International Journal of Biological Macromolecules 253 (2023) 127207

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Analyzing the aromatic-aromatic interactions in proteins: A^{2}ID 2.0

Y. Bhargav Kumar [a,b], Nandan Kumar [a], S. Vaikundamani [a], Selvaraman Nagamani [a,b], Hridoy Jyoti Mahanta [a,b], G. Madhavi Sastry [c], G. Narahari Sastry [a,b,*]

[a] Advanced Computation and Data Sciences Division, CSIR-North East Institute of Science and Technology, Jorhat 785006, Assam, India
[b] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, U. P., India
[c] Schrödinger Inc., HITEC City, Hyderabad, Telangana 500081, India

ARTICLE INFO

Keywords:
Proteins
Aromatic interactions
Centroid distance
Interplanar angle
Database

ABSTRACT

The Aromatic-Aromatic Interactions Database (A^{2}ID) is a comprehensive repository dedicated to documenting aromatic-aromatic (π-π) networks observed in experimentally determined protein structures. The first version of A^{2}ID was reported in 2011 [*Int J Biol Macromol*, 2011, 48, 540]. It has undergone a series of significant updates, leading to its current version, which focuses on the identification and analysis of 3,444,619 π-π networks from proteins. The geometrical parameters such as centroid-centroid distances (**r**) and interplanar angles (**ϕ**) were used to identify and characterize π-π networks. It was observed that among the 84,500 proteins with at least one aromatic π-π network, about 92.50 % of the instances are found to be either 2π (77.34 %) or 3π (15.23 %) networks. The analysis of interacting amino acid pairs in 2π networks indicated a dominance of PHE residues followed by TYR. The updated version of A^{2}ID incorporates analysis of π-π networks based on SCOP2 and ECOD classifiers, in addition to the existing SCOP, CATH, and EC classifications. This expanded scope allows researchers to explore the characteristics and functional implications of π-π networks in protein structures from multiple perspectives. The current version of A^{2}ID along with its extensive dataset and detailed geometric information is publicly accessible using https://acds.neist.res.in/a2idv2.

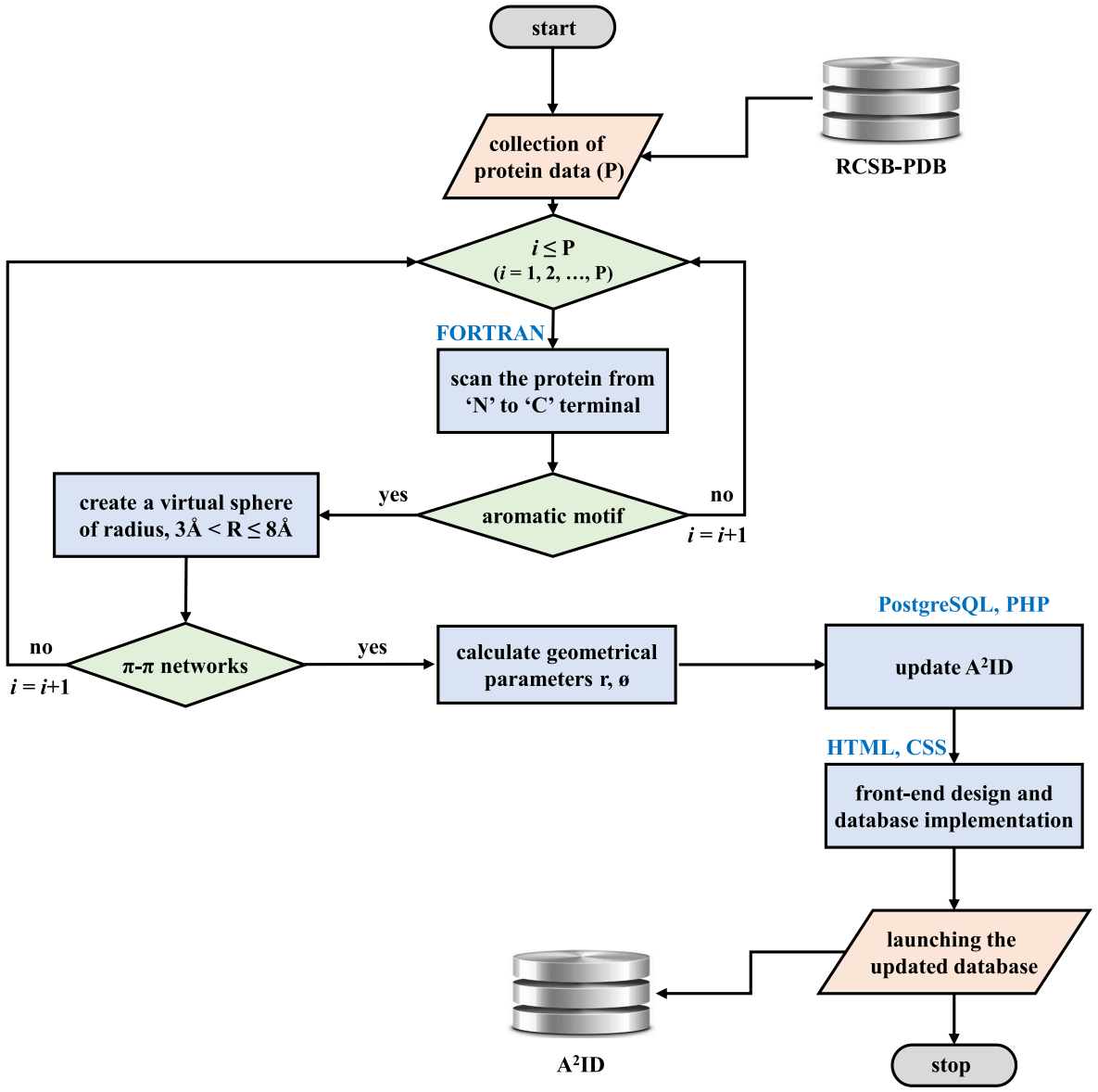
1. Introduction

Aromatic residues play a crucial role in protein structures due to their abundance and ability to engage in diverse intermolecular interactions and significantly contribute to the overall stability of proteins. These interactions play a pivotal role in vital processes such as protein folding, ligand binding, and molecular recognition [1–12]. Among the non-covalent interactions, the most crucial aspect is how the π-π interactions manifest in biomolecular structure is responsible for bestowing the protein structure, folding, and function. A^{2}ID offers extensive information on π-π networks in protein structures and was the first database to report the π-networks in proteins [1] comprehensively. It reported 7848 proteins that were deposited in the Protein Data Bank till September 2008 with their associated π-π networks. This database has significantly contributed to understanding the prevalence and characteristics of π-π networks in protein structures. Furthermore, significant advancements in our understanding of π-π interactions have been made through extensive research and computational studies [13–32]. Recent years have witnessed a tremendous increase in experimental sophistication and computational power which helped in providing precise information on the strength and nature of π-π interactions as well as their role in the structure and function of the proteins [15–25]. Investigations into the prevalence of these interactions in the Protein Data Bank (PDB) [15,26] and the extensive investigation of amino acids such as HIS [27–30], PHE [31], and TYR [29,32] have provided valuable insights. Beyond dimers, aromatic residues exhibit a strong tendency to form clusters or networks, which are key sources of cooperativity and have a profound influence on protein folding, structure, and stability [33–35]. Given the rapid increase in the number of protein structures deposited in the Protein Data Bank and the advancements in our understanding of protein structures and aromatic-aromatic interactions, an update to the A^{2}ID has become imperative.

The present study is an update of A^{2}ID, which included the updated information in the last 13 years and provides the data on the occurrence of 2π to 14π networks in proteins. Algorithm that is employed in the database provide quantitative information structural parameters of the π-networks. Furthermore, the preference of these networks is demonstrated based on five different classifiers: SCOP (Structural Classification

* Corresponding author at: Advanced Computation and Data Sciences Division, CSIR-North East Institute of Science and Technology, Jorhat 785006, Assam, India.
E-mail addresses: gnsastry@gmail.com, gnsastry@neist.res.in (G.N. Sastry).

https://doi.org/10.1016/j.ijbiomac.2023.127207
Received 30 June 2023; Received in revised form 9 September 2023; Accepted 30 September 2023
Available online 4 October 2023

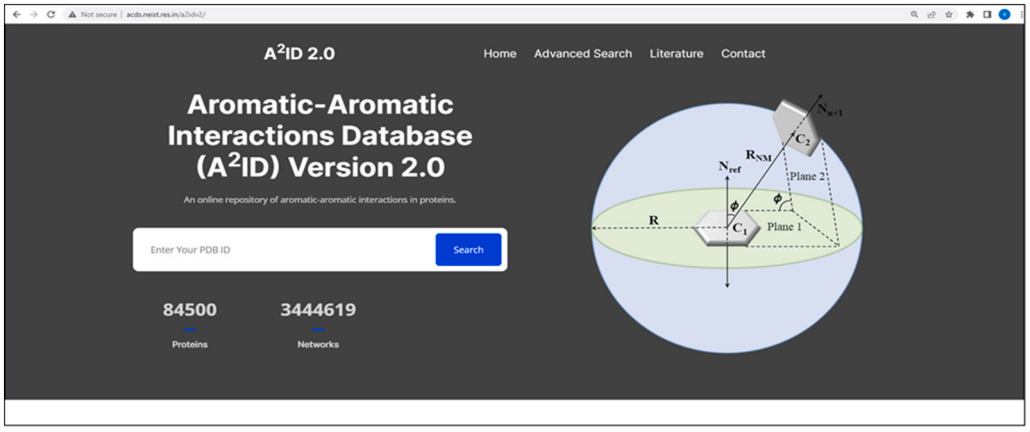


Scheme 1. The workflow for updating the Aromatic-Aromatic Interaction Database.

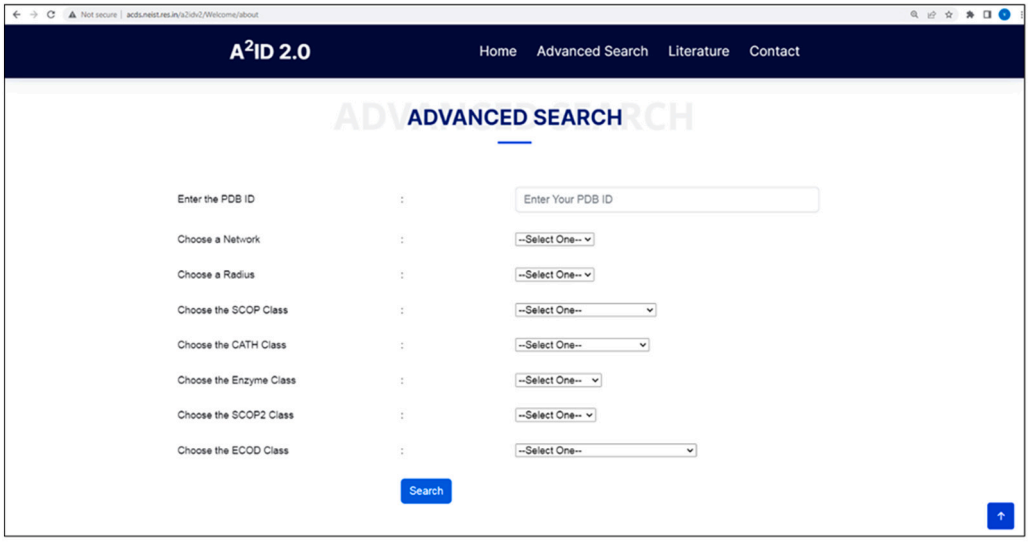
of Proteins) [33], SCOP2 (an updated version of SCOP) [33], ECOD (Evolutionary Classification of Protein Domains) [34], CATH (Class, Architecture, Topology, and Homologous superfamily) [35], and EC (Enzyme Classification) [36]. This updated A^{2}ID allows a more comprehensive understanding of π-π networks and their occurrence in different proteins. Through continuous improvement and expansion, our goal is to provide the scientific community with a reliable and indispensable resource for studying aromatic-aromatic interactions in protein structures and exploring specific networks and connectivity of aromatic rings in proteins.

2. Identifying the geometrical parameters

For the update of the A^{2}ID, the analysis of π-π networks in proteins was conducted in a similar manner to the initial release [1]. A set of 84,500 proteins deposited in the PDB until 20th June 2023 was considered that have an X-ray resolution of ≤2 Å. The analysis of the π-π networks has been done by considering the protein in its entirety. To avoid the bias created in the results due to the sequence identity cut-offs considered in the earlier version, we have dropped that sequence identity filter for the identification of the arrangements and the subsequent calculation of their geometrical parameters. The necessary modifications in the FORTRAN code have been made to be able to read the most recent PDB file format for proteins. The FORTRAN code identifies the centres of the three aromatic rings in phenylalanine (PHE), tyrosine (TYR), and the 5-membered histidine (HIS) and then creates a virtual sphere around these centres to identify aromatic-aromatic networks within. However, in the case of tryptophan (TRP) which contains a 6- and a 5-membered ring fused together as the indole ring, the code identifies the centres of both the rings instead of arbitrarily identifying the centre of the indole moiety. This enables for a better understanding on which ring of the TRP residue specifically interacts with the neighbouring aromatic residue. The two aromatic centres of The π-π network is represented by using three distinct parameters viz. (i) centroid-centroid distances (**r**) between the aromatic ring pairs, (ii) the angle between the planes of the two aromatic rings (interplanar angle, **ϕ**), and (iii) the centroid of one aromatic ring surrounded by multiple aromatic rings. Similar to the previous version, various cut-off radii (***R*** = 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0 Å) were considered to identify all the exclusive π-π networks in the proteins at different distance criteria. The aromatic networks were identified based on the number of aromatic residues (**n**) in a π-network. The FORTRAN code first finds a 2π network



(A)



(B)

Fig. 1. Snapshots of (A) home page and (B) advanced search page of the updated Aromatic-Aromatic Interactions Database Version 2.0 (A^2ID V2.0).

Table 1
The list of all the nπ (n = 2–14) networks identified at each cut-off radii, R = 3.0 Å–8.0 Å.

nπ	Cut-off Radius, R (in Å)											
	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	Sum
2π	505	2763	49,799	135,344	330,739	432,060	423,980	384,735	344,091	300,834	259,242	2,664,092
3π	40,118	40,012	37,486	34,233	33,546	47,901	60,879	63,114	60,917	56,377	50,121	524,704
4π	1	4	16	218	3852	14,175	23,737	30,152	32,230	33,163	30,595	168,143
5π	0	2	8	13	845	3749	7628	9553	10,590	11,150	10,879	54,417
6π	0	0	0	0	78	788	2754	3939	4659	4639	4520	21,377
7π	0	0	0	0	44	159	471	1340	1577	1729	1903	7223
8π	0	0	0	0	3	23	215	373	643	860	979	3096
9π	0	0	0	0	0	5	38	129	203	236	262	873
10π	0	0	0	0	0	9	13	71	96	130	169	488
11π	0	0	0	0	0	50	2	12	19	20	20	123
12π	0	0	0	0	0	3	0	6	10	18	20	57
13π	0	0	0	0	0	0	1	0	0	5	10	16
14π	0	0	0	0	0	0	0	0	0	4	6	10
Sum	40,624	42,781	87,309	169,808	369,107	498,922	519,718	493,424	455,035	409,165	358,726	**3,444,619**

by identifying an aromatic ring in the virtual sphere of cut-off radius **R**, generated from the centre of the first aromatic residue. The program further proceeds to scan for more aromatic rings by generating two virtual spheres, one from reach centre of the 2π network. In the event of finding one aromatic residue, the network becomes a 3π network and this process is repeated until no higher networks are found. It has been

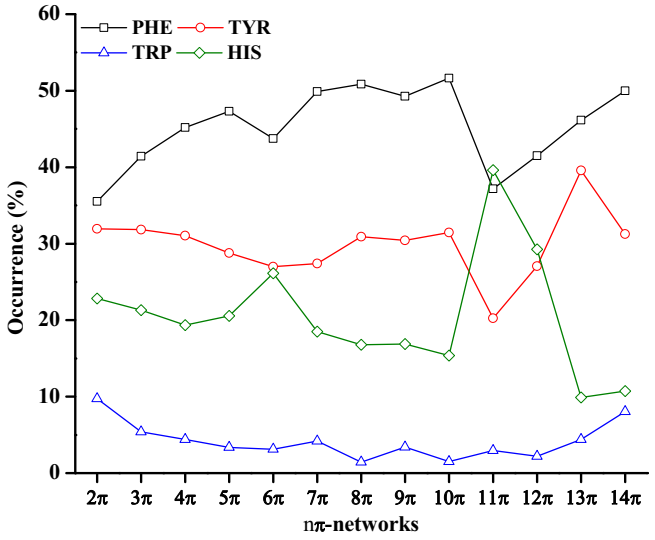


Fig. 2. The percentage occurrence of the four aromatic amino acid residues PHE, TYR, TRP, and HIS in the obtained nπ-networks (n = 2–14).

observed that the largest networks that were identified were the 14π networks. By considering the criteria of centroid-centroid distances (**r**) and the interplanar angle (**ϕ**), the overall arrangement and nature of interactions within the π-networks were estimated.

3. Updating the database

The overall workflow for updating the database involves several steps, as shown in Scheme 1. The database has been arranged into six tables, the first table has the details of all the aromatic networks and their geometrical arrangement information. This table has been organized into 9 columns i.e., (i) PDB ID, (ii) protein name, (iii) deposition date, (iv) cut-off radius, (v) nπ network, (vi) connectivity pattern of the nπ network, (vii) centroid-centroid distance between the rings, (viii) interplanar angle between all the interacting pair of aromatic residues, and (ix) residues involved in the nπ network. The remaining five tables correspond to the entries of SCOP, CATH, EC, SCOP2, and ECOD classification schemes and contain information on all the proteins available in each subclass of the classes, as additional refinement criteria to the basic query options. PostgreSQL (version 14) is installed and employed on Linux (Cent 7.9) Operating System for implementing the database. HTML, PHP, and CSS with Apache server (6.0.16) have been used for the design and development of the interactive user interface. PHP to PostgreSQL connector has been used to connect the back-end database server to the front-end. The front-end has been transformed into a user-friendly graphical user interface (GUI), allowing users to easily extract the desired information. This can be achieved by utilizing a range of refinement parameters provided on the advanced search page. The refinement filters include class (or subclass in a certain class), radius, and network, which will provide the output of choice as per the user's requirement. The π-π networks in proteins and their connectivity information can be retrieved by the user either by using the basic query option or the advanced search tool. The basic query requires the user to simply give the four letter PDB ID to retrieve the information related to the protein. The advanced search offers eight different search criteria as query options for the user which include, PDB ID, cut-off radius (R), length of the π-π networks (nπ), classes of the SCOP, CATH, EC, SCOP2, ECOD classification schemes. Fig. 1 shows the snapshot of the home page and advanced search page of the updated database (A^2ID 2.0) which also displays the number of proteins currently included in the database and the number of the π-π networks. An additional feature has been provided for the user to upload either new proteins from PDB or modelled/computationally predicted structures into the webserver for identifying the occurrence of different aromatic networks and calculating their corresponding geometrical parameters since this would

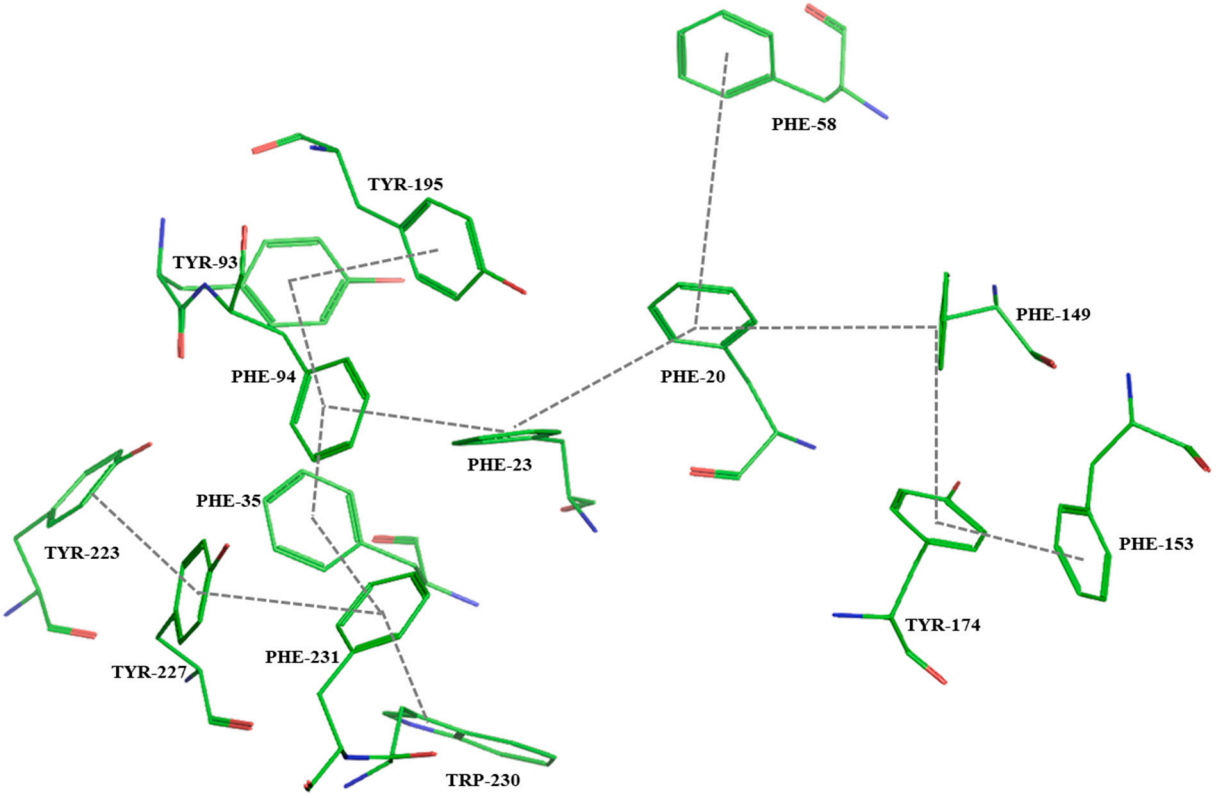


Fig. 3. Representation of a 14π network observed in the protein 4ZV2 along with the residues involved in the network.

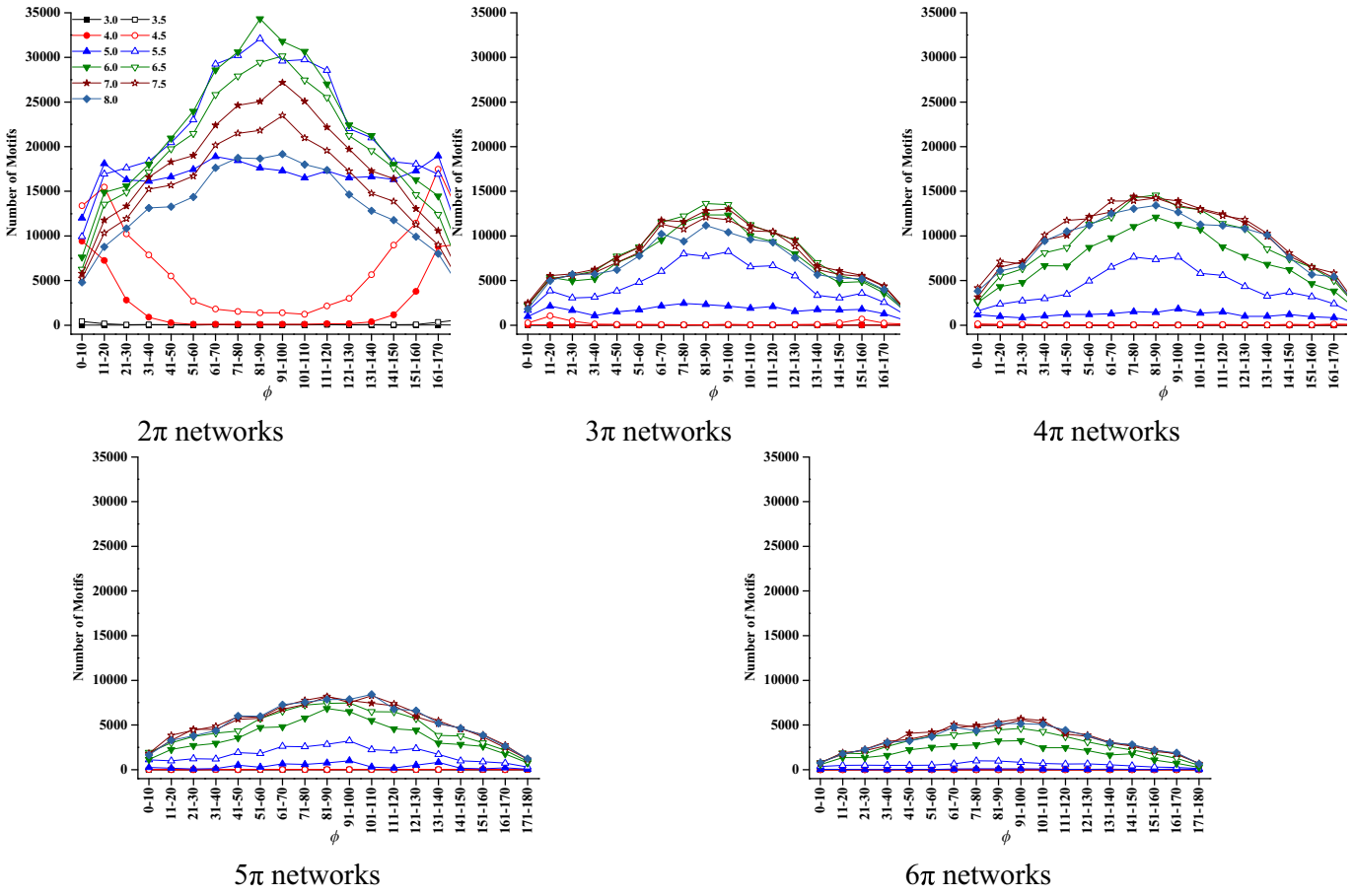


Fig. 4. The distribution of plane angles (ϕ in degree) in the π-motifs obtained at different cut-off radii (R = 3.0 Å–8.0 Å) for 2π, 3π, 4π, 5π, and 6π networks.

provide valuable insights, especially in the context of protein design and understanding the impact of mutations on the protein structure and its networks.

4. Distribution of the amino acids in π-π networks

The analysis of 84,500 proteins has resulted in a total of 3,444,619 π-π networks where the 2π and 3π networks constitute >90 % of the overall networks. The number of proteins which has at least one π-network along with the year of the protein deposition has been provided in Table S1 while an exhaustive list of individual π-networks at all the cut-off radii considered has been presented in Table 1. The occurrence and distribution of the four aromatic amino acids viz. PHE, TYR, TRP, and HIS in the π-π networks have been analyzed (Fig. 2 and Table S2). Aromatic amino acids, namely PHE, TYR, HIS, and TRP, are found abundantly in nature, with occurrence percentages of 4 %, 4 %, 3 %, and 2 % respectively. Collectively, these aromatic amino acids contribute to an overall occurrence of 13 % in protein structures, highlighting their significant presence in biological systems. Through analysis, it has been revealed that PHE exhibits a greater prevalence in π-π networks among the four aromatic residues in proteins. This dominance can be attributed to the smaller size of the PHE residue, which results in fewer steric hindrances compared to the other aromatic residues [36]. The frequency of occurrence of aromatic amino acids follows a distinct trend: PHE > TYR > HIS > TRP. Notably, the relatively lower abundance of TRP residues can be attributed to the larger size of this fused aromatic residue. The occurrence of PHE residue varies from 35 to 52 %, while the TYR, HIS, and TRP residues have always been observed to be occurring up to 40 % in most cases.

5. Analysis of π-π networks

The arrangement of the π-π networks in the proteins, which is one of the important features is identified by scanning the protein from the N- to C-terminal and identifying the aromatic rings that form networks. The preferences of the aromatic rings to either cluster together to form interconnected networks or to spread out on a larger area help in understanding the protein stability. The 2π network can only be linear, while the 3π network can either be found in linear or triangular arrangements. The 4π or any other higher network will have a greater number of possible arrangement patterns and the number increases with the increase in the length of the **n**π network. The connectivity information has been incorporated in the form of a numerical string of length **n** for any **n**π network where the position (place) of each number indicates the number of the aromatic residue in the network and the number at each position defines the number of contacts (interactions) made by the ring with its neighbouring aromatic residues. The largest exclusive **n**π networks i.e., 14π networks were identified in only ten instances and involved only six proteins with PDB IDs 4AE8, 4ZV2, 6BJ2, 6GPC, 6W5A, and 7VWB. A representative 14π network of 4ZV2 protein has been depicted in Fig. 3. The protein 4AR8 displays four 14π networks, showcasing two distinct connectivity patterns: 55522164522111 and 65522165522111. The details of these connectivity patterns are explained in the previous version [1]. Furthermore, it has been observed that the 14π networks in proteins 6BJ2 and 6W5A are consistently arranged in a 78544423455531 pattern. In the case of protein 7VWB, two distinct connectivity patterns have been identified: 44332443322121 and 54433443332222. Additionally, the remaining two connectivity patterns, namely 45713632242452 and 47314854731485, have been discovered in proteins 4ZV2 and 6GPC,

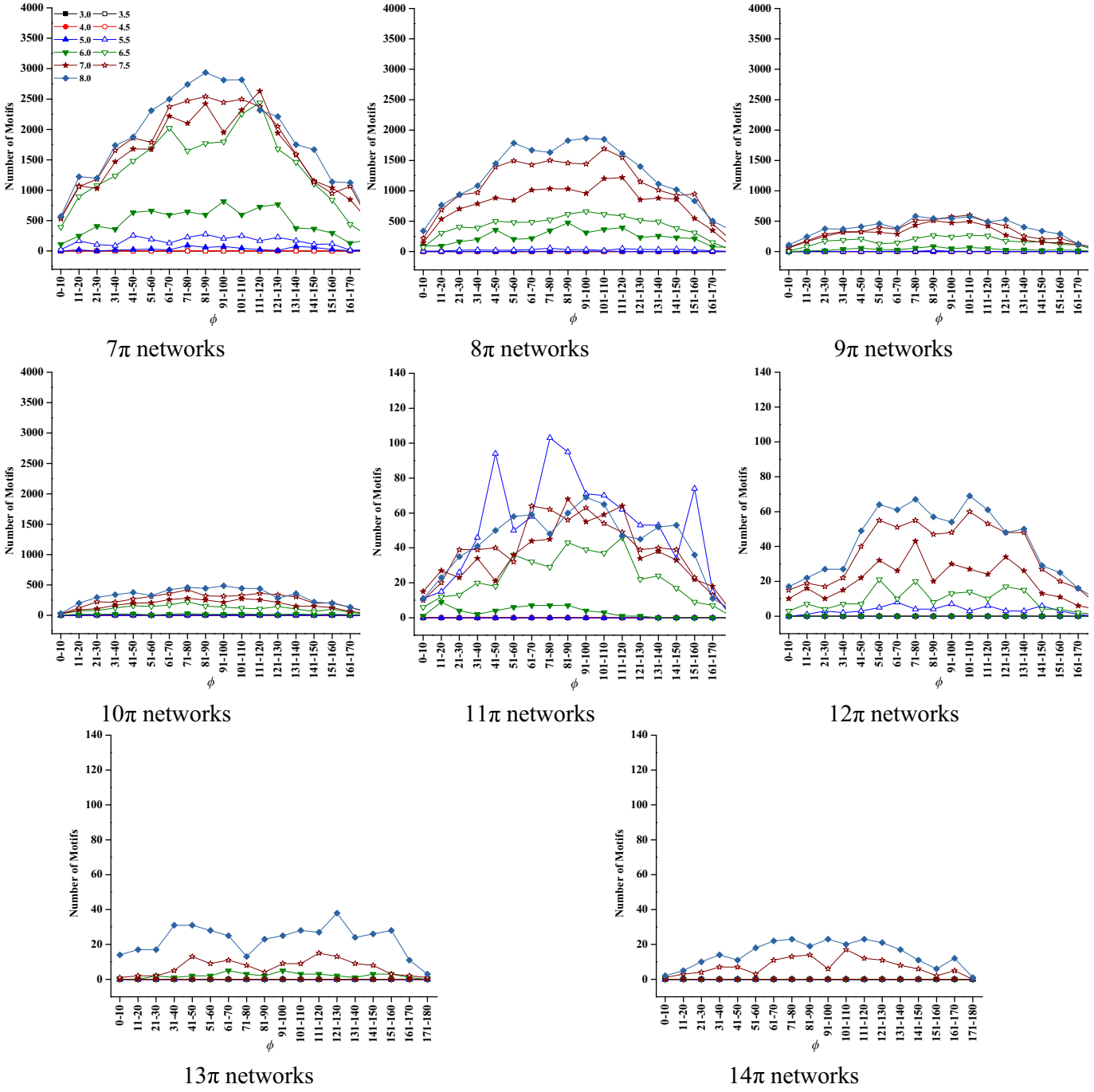


Fig. 5. The distribution of plane angles (ϕ in degree) in the π-motifs obtained at different cut-off radii (R = 3.0 Å–8.0 Å) for 7π, 8π, 9π, 10π 11π, 12π, 13π, and 14π networks.

respectively. Other large **n**π networks include the 13π networks, which were observed in merely 16 instances, and the 12π networks which were obtained in 57 instances in the overall database. Interestingly, the higher π-π networks (13π and 14π) were only identified at larger cut-off distances i.e., 7.5 and 8.0 Å, while the 12π networks were observed at a much diverse range of cut-off radii. Overall, we have identified only 694 (~0.02 %) 10π or higher networks among the ~3.4 million π-π networks as shown in Table 1. Within the scope of our investigation into oligomeric proteins, a comprehensive evaluation of our analysis reveals the adept capture of inter-monomer networks. We looked at the aromatic-aromatic interactions in various protein structures like lactose dehydrogenase (9LDT), *E. coli* L-asparaginase II mutant (8ECE), Concanavalin A bound to a DNA glycoconjugate structure (7MG7), human transthyretin (6SUH), beta-ketoadipyl-CoA thioloase (6PCD), oxido reductase (6LGK), and short-chain dehydrogenase reductase (5JY1). For example, in protein 9LDT, we found that parts called TRP 248 in chain B and TRP 248 in chain A interact with each other at a distance of 6.11 Å and an angle of 96.169°. Another interesting discovery was in protein 8ECE, where a network involving parts TYR 250, TYR 220, and TYR 218 in chains B and A had angles ranging from 5.779 to 26.425°. These findings are just a glimpse of the intricate connections within these proteins. If user wants to explore more details about how different parts of oligomeric proteins interact, user can search for their protein of interest using its PDB ID to discover helpful insights.

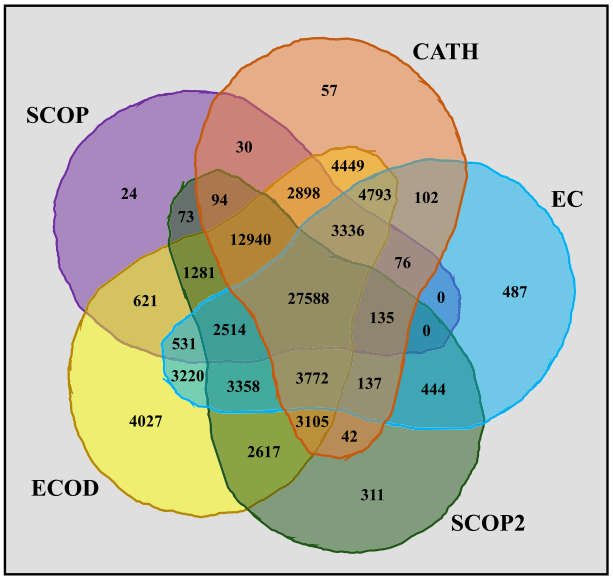


Fig. 6. Venn diagram representing the occurrence of proteins in the classification schemes, i.e., SCOP, CATH, EC, SCOP2, and ECOD.

6. Analysis of plane-plane angles

We have looked closely into the interplanar angles of the obtained π-π networks to understand the preferential arrangements of the aromatic residues in the protein networks. The distribution of the range of interplanar angles observed in each of the nπ networks (n = 2–14) at different cut-off radii (Figs. 4 and 5) has shown a general trend of the aromatic residues having a higher preference to accommodate themselves in T-shape geometries (perpendicular) rather than the stacked orientation. Similar observations have been reported by Živković et al. in their study of aromatic-aromatic interactions in crystal structures of proteins [37]. The interplanar angles in the 2π networks have shown a greater preference for a C—H···π type of geometry at higher cut-off distances i.e., ≥5.5 Å. At the cut-off radius of 5.0 Å, there is an uneven distribution of the interplanar angles with specific peaks at 11–20° and 160–170° angles rather than a right-angled orientation. The most interesting observations were made at the cut-off distances ≤4.5 Å, where the most preferred angles were the parallel or anti-parallel stacked geometries, while the C—H···π type geometries are the least favorable, forming reversed-gaussian distributions. Ninković et al. have reported substantial stability in stacked benzene dimers at large horizontal displacements, suggesting that the overall stability of the stacked aromatic residues in proteins is almost undisturbed, even at larger distances [25]. A closer inspection into the distribution of interplanar

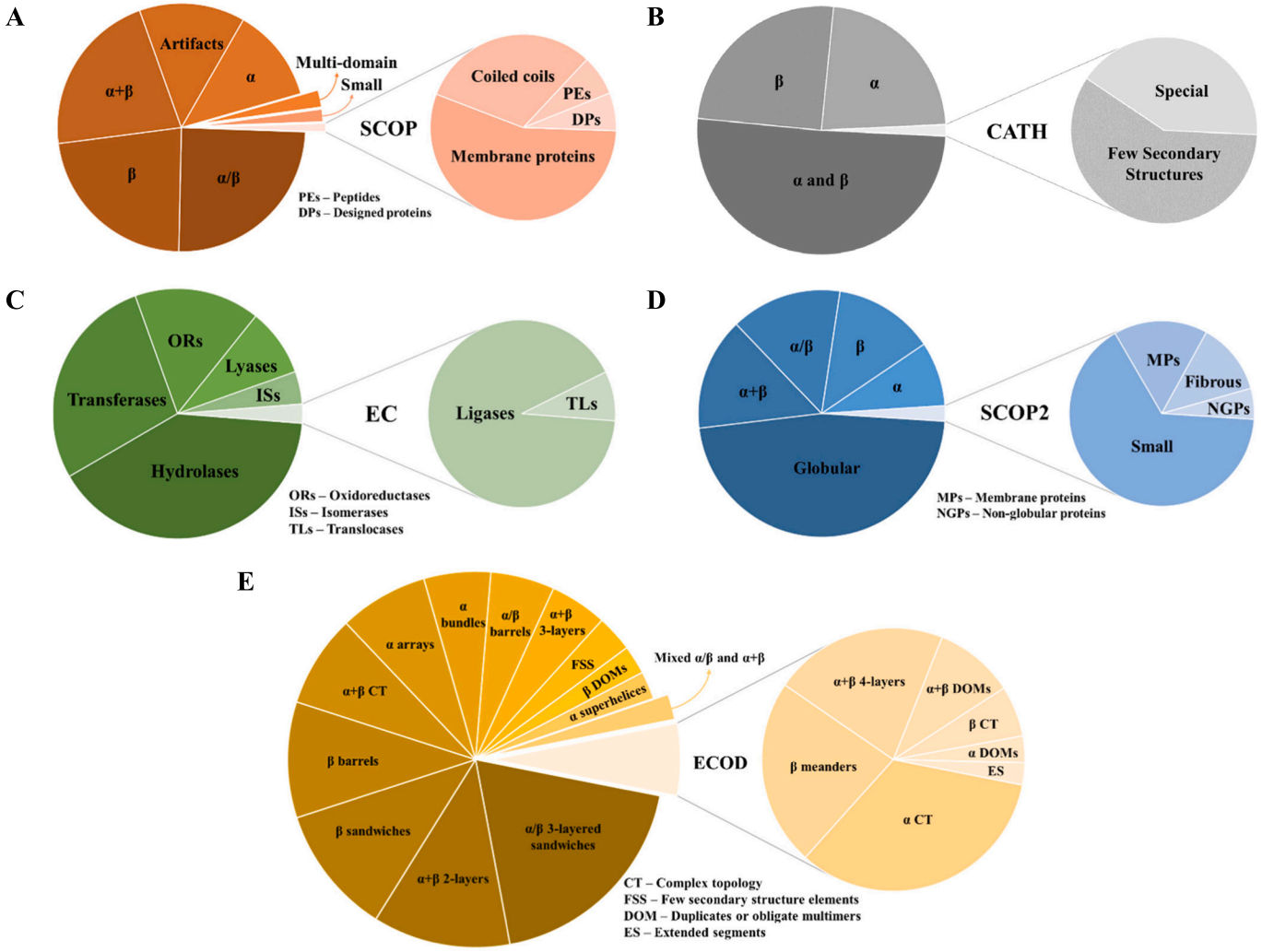


Fig. 7. The distribution of different classes in the (A) SCOP, (B) CATH, (C) EC, (D) SCOP2, and (E) ECOD classification schemes for the proteins considered for this study.

Table 2
Mutations in π-π networks and their effects on mutant variants of Human p53 core domain protein.

PDB IDs	Name	Mutated Residues in π-π networks	Network Type	Mutation Effects
2WGX	M133L, V203A, Y236F, N239Y, T253I, N268D	F236-Y234	2π	Y236F and T253I, improve hydrophobic core packing, which in turn leads to higher stability of the core domain.
6SI1	Y220H	H220-F109	2π	Y220H mutations drastically impair the thermostability of the p53
2J21	M133L, V203A, N239Y, N268D, R282W	W282-F134	2π	The R282W Mutation Affects the Packing of the Loop–Sheet–Helix Motif.
2BIM	M133L, V203A, N239Y, N268D, R273H	H273-F134	2π	DNA contact mutations R273H, with little effect on the overall stability of the protein but impaired function because of the loss of a residue mediating DNA contacts.
4LOF	V157F, N235K, N239Y	F157-F109	2π	V157F mutation in p53 protein leads to instability, but suppressor mutations can restore stability.
4LOE	N239Y	H179 -H178- Y107-Y239	4π	Biochemical evidence shows that suppressor mutations N235K and N239Y individually restore stability to V157F, and their combined effect is stronger than either mutation alone.
4KVP	V157F	Y236-Y234-Y220-F157	4π	V157F mutation in p53 protein leads to instability, but suppressor mutations can restore stability.

angles over the diverse range of cut-off radii considered (Figs. 4 and 5), has shown that the majority of aromatic-aromatic networks have an abundance of the arene motifs interacting side-ways leading to the C—H···π interactions rather than forming stacked motifs, while this curve often flattened or is inverted at lower cut-off radii.

7. Distribution of π-π networks in different protein classes

In addition to analyzing the geometrical parameters of π-π networks, we also examined their distribution across various classification schemes (SCOP, CATH, EC, SCOP2, and ECOD). The number of proteins in each class is represented in Figs. 6–7 and Tables S3–S12. In SCOP 67,337 proteins were categorized into 12 classes, but low-resolution proteins were excluded, leaving 2,661,069 π-π networks. CATH classified 75,326 proteins into five categories, resulting in 3,098,421 π-π networks. The EC classification scheme included six enzyme classes, with an additional class for translocases. A total of 56,591 proteins were classified, generating 2,472,856 π-π networks, where only 125 proteins belong to translocases forming 18,203 networks. SCOP2, a more comprehensive version of SCOP, had five classes and an additional four integrated classes. There were 122,908 non-unique proteins in SCOP2, resulting in 4,867,376 π-π networks. ECOD, with 20 classes, identified 102,665 non-unique proteins and 4,685,516 π-π networks. The analysis of aromatic-aromatic networks per protein revealed insights into their importance in protein stability. The number of networks per protein was higher in α and β (α/β) proteins compared to α, β, or α and β (α + β) proteins according to SCOP, SCOP2, and ECOD. Multi-domain and membrane proteins exhibited a higher number of networks per protein. In the enzyme classes, translocases and oxidoreductases had a relatively higher number of proteins and networks. The advanced search feature of the database allows users to retrieve specific information for each class without conflicts caused by non-unique protein structures in the classification schemes.

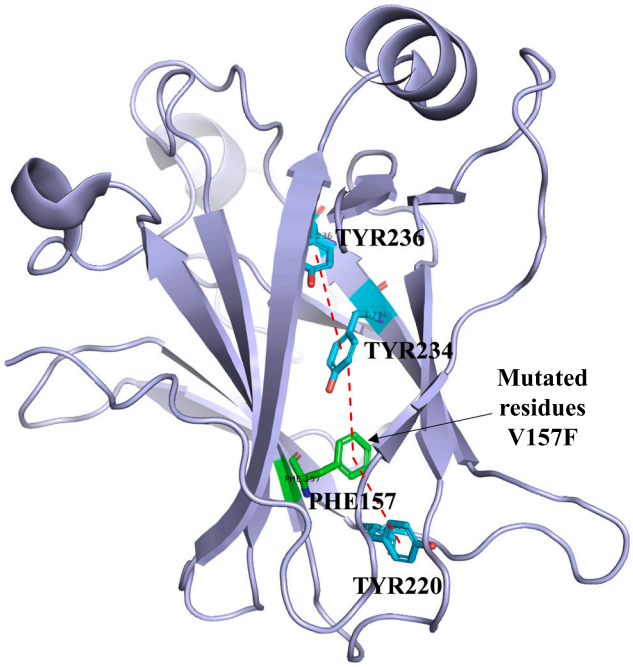


Fig. 8. 4π-networks in the Human p53 Core Domain Mutant V157F (4KVP) involving residues 236(TYR), 234(TYR), 220(TYR), and 157(PHE).

8. Case study: Impact of π-π network disruptions on mutant p53 variants

This case study examines the effects of disruptions in the π-π network on mutant variants of the human p53 core domain protein. Our analysis reveals that mutations in p53 significantly affect the π-π networks in the protein, leading to diverse consequences (Table 2). For instance, in the case of the 2WGX mutant structure, the Y236F and T253I mutations have been found to improve the packing of aromatic residues in the hydrophobic core. This enhancement leads to higher stability of the core domain, imparting favorable effects on the protein. On the other hand, the Y220H mutation in the 6SI1 mutant structure significantly impairs the thermo-stability of p53. This emphasizes the crucial role of π-π interactions involving residue 220 (HIS) in maintaining protein stability. Disruption of these interactions may negatively impact the overall structural integrity of the protein. Another notable example is the R282W mutation in the 2 J21 mutant structure, which affects the packing of the Loop-Sheet-Helix motif. This alteration potentially disrupts the π-π interactions within this structural motif, thereby undermining the stability of the protein. Conversely, the 2BIN mutant structure with the H168R mutation does not exhibit any identified effects on the π-π networks.

In the 2BIM mutant structure, the R273H mutation impairs DNA contacts, indicating that π-π interactions involving residue 273 (HIS) play a vital role in mediating these contacts. Disruption of these interactions leads to a compromised ability of the protein to interact with DNA, thereby affecting its function. The V157F mutation is commonly observed in mutant structure of proteins such as 4LOF, 4LOE, and 4KVP. This mutation induces minor changes in the side chains within the hydrophobic core, resulting in protein instability (Fig. 8). The database provides a platform to corroborate the loss or gain of π-π interactions

upon mutation and their consequences on the stability. The N239Y mutation is another common occurrence observed in various mutant structure of proteins, including 2WGX, 2 J21, 2BIN, 2BIM, 4LOF, 4LOE, and 4KVP. This mutation potentially affects π-π interactions involving residue 239 (tyrosine) and can influence the stability and function of the protein. Investigating the precise role of these π-π interactions is essential for a comprehensive understanding of the structural stability, functional implications, and potential associations with diseases involving the p53 protein. Overall, the disruptions in the π-π network within the p53 core domain have significant implications for the structural and functional properties of the protein.

9. Conclusions

In the current update, about 3,444,619 π-π networks of varying lengths and arrangements have been identified from 84,500 protein structures. We have further probed into the occurrence of different aromatic residues and looked into understanding the differences in proteins at the topology level by analyzing the π-π networks at different classifications. An in-house FORTRAN program is used to identify the aromatic residues within a given protein and determine their connectivity and networks. This program enables the calculation of important geometric parameters such as **r** and **ϕ**. This information is then utilized to identify the arrangement and structure of these networks. While the **r** provides information about the proximity between arene ring pairs within a network, they do not offer insight into the nature of interactions. On the other hand, **ϕ** for each pair of π-π networks offers a comprehensive understanding of the interactions involved. It reveals whether the interaction is a π-π interaction or a C—H···π interaction, providing a more complete picture of the nature of these interactions. It has been observed that the orientation of the π-π networks tends to be perpendicular to each other rather than either parallel or antiparallel orientation, meaning that a vast majority of these networks are stabilized by C—H···π type of interactions rather than π-stacking. The abundance of PHE and TRP residues is found to be the highest and lowest in the π-π networks, respectively. The π-π networks at lower cut-off radii do not exhibit a distinct preference for either stacking or right-angled interactions. However, as the cut-off values increase, there is a noticeable shift in the tendency to form either C—H···π or N—H···π type interactions, compared to stacking interactions. It is observed that the translocases from the EC and extended segments from ECOD have given the highest number of networks per protein with 145.62 and 134.41 networks per protein respectively suggesting that these protein classes make use of the extra stability imparted by the π-π interactions to attain a stable structure.

Author contribution statement

Y. Bhargav Kumar; Manuscript draft preparation, Database development, Formal analysis, Reviewing and editing.

Nandan Kumar; Manuscript draft preparation, Formal analysis.

S. Vaikundamani; Database development.

Selvaraman Nagamani; Reviewing and editing, Visualization.

Hridoy Jyoti Mahanta; Reviewing and editing, Visualization.

G. Madhavi Sastry; Methodology, Original code, Reviewing and editing.

G. Narahari Sastry; Conceptualization, Methodology, Supervision, Manuscript writing and Finalizing.

Declaration of competing interest

Authors have no conflict of interest.

Data availability

Data will be made available on request.

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).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2023.127207.

References

[1] M. Chourasia, G.M. Sastry, G.N. Sastry, Int. J. Biol. Macromol. 48 (2011) 540–552.
[2] A.S. Mahadevi, G.N. Sastry, Chem. Rev. 113 (2013) 2100–2138.
[3] A.S. Mahadevi, G.N. Sastry, Chem. Rev. 116 (2016) 2775–2825.
[4] J. Singh, J.M. Thornton, FEBS Lett. 191 (1985) 1–6.
[5] S.K. Burley, G.A. Petsko, FEBS 203 (1986) 139–143.
[6] C.A. Hunter, J.K.M. Sanders, J. Am. Chem. Soc. 112 (1990) 5525–5534.
[7] C.A. Hunter, J. Singh, J.M. Thornton, J. Mol. Biol. 218 (1991) 837–846.
[8] L. Serrano, M. Bycroft, A.R. Fersht, J. Mol. Biol. 218 (1991) 465–475.
[9] G.B. McGaughey, M. Gagne, A.K. Rappe, J. Biol. Chem. 273 (1998) 15458–15463.
[10] N. Kannan, S. Vishveshwara, Protein Eng. 13 (2000) 753–761.
[11] K.S. Kim, P. Tarakeshwar, J.Y. Lee, Chem. Rev. 100 (2000) 4145–4185.
[12] A.S. Reddy, G.M. Sastry, G.N. Sastry, Proteins: Struct., Funct., Bioinf. 67 (2007) 1179–1184.
[13] R.M. Vernon, P.A. Chong, B. Tsang, T.H. Kim, A. Bah, P. Farber, H. Lin, J. D. Forman-Kay, elife 7 (2018), e31486.
[14] Z. Xu, Q. Zhang, J. Shi, W. Zhu, J. Chem. Inf. Model. 59 (2019) 3389–3399.
[15] A. Jain, R. Sankararamakrishnan, J. Chem. Inf. Model. 51 (2011) 3208–3216.
[16] K.A. Wilson, J.L. Kellie, S.D. Wetmore, Nucleic Acids Res. 42 (2014) 6726–6741.
[17] B. Sharma, H.K. Srivastava, G. Gayatri, G.N. Sastry, J. Comput. Chem. 36 (2015) 529–538.
[18] N. Kumar, A.S. Gaur, G.N. Sastry, J. Chem. Sci. 133 (2021) 97.
[19] D. Vijay, H. Sakurai, G.N. Sastry, Int. J. Quantum Chem. 111 (2011) 1893–1901.
[20] D. Umadevi, G.N. Sastry, ChemPhysChem 14 (2013) 2570–2578.
[21] S. Saha, G.N. Sastry, J. Phys. Chem. B 119 (2015) 11121–11135.
[22] K. Ding, S. Yin, Z. Li, S. Jiang, Y. Yang, W. Zhou, Y. Zhang, B. Huang, J. Chem. Inf. Model. (2022) 1734–1743.
[23] E.C. Lee, D. Kim, P. Jurečka, P. Tarakeshwar, P. Hobza, K.S. Kim, J. Phys. Chem. A 111 (2007) 3446–3457.
[24] M.O. Sinnokrot, E.F. Valeev, C.D. Sherrill, J. Am. Chem. Soc. 124 (2002) 10887–10893.
[25] D.B. Ninković, J.P. Blagojević Filipović, M.B. Hall, E.N. Brothers, S.D. Zarić, ACS Cent. Sci. 6 (2020) 420–425.
[26] V. Philip, J. Harris, R. Adams, D. Nguyen, J. Spiers, J. Baudry, E.E. Howell, R. J. Hinde, Biochemistry 50 (2011) 2939–2950.
[27] E. Kuzniak-Glanowska, M. Glanowski, R. Kurczab, A.J. Bojarski, R. Podgajny, Chem. Sci. 13 (2022) 3984–3998.
[28] U. Purushotham, G.N. Sastry, Theor. Chem. Accounts 131 (2012) 1–4.
[29] U. Purushotham, G.N. Sastry, J. Comput. Chem. 35 (2014) 595–610.
[30] U. Purushotham, ChemistrySelect 3 (2018) 3070–3078.
[31] U. Purushotham, D. Vijay, G.N. Sastry, J. Comput. Chem. 33 (2012) 44–59.
[32] J. Shao, B.P. Kuiper, A.M. Thunnissen, R.H. Cool, L. Zhou, C. Huang, B.W. Dijkstra, J. Broos, J. Am. Chem. Soc. 144 (2022) 13815–13822.
[33] A. Andreeva, D. Howorth, C. Chothia, E. Kulesha, A.G. Murzin, Nucleic Acids Res. 42 (2014) D310–D314.
[34] H. Cheng, R.D. Schaeffer, Y. Liao, L.N. Kinch, J. Pei, S. Shi, B.H. Kim, N.V. Grishin, PLos Comput. Biol. 10 (2014), e1003926.
[35] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, J. Mol. Biol. 247 (1995) 536–540.
[36] A. Bairoch, Nucleic Acids Res. 28 (2000) 304–305.
[37] J.M. Živković, I.M. Stanković, D.B. Ninković, S.D. Zarić, Cryst. Growth Des. 21 (2021) 1898–1904.