

RESEARCH ARTICLE

Analyzing the cation-aromatic interactions in proteins: Cation-aromatic database V2.0

Y. Bhargav Kumar^{1,2} | Nandan Kumar¹ | Lijo John¹ | Hridoy Jyoti Mahanta^{1,2} |
S. Vaikundamani¹ | Selvaraman Nagamani^{1,2} | G. Madhavi Sastry³ |
G. Narahari Sastry^{1,2} 

¹Advanced Computation and Data Sciences Division, CSIR-North East Institute of Science and Technology, Jorhat, Assam, India

²Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh, India

³Schrödinger Inc., HITEC City, Hyderabad, Telangana, India

Correspondence

G. Narahari Sastry, Advanced Computation and Data Sciences Division, CSIR-North East Institute of Science and Technology, Jorhat 785006, Assam, India.
(gnsastry@gmail.com)

Funding information

Department of Biotechnology, Ministry of Science and Technology, India, Grant/Award Number: BT/PR40188/BTIS/137/27/2021

Abstract

The cation-aromatic database (CAD) is a comprehensive repository of cation-aromatic motifs found in experimentally determined protein structures, first reported in 2007 [Proteins, 2007, 67, 1179]. The present article is an update of CAD that contains information of approximately 27.26 million cation-aromatic motifs. CAD uses three distance parameters (r , $d1$, and $d2$) to determine the position of the cation relative to the centroid of the aromatic residue and classifies the motifs as cation- π or cation- σ interactions. As of June 2023, about 193 936 protein structures were retrieved from Protein Data Bank, and this resulted in the identification of an impressive number of 27 255 817 cation-aromatic motifs. Among these motifs, spherical motifs constituted 94.09%, while cylindrical motifs made up the remaining 5.91%. When considering the interaction of metal ions with aromatic residues, 965 564 motifs are identified. Remarkably, 82.08% of these motifs involved the binding of metal ions to the amino acid HIS. Moreover, the analysis of binding preferences between cations and aromatic residues revealed that the HIS-HIS, PHE-ARG, and TRP-ARG pairs exhibited a preferential geometry. The motif pair HIS-HIS was the most prevalent, accounting for 19.87% of the total, closely followed by TYR-LYS at 10.17%. Conversely, the motif pair TRP-HIS had the lowest occurrence, representing only 4.20% of the total. The data generated help in revealing the characteristics and biological functions of cation-aromatic interactions in biological molecules. The updated version of CAD (Cation-Aromatic Database V2.0) can be accessed at <https://acds.neist.res.in/cadv2>.

KEYWORDS

amino acids, cation, cation-aromatic motifs, proteins

1 | INTRODUCTION

The cation-aromatic database (CAD) is a resource that provides comprehensive information about the cation-aromatic motifs in protein structures. The first version of CAD was introduced in 2007, with about 6 124 083 cation-aromatic motifs from 34 917 protein structures in the Protein Data Bank (PDB).^{1,2} It has provided valuable

insights into the prevalence and characteristics of cation-aromatic motifs. Since its inception, CAD has played a pivotal role in unraveling the structural and functional implications of cation-aromatic interactions.³⁻⁸

The understanding of cation- π interactions has significantly advanced due to the tremendous development in the structural characterization techniques, other experimental advances, and

computational studies. Computations played a critical role in deciphering the roles of structural and energetic factors and their manifestation in controlling the macromolecular assembly. Exploring various aspects of cation- π interactions, such as dependence on curvature and electronic factors,⁹ energy decomposition analysis,¹⁰ solvation effects,¹¹ complexation with metal ions,¹² contrasting preferences between nitrogen- and phosphorous-substituted rings,¹³ and structural and energetic preferences of cation binding.¹⁴ Furthermore, explicit solvent effects on cation- π interactions,¹⁵ size dependence of cyclic and acyclic π -systems,¹⁶ contrasting structural and energetic characteristics of bare and coordinatively saturated metal ions,¹⁷ and analysis of aromatic π -networks in proteins¹⁸ have been investigated^{19–27} that have collectively contributed to an enhanced understanding of cation- π interactions in diverse biological and chemical contexts.

Thus, over the past decade, our understanding of protein structures and cation-aromatic interactions has advanced significantly, necessitating an update to the CAD. Figure 1 represents the number of articles published over time on studying the cation- π interactions and highlighting their nature and role in the protein structure and function.^{8–20} This manuscript presents the updated version of CAD (CAD V2.0), which represents a major update since the original release, incorporating advancements in methodologies and a substantially expanded dataset. CAD V2.0 now encompasses an extensive collection of over 193 936 proteins, ensuring a comprehensive coverage of cation-aromatic interactions across diverse protein structures and families. Furthermore, we present a comprehensive analysis of the distribution, prevalence, and characteristics of cation-aromatic motifs based on the enriched dataset within CAD V2.0. The updated database is accompanied by improved search functionalities, enhanced performance, and increased security measures, offering researchers an even more robust and user-friendly platform. CAD V2.0, which is a greatly improved and expanded version, provides not only the data augmentation of the last 15 years but also has many

new features, in the categories of data acquisition, filtering, and classification. Clearly, expansion of CAD leads to comprehensive and exhaustive repository and the data generated are of tremendous value for bringing out novel structure-function correlations, primarily guided by cation- π interactions.

2 | IDENTIFICATION AND CHARACTERIZATION OF CATION-AROMATIC MOTIFS

The cation-aromatic motifs in the CAD were generated using a custom FORTRAN code developed in-house. The code was modified further to be able to account for the changes in the PDB file format after the release of the earlier version of CAD. The protocol details can be found in the first version.¹ Scheme 1 illustrates the workflow for extracting the geometrical features of the motifs and applying classification filters to update the database. The position of the cationic moiety relative to the arene is determined using three parameters namely, cation-centroid (of the arene) distance (r), perpendicular distance to the cation from the molecular plane of the aromatic ring (d_1), and distance between the cation and the principal axis of the arene (d_2) as shown in Scheme 2. The identified motifs are classified as either spherical motifs (within a virtual sphere and outside the cylinder) or cylindrical motifs (those lie within a cylinder). Cations located inside the cylinder are categorized as cation- π interactions, whereas spherical motifs are recognized as cation- σ interactions.

3 | UPDATION AND ANALYSIS

The first version of CAD was developed using MySQL, and for the current updated version, we have moved to PostgreSQL 14 database

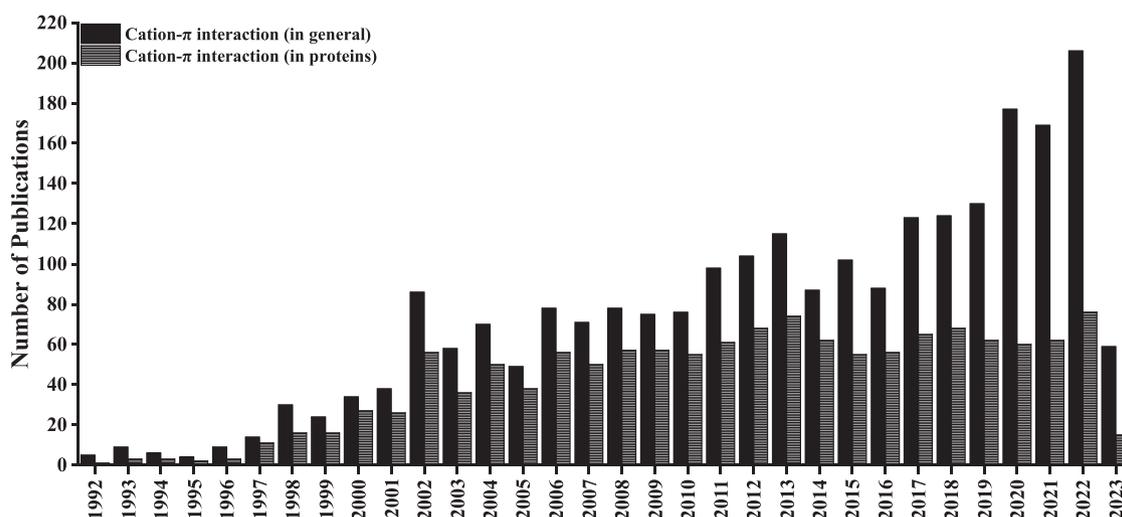
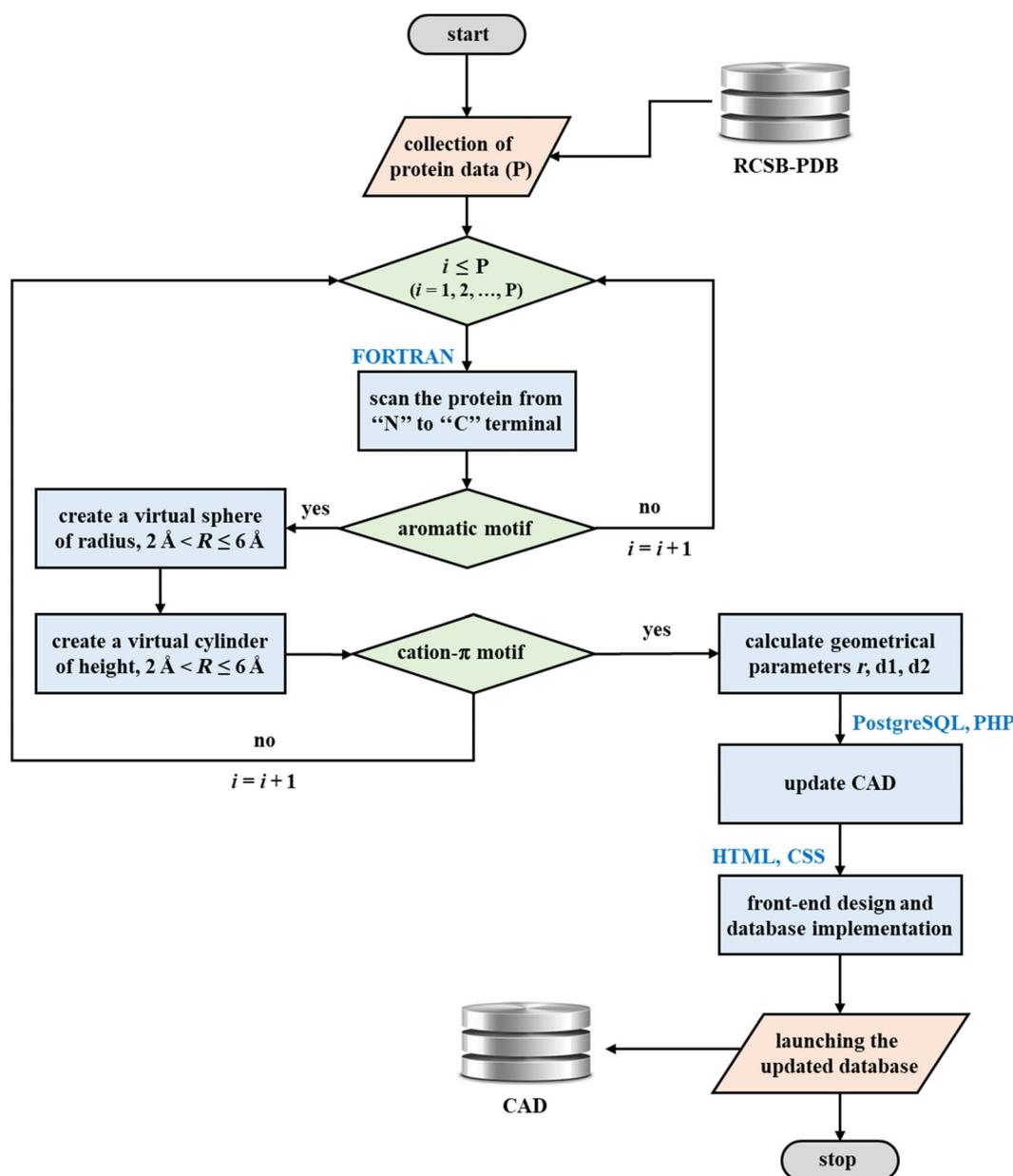


FIGURE 1 Number of publications related to cation- π interaction in general and in proteins from 1992 to 2023, obtained from Web of Science.

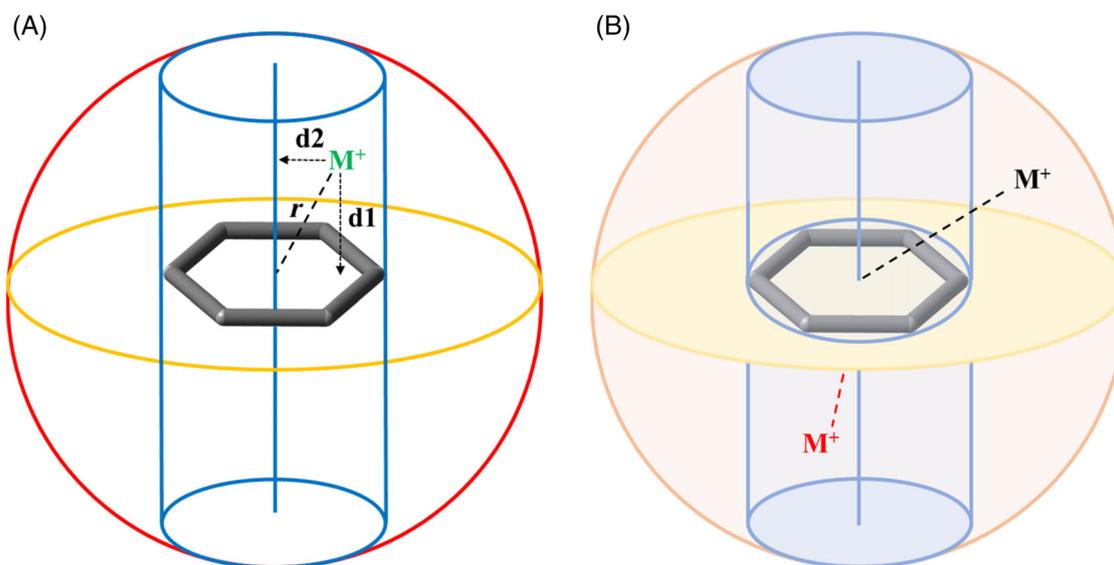


SCHEME 1 Workflow for updating the cation-aromatic database (CAD).

server, installed and configured on the Cent 7.9 Linux Operating System. This transition to PostgreSQL brings several advantages, including enhanced performance, flexibility, and improved security for researchers utilizing the database. To ensure efficient organization and retrieval of data, each identified motif is meticulously stored in a dedicated 20-column table within the CAD. This table includes crucial information such as the PDB ID, cation details (residue, atom type, chain, and chain ID), cation-aromatic distances, aromatic residue information, cation position (spherical/cylindrical motif, $d1$, and $d2$), protein class, as well as additional details like resolution and experimental techniques. Notably, all cations identified within the cylinder are categorized as "putative cations" while those outside the cylinder are not included in this category. To facilitate user-friendly access to the

CAD, we have developed a graphical user interface (GUI) frontend. This frontend enables users to retrieve information using various refinement filters accessible through the advanced search page. The frontend design incorporates HTML, PHP, JavaScript, and CSS, and is hosted on an Apache server. The backend database server is seamlessly connected to the frontend user interface using the PHP to PostgreSQL connector, ensuring smooth and reliable data retrieval and interaction.

The attractive interaction between the aromatic side chains of residues (PHE, TYR, TRP, and HIS) and cationic moieties of amino acids (ARG, LYS, and HIS) or metal ions leads to the formation of cation-aromatic motifs. Our analysis encompassed the examination of all 193 936 proteins, resulting in the identification of 27 255 817



SCHEME 2 Schematic representation of the (A) virtual sphere and cylinder with a representative cation (M^+) specified by distances r , $d1$, and $d2$ (B) spherical (black) and cylindrical (red) cation-aromatic motifs.

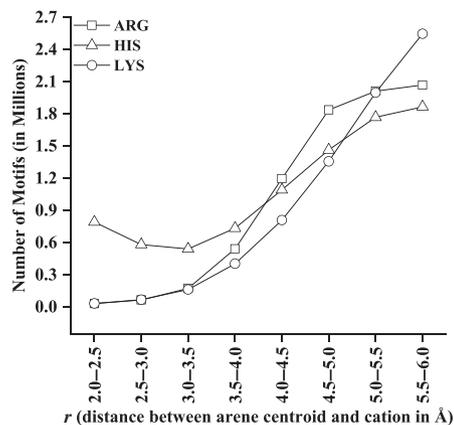
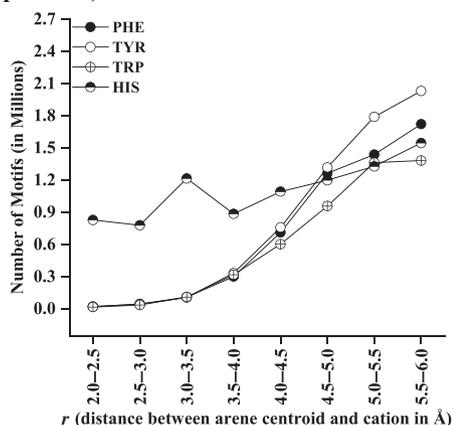
cation-aromatic motifs. Among these motifs, 94.09% (25 644 677) motifs were identified as spherical motifs while the remaining 5.91% (1 611 140) were identified as cylindrical motifs. Tables S1–S9 provide the number and distribution of spherical and cylindrical motifs based on distance parameters. The analysis reveals that the number of motifs increases gradually with increasing $d1$ values in spherical motifs involving only amino acid side chain residues, reaching a maximum between 2.5 and 4.5 Å, followed by a decline. Metal cations show a decrease in motifs with increasing cation-centroid distance, indicating a preference for strong cation-aromatic interactions. Cylindrical motifs formed by amino acid side chains primarily occur after 3.0 Å, suggesting a specific spatial arrangement for these interactions. The distribution of motifs based on $d2$ follows a Gaussian pattern, with a significant number of motifs observed between 2.0 and 5.0 Å. Notably, there is a sharp increase up to 2.0 Å, followed by a rise and subsequent decrease beyond 4.5 Å. Inorganic metal cations, particularly zinc and iron, dominate the motifs abundantly available between 2.5 and 4.0 Å, suggesting a preference for cation- π or cation- σ interactions at an average distance from the aromatic residues (Figures 2–4, and S1). Consequently, the abundance of spherical motifs (cation- σ interaction) surpasses that of cylindrical motifs (cation- π interaction), emphasizing the stabilization of proteins primarily through cation- σ interactions.

4 | DISTRIBUTION OF AMINO ACID RESIDUES IN CATION-AROMATIC MOTIFS

We investigated the distribution of aromatic residues (PHE, TYR, TRP, and HIS) and cationic residues (ARG, HIS, and LYS) in the

identified cation-aromatic motifs. The abundance of these residues is depicted in Figure 5. Figure 6 depicts the representative structures of the 12 cation-aromatic motifs formed from the combination of the arene and basic amino acid residues, while Table S10 provides the actual number of motifs for aromatic and cationic residues. A perusal of Figure 5 reveals that the most abundant cationic amino acid is HIS (36.71%), followed by ARG (32.90%) and LYS (30.38%), collectively accounting for 25 741 650 (94.44%) of all cation-aromatic motifs. The majority of ARG residues (93.60%) are found in spherical motifs, while a smaller portion (6.40%) is present in cylindrical motifs. Similarly, 93.55% of motifs containing HIS and 94.32% of motifs with LYS are observed to be spherical. Among the aromatic residues, HIS is the most abundant (33.87%), followed by TYR (24.97%), PHE (22.17%), and TRP (18.99%). Interestingly, only a small percentage (1%–2%) of each aromatic residue is found in cylindrical motifs, indicating a higher preference for cation- σ interactions rather than cation- π interactions. We further explored the combinations of aromatic and basic amino acid pairs in cation-aromatic motifs whose frequencies of occurrence are given in Table 1. The most frequent pair observed was HIS-HIS (19.87%), followed by LYS-TYR (10.77%) and ARG-TYR (9.31%) as given in Figure 7. This can be attributed to the unique properties of the imidazole ring in HIS, which can form both hydrogen bonding and π -cation interactions, making it versatile in forming stable interactions with the aromatic rings of other amino acid residues, and its higher relative abundance compared with the other amino acid residues considered here. HIS acting as a cation or an arene has shown different trends compared with their other counterparts in terms of their relative abundance, which may be traced to their size and compactness (Figures 2–4). While LYS-TYR and ARG-TYR are less

(A) (Spherical)



(B) (Cylindrical)

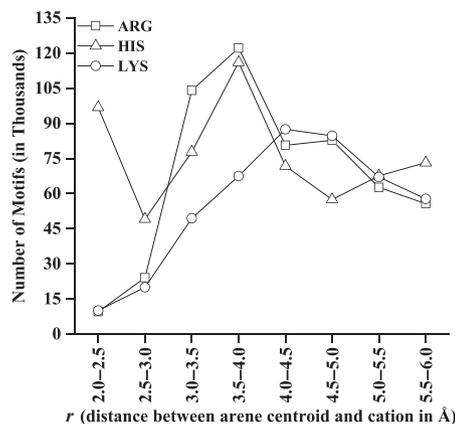
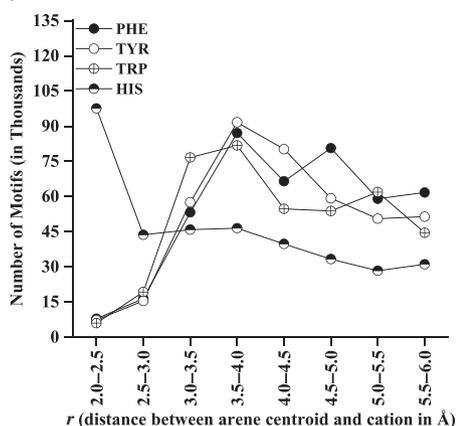
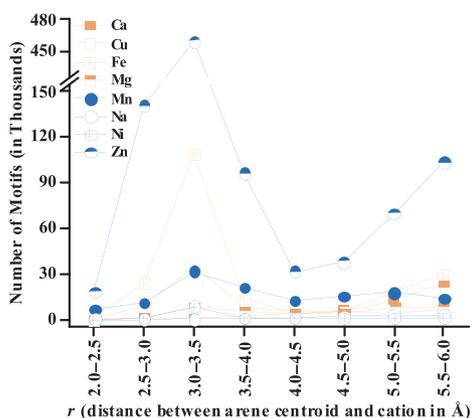
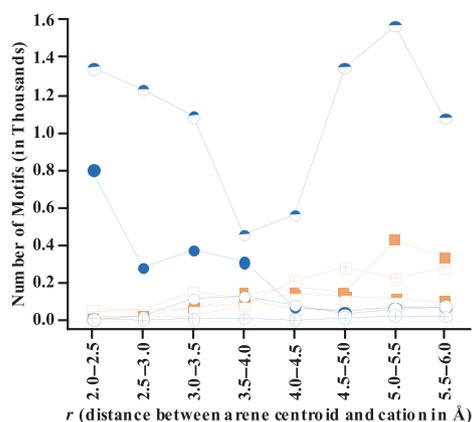


FIGURE 2 The distribution of aromatic and cationic residues in (A) spherical and (B) cylindrical motifs according to their cation-centroid distances.



(A) (Spherical)



(B) (Cylindrical)

FIGURE 3 The distribution of most common metal cations in (A) spherical and (B) cylindrical motifs according to their cation-centroid distances.

frequent, they are still significant. Other pairs are less common, possibly due to the size and structural constraints of the residues

involved, as well as specific requirements of the structure and function of proteins.

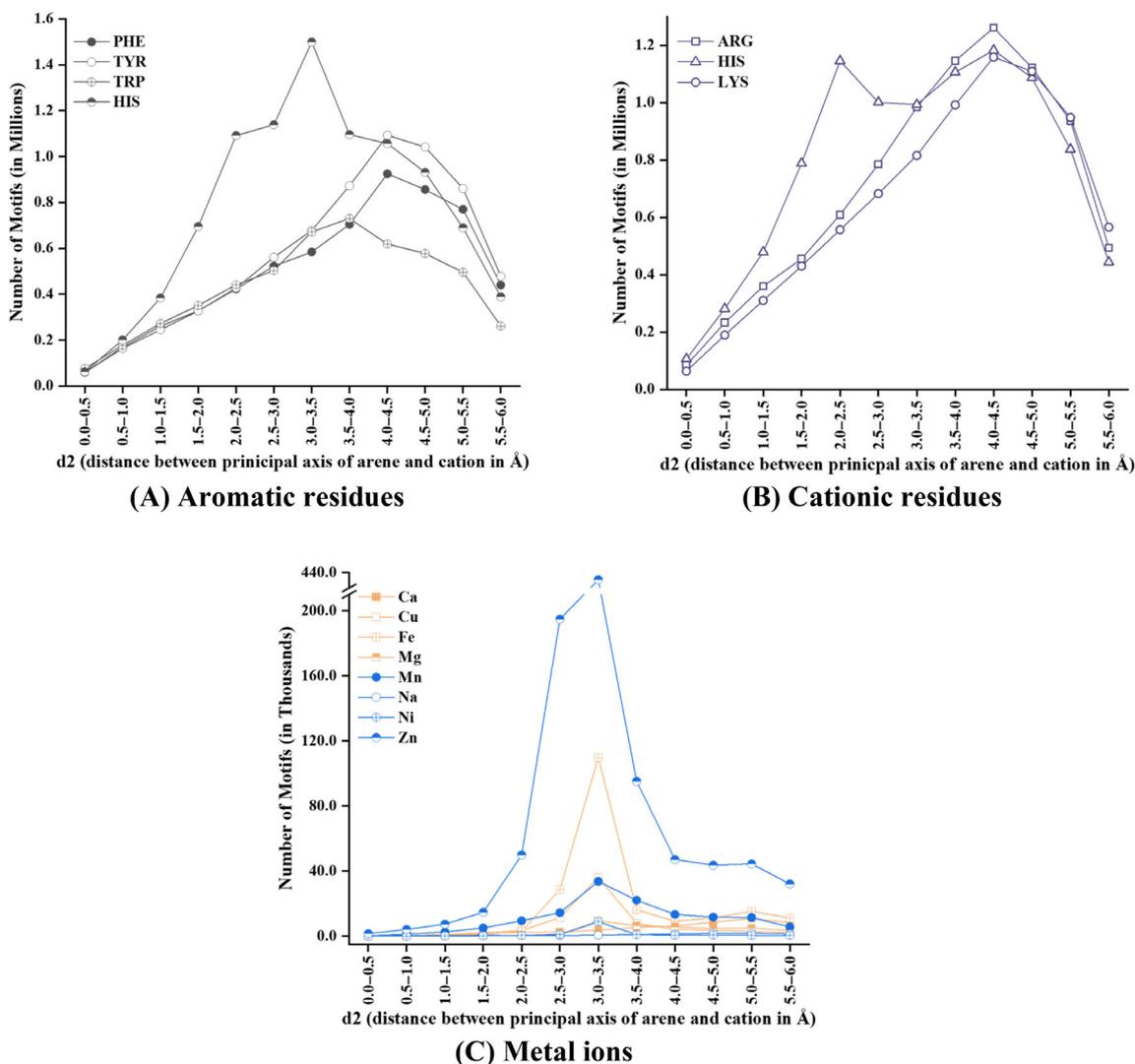


FIGURE 4 The distribution of (A) aromatic residues, (B) cationic residues, and (C) metal ions in cation-aromatic motifs with respect to the d_2 distances.

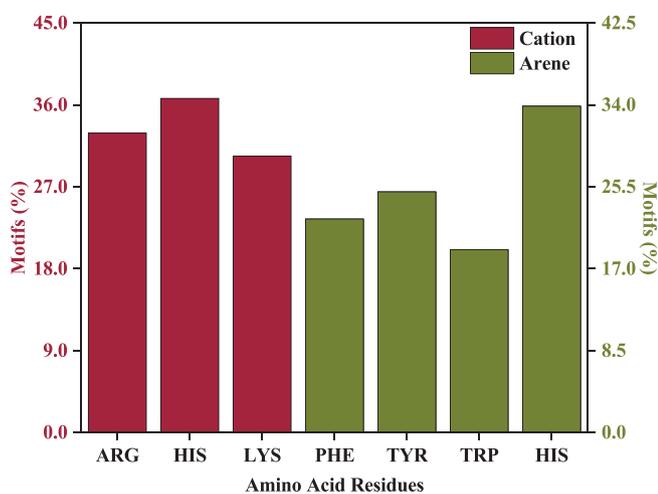


FIGURE 5 The distribution of different cationic and aromatic amino acids in the cation-aromatic motifs observed in proteins.

5 | DISTRIBUTION OF METAL CATIONS IN CATION-AROMATIC MOTIFS

The cation-aromatic motifs include various metal ions interacting with aromatic residues. We have identified 16 metal ions in addition to the three basic amino acids. For the database analysis, the oxidation state information of the metal ions was excluded, and only the names of the metal ions were considered in the analysis. Figure 8 provides an overview of the frequency of occurrence of metal ion-arene motifs, as well as the percentage occurrence of different aromatic residues and metal ions with each arene in the cation-aromatic motifs. Table S11 presents the abundance of metal ions in terms of the number of motifs, with Hg showing the highest occurrence among the cations with fewer than 2000 motifs. Among the identified motifs, 0.97% of cylindrical motifs and 5.84% of spherical motifs contain metal ions as cations, making up 1.04% and 5.56% of the total motifs, respectively. The analysis revealed that the majority of

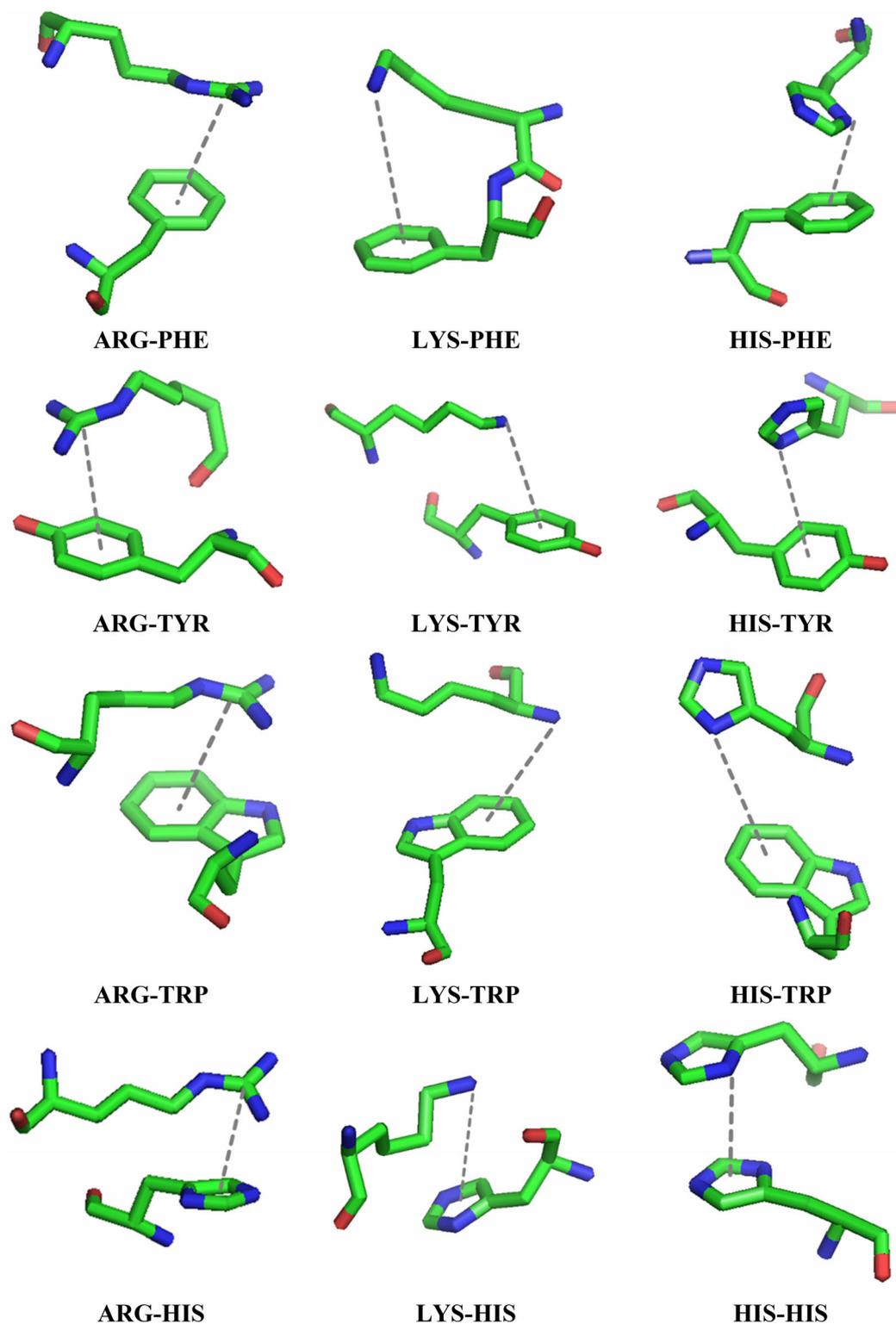


FIGURE 6 The representative structures of the 12 cation-aromatic pairs formed between three basic amino acids (ARG, LYS, HIS) and four aromatic amino acids (PHE, TYR, TRP, HIS).

metal ions (82.08%) bind to HIS, and mostly the binding is with the free N atom of the histidine ring. This preference can be attributed to the presence of heteroatoms in the imidazole ring of HIS, providing coordination sites for metal ions. Moreover, quantum chemical

studies have suggested a preference for cation- σ interactions between metal ions and hetero-aromatics, further supporting the abundance of spherical motifs over cylindrical motifs. The preference of metal cations to bind aromatic residues follows the order of

Arene	Cation	# cylindrical motifs	# spherical motifs	# total motifs
PHE	ARG	160 551	2 148 164	23,08715
PHE	HIS	132 891	1 508 771	1 641 662
PHE	LYS	135 818	1 830 625	1 966 443
TYR	ARG	154 061	2 242 775	2 396 836
TYR	HIS	120 157	1 428 813	1 548 970
TYR	LYS	136 349	2 635 393	2 771 742
TRP	ARG	158 762	2 156 803	2 315 565
TRP	HIS	116 808	1 029 229	1 146 037
TRP	LYS	122 083	1 533 450	1 655 533
HIS	ARG	68 334	1 380 259	1 448 593
HIS	HIS	2,39 927	4 873 778	5 113 705
HIS	LYS	49 676	1 378 173	1 427 849

TABLE 1 The frequency of pair-wise occurrence of basic (ARG, HIS, LYS) amino acid and aromatic (PHE, TYR, TRP, HIS) amino acids in the cation-aromatic motifs.

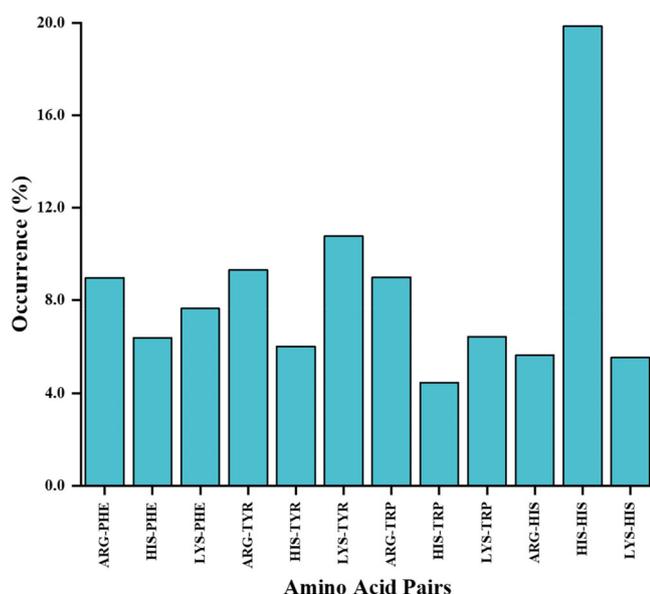


FIGURE 7 The percentage occurrence of the pairs of amino acid residues PHE, TYR, TRP, HIS, ARG, and LYS in the cation-aromatic motifs.

HIS > TYR > PHE > TRP. Metal cations preferentially interact with aromatic residues, but TYR and PHE have limited binding ability due to their lack of nitrogen atoms and functional groups, as well as their larger size. TRP's complex indole ring also reduces its affinity for metal ion interactions.

6 | BINDING PREFERENCES OF CATIONS TOWARD DIFFERENT AROMATIC RESIDUES

The geometrical parameters of cation-aromatic motifs were examined to elucidate the binding preferences of cations toward aromatic

residues using contour plots. The plots in Figures 8 and 9 and Figure S2 show the density of observed cation-aromatic pairs, and Table S12 provides a glimpse of the statistical descriptors applied to the distance parameters. A closer inspection of Table S12 suggests that in addition to the minor skew in the distribution of r , $d1$, and $d2$ values, the variance for the perpendicular distances is much higher compared with the r values suggesting a lesser spread around the mean in the latter. This combined with the frequency of occurrence of these distance parameter pairs has shown some valuable insights. The highest density of pairs occurs within specific distance ranges: 3.0–4.5 for $d1$ and 5.0–6.0 for r (Figure 9A), 5.0–6.0 for $d2$ and r (Figure 9B), 2.0–4.0 for $d1$ and 3.5–6.0 for $d2$ (Figure 9C), which suggests a potential optimal distance for cation-arene interactions. This finding could have implications for understanding the mechanism of cation-arene interactions and their role in biological systems. Given that the $d2$ parameter provides valuable insights into the preferences of cationic-aromatic residue pairs toward cation- π and cation- σ interactions, and r provides the distance between arene centroid and cation, we conducted a more detailed examination of r versus $d2$ for all cationic-aromatic residue pairs generating the contour plots, as depicted in Figure 10. PHE-ARG and TRP-ARG pairs showed a sharp peak around 4.5 and 5.0 Å for PHE-ARG, and 3.9 and 5.5 Å for TRP-ARG, indicating a preferential geometry. PHE and TRP pairs with LYS and HIS exhibited a scattered distribution, suggesting multiple interacting poses. TYR-based cation-arene pairs showed discrete distributions with major and minor peaks. TRP-ARG and TRP-HIS pairs displayed prominent peaks at around 5.5 and 5.0 Å, respectively. HIS-HIS pairs showed a major peak at smaller distances (around 2.2 Å), indicating a preferential geometry. HIS-ARG and HIS-LYS pairs displayed spread-out distributions with smaller peaks, suggesting multiple preferred geometrical arrangements. Overall, these contour plots provide valuable insights into the binding preferences and geometrical arrangements of cationic and aromatic residues, with potential implications for understanding cation-arene interactions in biological systems.

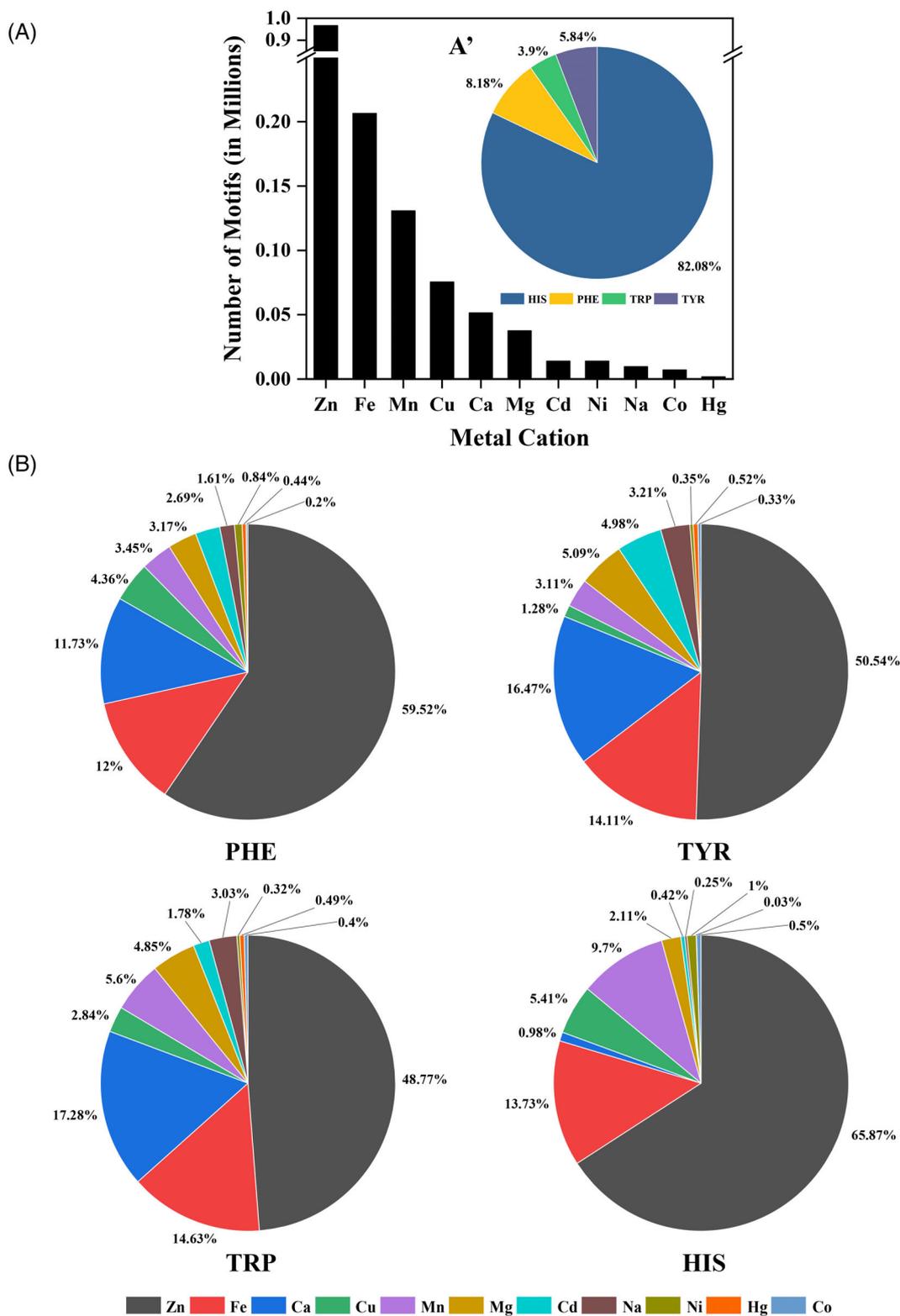


FIGURE 8 The frequency of occurrence of (A) metal cations, (A') aromatic amino acids in the metal atom containing cation-aromatic motifs and (B) occurrences of metal cations in with the respective aromatic residues viz. PHE, TYR, TRP, and HIS.

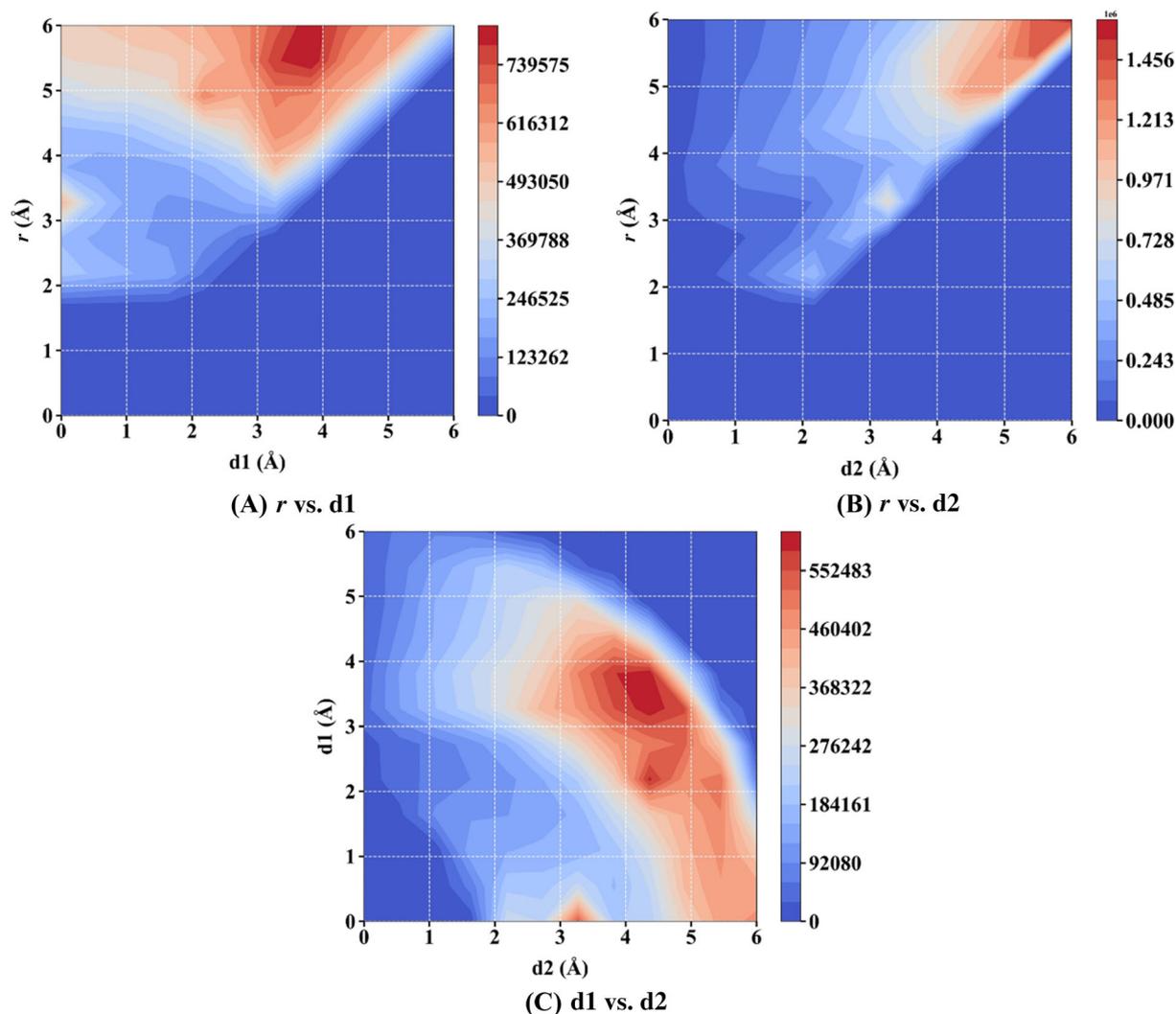


FIGURE 9 Contour maps representing the frequency of distribution of the *r*-*d1*, *r*-*d2*, and *d1*-*d2* values in the observed cation (ARG, LYS, HIS)-arene (PHE, TYR, TRP, HIS) pairs. The distance parameters *r*, *d1*, and *d2* are given in Å.

7 | CLASSIFICATION OF CATION-AROMATIC MOTIFS BASED ON ENZYME CLASSIFICATION

The cation-aromatic motifs were categorized based on the enzyme classification entries found in the protein databank. Out of the 193 936 unique proteins involved in cation-aromatic interactions, 104 964 proteins are associated with one or more classes in the Enzyme Classification (EC) scheme.²¹ The proteins in each enzyme class can be further classified as oxidoreductase (17 678), transferase (37 350), hydrolase (44 001), lyase (10 083), isomerase (4225), ligase (3603), and translocase (1660). A table in the database contains the details of these proteins, and the cation-aromatic motifs belonging to these classes can be accessed using the “Class” option on the CAD webpage.

7.1 | A comparison with other resources

There are several resources available for obtaining information on noncovalent interactions, including cation- π interactions in

proteins and other molecular systems. These resources include CIPDB, RING, ExptNCI, PLIP, MolADI, ProteinTools, and others.²²⁻³⁴ In Table S13, you can find a comprehensive list of databases, web servers, and software programs specifically designed for identifying and analyzing these interactions. However, among these resources, CAD V2.0 stands out as a remarkable database of cation-aromatic motifs. It contains information on approximately 27.26 million such motifs, identified within the 3D structures of proteins. CAD V2.0 employs three distance parameters, namely *r*, *d1*, and *d2*, to determine the position of the cation in relation to the arene. It further categorizes cation-aromatic motifs as either cation- π or cation- σ motifs. The cation-aromatic database not only offers an extensive collection of information on cation-aromatic motifs but also provides a comprehensive analysis of their occurrence across different protein classes. This wealth of information can greatly contribute to our understanding of the structural and functional implications of cation-aromatic interactions in proteins. Additionally, it can facilitate the development of new methods for predicting and analyzing such interactions.

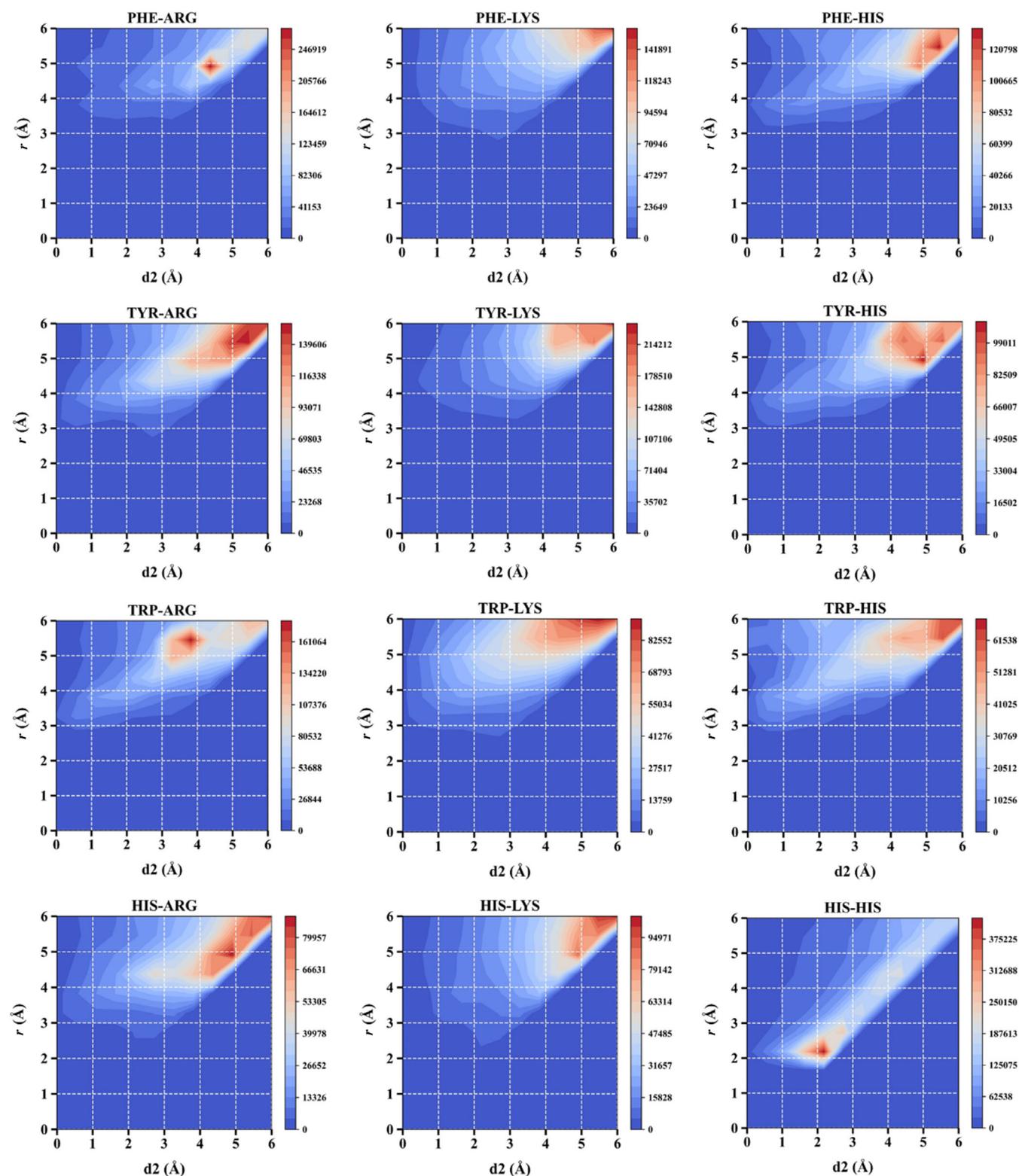


FIGURE 10 Contour maps representing the frequency of distribution of $d2$ with respect to r for cation (ARG, LYS, HIS)-arene (PHE, TYR, TRP, HIS) pairs. The distance parameters r and $d2$ are given in Å.

8 | CONCLUSIONS

The aromatic side chain residues of the proteins interact attractively with the cationic moieties like basic amino acids or metal ions forming

cation-aromatic motifs. The mode of interaction between the cation and the aromatic residue not only determines the preference of forming a cation- π or cation- σ interaction but also the strengths of these interactions. Analyzing the occurrence of these cation-aromatic motifs

in proteins can underline the influence of aromatic and other non-covalent interactions in the structural stability of the proteins. This update of the cation-aromatic database put forth an extensive analysis of the cation-aromatic motifs and their frequency of occurrence in different types of protein classes. Out of the 27 255 817 cation-aromatic motifs identified, 94.09% were spherical motifs, and the remaining 5.91% were cylindrical motifs. This suggests a distinct preference for cation- σ interactions over cation- π interactions in proteins. The distribution of cation-aromatic motifs based on distance parameters showed a general increase in motif count with distance, leading to Gaussian distributions. This suggests that most cation-aromatic interactions are either weak or moderately strong. The distribution of aromatic residues and cations in the set of identified cation-aromatic motifs suggested that the order of relative abundance of the arene residues in the motifs is HIS > TYR > PHE > TRP with a range of occurrence to between 19% and 34%. Among the basic amino acids, the HIS molecules occur in \sim 37% of the residues followed by ARG and LYS. As anticipated, when the basic amino acid and aromatic residue pairs were investigated, it was observed that the HIS-HIS pair had the highest relative abundance followed by LYS-TYR and ARG-TYR. Although the overall number of metal cations in the motifs is relatively low, the Zn ion was found in \sim 1 million cation-aromatic motifs, surpassing the occurrences of Fe and Mn by a large margin. Interestingly, the HIS residue has occurred most frequently as the prominent arene even in the metal-arene motifs.

AUTHOR CONTRIBUTIONS

Y. Bhargav Kumar: Data curation; investigation; formal analysis; methodology; validation; writing – original draft. **Nandan Kumar:** Investigation; validation; data curation; writing – review and editing. **Lijo John:** Writing – original draft; data curation. **Hridoy Jyoti Mahanta:** Writing – original draft; validation; writing – review and editing. **S. Vaikundamani:** Software; formal analysis; data curation. **Selvaraman Nagamani:** Writing – original draft; writing – review and editing; validation. **G. Madhavi Sastry:** Software; conceptualization. **G. Narahari Sastry:** Conceptualization; writing – review and editing; supervision; project administration.

ACKNOWLEDGMENTS

DBT is thanked for the financial support in the form of Centre of Excellence in Advanced Computation and Data Sciences (Ref. No: BT/PR40188/BTIS/137/27/2021).

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

G. Narahari Sastry  <https://orcid.org/0000-0003-3181-7673>

REFERENCES

- Reddy AS, Sastry GM, Sastry GN. Cation-aromatic database. *Proteins*. 2007;67(4):1179-1189.
- Burley SK, Bhikadiya C, Bi C, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res*. 2021;49:D437-D451.
- Mahadevi AS, Sastry GN. Cation- π interaction: its role and relevance in chemistry, biology, and material science. *Chem Rev*. 2013;113(3):2100-2138.
- Mahadevi AS, Sastry GN. Cooperativity in non-covalent interactions. *Chem Rev*. 2016;116(5):2775-2825.
- Kumar N, Gaur AS, Sastry GN. A perspective on the nature of cation- π interactions. *J Chem Sci*. 2021;133:1-3.
- Gallivan JP, Dougherty DA. A computational study of cation- π interactions vs salt bridges in aqueous media: implications for protein engineering. *J Am Chem Soc*. 2000;122(5):870-874.
- Sharma B, Umadevi D, Sastry GN. Contrasting preferences of N and P substituted heteroaromatics towards metal binding: probing the regioselectivity of Li⁺ and Mg²⁺ binding to (CH)₆-m-nNmPn. *Phys Chem Chem Phys*. 2012;14(40):13922-13932.
- Kumar N, Saha S, Sastry GN. Towards developing a criterion to characterize non-covalent bonds: a quantum mechanical study. *Phys Chem Chem Phys*. 2021;23(14):8478-8488.
- Priyakumar UD, Punnagai M, Mohan GK, Sastry GN. A computational study of cation- π interactions in polycyclic systems: exploring the dependence on the curvature and electronic factors. *Tetrahedron*. 2004;60(13):3037-3043.
- Sharma B, Srivastava HK, Gayatri G, Sastry GN. Energy decomposition analysis of cation- π , metal ion-lone pair, hydrogen bonded, charge-assisted hydrogen bonded, and π - π interactions. *J Comput Chem*. 2015;36(8):529-538.
- Sharma B, Rao JS, Sastry GN. Effect of solvation on ion binding to imidazole and methylimidazole. *J Phys Chem A*. 2011;115(10):1971-1984.
- Sharma B, Neela YI, Narahari SG. Structures and energetics of complexation of metal ions with ammonia, water, and benzene: a computational study. *J Comput Chem*. 2016;37(11):992-1004.
- Vijay D, Sastry GN. A computational study on π and σ modes of metal ion binding to heteroaromatics (CH)_{5-m}X_m and (CH)_{6-m}X_m (X = N and P): contrasting preferences between nitrogen- and phosphorous-substituted rings. *J Phys Chem A*. 2006;110(33):10148-10154.
- Rao JS, Sastry GN. Structural and energetic preferences of π , σ , and bidentate cation binding (Li⁺, Na⁺, and Mg²⁺) to aromatic amines (Ph-(CH₂)_n-NH₂, n = 2-5): a theoretical study. *J Phys Chem A*. 2009;113(18):5446-5454.
- Rao JS, Zipse H, Sastry GN. Explicit solvent effect on cation- π interactions: a first principle investigation. *J Phys Chem B*. 2009;113(20):7225-7236.
- Vijay D, Sastry GN. Exploring the size dependence of cyclic and acyclic π -systems on cation- π binding. *Phys Chem Chem Phys*. 2008;10(4):582-590.
- Reddy AS, Zipse H, Sastry GN. Cation- π interactions of bare and coordinatively saturated metal ions: contrasting structural and energetic characteristics. *J Phys Chem B*. 2007;111(39):11546-11553.
- Chourasia M, Sastry GM, Sastry GN. Aromatic-aromatic interactions database, A2ID: an analysis of aromatic π -networks in proteins. *Int J Biol Macromol*. 2011;48(4):540-552.
- Spontarelli K, Infield DT, Nielsen HN, et al. Role of a conserved ion-binding site tyrosine in ion selectivity of the Na⁺/K⁺ pump. *J Gen Physiol*. 2022;154(7):e202113039.
- Sparrow ZM, Ernst BG, Joo PT, Lao KU, DiStasio RA Jr. NENCI-2021. I. A large benchmark database of non-equilibrium non-covalent

- interactions emphasizing close intermolecular contacts. *J Chem Phys*. 2021;155(18):184303.
21. Kriz K, Novacek M, Rezac J. Non-covalent interactions atlas benchmark data sets 3: repulsive contacts. *J Chem Theory Comput*. 2021; 17(3):1548-1561.
 22. Novikov AS. IsoStar program suite for studies of noncovalent interactions in crystals of chemical compounds. *Crystals*. 2021; 11(2):162.
 23. Bietz S, Urbaczek S, Schulz B, Rarey M. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J Chem*. 2014;6:1-12.
 24. Li S, Wan F, Shu H, Jiang T, Zhao D, Zeng J. MONN: a multi-objective neural network for predicting compound-protein interactions and affinities. *Cell Syst*. 2020;10(4):308-322.e11.
 25. Biot C, Buisine E, Kwasigroch JM, 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.
 26. Crowley PB, Golovin A. Cation- π interactions in protein-protein interfaces. *Proteins*. 2005;59(2):231-239.
 27. Reddy AS, Sastry GN. Cation [M= H⁺, Li⁺, Na⁺, K⁺, Ca²⁺, Mg²⁺, NH₄⁺, and NMe₄⁺] interactions with the aromatic motifs of naturally occurring amino acids: a theoretical study. *J Phys Chem B*. 2005; 109(39):8893-8899.
 28. Enzyme Nomenclature. <https://iubmb.qmul.ac.uk/enzyme/>. Accessed March 20, 2023.
 29. Yang JF, Wang F, Wang MY, et al. CIPDB: a biological structure database for studying cation- π interactions. *Drug Discov Today*. 2023;28: 103546.
 30. Clementel D, Del Conte A, Monzon AM, et al. RING 3.0: fast generation of probabilistic residue interaction networks from structural ensembles. *Nucleic Acids Res*. 2022;50(W1):W651-W656.
 31. Ding K, Yin S, Li Z, et al. Observing noncovalent interactions in experimental electron density for macromolecular systems: a novel perspective for protein-ligand interaction research. *J Chem Inf Model*. 2022;62(7):1734-1743.
 32. Adasme M, Linnemann KL, Bolz SN, et al. PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA. *Nucleic Acids Res*. 2021;49(W1):W530-W534.
 33. Bai B, Zou R, Chan HCS, Li H, Yuan S. MolADI: a web server for automatic analysis of protein-small molecule dynamic interactions. *Molecules*. 2021;26(15):4625.
 34. Ferruz N, Schmidt S, Hocker B. ProteinTools: a toolkit to analyze protein structures. *Nucleic Acids Res*. 2021;49(W1):W559-W566.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kumar YB, Kumar N, John L, et al. Analyzing the cation-aromatic interactions in proteins: Cation-aromatic database V2.0. *Proteins*. 2023;1-13. doi:10.1002/prot.26600