the aromatic motifs of naturally occurring amino acids: a theoretical study. J Phys Chem A 2005;109:8893–8903.
20. Reddy AS, Vijay D, Sastry GM, Sastry GN. From subtle to substantial: role of metal ions on π–π Interactions. J Phys Chem B 2006;110:2479–2481.
21. Vaden TD, Lisy JM. Competition between cation–π interactions and intermolecular hydrogen bonds in alkali metal ion-phenol clusters. II. Phenol trimer. J Chem Phys 2006;124:214315–214318.
22. Zhu W, Luo X, Puah CM, Tan X, Shen J, Gu J, Chen K, Jiang H. The multiplicity, strength, and nature of the interaction of nucleobases with alkaline and alkaline earth metal cations: a density functional theory investigation. J Phys Chem A 2004;108:4008–4018.
23. Priyakumar UD, Sastry GN. Cation–π interactions of curved polycyclic systems: M^+ (M=Li and Na) ion complexation with buckybowls. Tetrahedron Lett 2004;44:6043–6046.
24. Priyakumar UD, Punnagai M, Krishnamohan GP, Sastry GN. A computational study of cation–π interactions in polycyclic systems: exploring the dependence on the curvature and electronic factors. Tetrahedron 2004;60:3037–3043.
25. Gromiha MM, Suwa M. Structural analysis of residues involving cation–π interactions in different folding of membrane proteins. Int J Mol Biol 2005;35:55–62.
26. Gokel GW. The aromatic sidechains of amino acids as neutral donor groups for alkali metal cations. Chem Commun 2003;2847–2852.
27. Reddy AS, Vijay D, Sastry GM, Sastry GN. From subtle to substantial: role of metal ions on π–π interactions. J Phys Chem B 2006;110:10206–10207 (Reply).
28. Chelli R, Procacci P. From subtle to substantial: role of metal ions on π–π Interactions. J Phys Chem B 2006;110:10204–10205 (comment).