Received: 31 May 2023 | Revised: 9 August 2023 | Accepted: 11 August 2023

DOI: 10.1002/prot.26581

RESEARCH ARTICLE

PROTEINS WILEY

Comparative analysis of permanent and transient domain–domain interactions in multi-domain proteins

Swayam Prakash Das Sidhanta[1] | Ramanathan Sowdhamini[1,2,3] | Narayanaswamy Srinivasan[1†]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India

[2]Computational Approaches to Protein Science, National Centre for Biological Sciences, Bangalore, India

[3]Computational Biology, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India

Correspondence
Ramanathan Sowdhamini, Molecular Biophysics Unit, Indian Institute of Science, Bangalore, 560012, India.
Email: mini@ncbs.res.in

Funding information
Department of Biotechnology, Ministry of Science and Technology, India, Grant/Award Number: BT/PR40187/BTIS/137/9/2021; Institute of Bioinformatics and Applied Biotechnology, Grant/Award Number: IBAB/MSCB/182/2022; Science and Engineering Research Board, Grant/Award Number: JBR/2021/000006

Abstract

Protein domains are structural, functional, and evolutionary units. These domains bring out the diversity of functionality by means of interactions with other co-existing domains and provide stability. Hence, it is important to study intra-protein inter-domain interactions from the perspective of types of interactions. Domains within a chain could interact over short timeframes or permanently, rather like protein–protein interactions (PPIs). However, no systematic study has been carried out between two classes, namely permanent and transient domain–domain interactions. In this work, we studied 263 two-domain proteins, belonging to either of these classes and their interfaces on the basis of several factors, such as interface area and details of interactions (number, strength, and types of interactions). We also characterized them based on residue conservation at the interface, correlation of residue motions across domains, its involvement in repeat formation, and their involvement in particular molecular processes. Finally, we could analyze the interactions arising from domains in two-domain monomeric proteins, and we observed significant differences between these two classes of domain interactions and a few similarities. This study will help to obtain a better understanding of structure–function and folding principles of multi-domain proteins.

KEYWORDS
domain repeats, hydrophobic interactions, multi-domain proteins, normal mode analysis, protein folds, residue conservation

1 | INTRODUCTION

The existence and functioning of any organism can be seen to be solely due to proteins in its cellular environment. Most of the functionalities arise due to several interactions of proteins with various macromolecular entities like nucleic acids, lipids, carbohydrates, and so forth, or with other proteins. Among these interactions, protein–protein interactions (PPIs) are of primary importance as these interacting complexes play crucial roles in several cellular processes like replication, transcription, translation, regulation, signaling, and so forth.[1–3] These protein–protein complexes (PPCs) can be categorized into different groups based on the proportion of interacting protomers or stability of protomers or the lifetime of interactions into homo/hetero complexes or obligate/non-obligate complexes or permanent/transient complexes, respectively.[4] The complexes where the protomers become unstable when they are separated are obligates, while the complexes where the interactors remain stable even though they are separated are non-obligates. On the other hand, the complexes where the protomers interact throughout their functional lifetime are permanent, while the complexes where the protomers associate and dissociate temporarily are transient complexes. In

This article is dedicated to the memory of late Prof. Narayanaswamy Srinivasan.

† Deceased.

general, obligate complexes are permanent both structurally and functionally,[4] while non-obligate complexes are mostly transient associations[5] with a few permanent associations[6] like antibody–antigen complexes.[4] One of the great examples of such interaction could be the heterotrimeric G protein. G protein consists of three subunits: α, β, and γ, where β and γ subunits interact throughout their lifetime, making it a permanent interaction. Instead, the α subunit interacts transiently to βγ complex when inactive and dissociates when active, making it a transient interaction. Among these, transient interactions are of utmost importance as these complexes are crucial for various biological processes as they act as hubs in protein–protein interaction networks,[7] are multi-specific,[4,7] are great drug targets[6,8,9] and are involved in various cellular processes.[10]

There are several studies that distinguish structural characteristics of such interaction types amongst proteins and PPIs. Few such physicochemical properties which discriminate permanent and transient interactions are contact area,[4,10,11] interface shape and size,[12] number of contacts,[11] polarity,[4] hydrophobicity,[4,12,13] complementarity of the interface,[12,14] involvement of secondary structures at the interface,[11] evolution of the interface,[15] and so forth. Based on these properties, several groups focused on distinguishing these two types of PPIs. Few groups focused solely on the physicochemical properties[16,17] or interfacial properties[18] to predict permanent and transient PPIs. Some groups represented these physicochemical properties into vectors[19–21] for better prediction using machine learning approaches, while some researchers used desolvation energy explicitly to predict permanent and transient.[22–25] Apart from these, some used sequence features to predict permanent or transient[26,27] and few developed algorithms which do not require information about binding partners for prediction.[28]

After the first enzyme was solved,[29] it was found that there were some distinct lobes present within the protein. However, the term "domain" was coined by Wetlaufer[30] by defining these entities as structurally independent regions within proteins. These domains are also referred to as units of protein evolution.[31] Several structure-based identification and analysis of domains have been performed and organized as databases.[32,33] Many algorithms such as DIAL,[33] PUU,[34] DETECTIVE,[35] DOMAK,[36] Protein Peeling,[37] KluDo,[38] a method by Islam et al.[39] and so forth, use different approaches to delineate protein structural domains. Despite the increase in the number of methods to assign structural domains, there has been a gap in the consensus of domain assignments by different algorithms, and still an active area of research till now.

The vast functional diversity of a protein arises by combining such domains into a single polypeptide chain, calling it a multi-domain protein, and most proteins, even in a simple proteome, are multi-domain proteins.[40] The interactions amongst multi-domain proteins with other such proteins are mainly carried out by a portion of the protein structure, a protein domain, rather than the whole protein.[41] The interactions between domains are called domain–domain interactions (DDIs), and they generally facilitate protein interactions. It is also observed that interacting domain pairs tend to co-evolve with each other in an interaction[40] in order to maintain a better interaction. The domain pairs are also consistent with their parent protein interactions.[42] There are few studies that take help of known structural DDIs to predict PPIs, whether these are input sequences[43] or structures.[44] Deng et al.[45] used maximum likelihood approach to estimate the probabilities of domain pairs in protein interactions to predict PPIs. Gonzalez and Liao[46] used fisher scores derived from the domain interaction profiles as features to predict DDI using SVM, which can be used to predict PPI. Instead of using generative methods of predicting PPI, Zhao et al.[47] used information of both PPI and non-PPI to infer DDI, which in turn can be used again to predict PPI from the inferred DDIs. Similarly, Sprinzak and Margalit[48] used correlated sequence signatures in proteins to predict DDI.

Often, functional characteristics of a multi-domain protein are dependent on the arrangement of the domains in it and interactions among them,[49,50] which can be compared to the arrangement of words to form meaningful sentences in natural languages.[51] Interactions among the domains facilitate proper functioning of multi-domain proteins. These domains are also known to be responsible for functional and evolutionary relationships of proteins. The occurrence of multiple domains also confers additional stability to individual domains[52,53] and hence the whole protein. Hence, there is a need to study inter-domain interactions, mostly in monomeric proteins, for their resident time of interactions or strength of interactions which could provide immense knowledge about the functional and structural aspects of multi-domain proteins. However, unlike studies differentiating PPIs into permanent and transient interactions, there is no systematic and organized approach[54] to classify DDIs into permanent and transient interactions. Instead, there are a few studies which investigate DDIs in a single polypeptide chain and regard such interactions to be either permanent interactions[26,55] or to have characteristics intermediate between PPIs.[56]

In this work, we extended the concept of permanent and transient interactions to intra-protein inter-domain interactions and characterized the underlying interaction types. We defined permanent domain interactions as those which interact throughout their functional and structural lifetime and transient domain interactions as those which often do not. Using a dataset of monomeric two-domain proteins whose domain definitions are taken from SCOPe,[57] we could identify such domain interactions to be either permanent or transient. Permanent and transiently interacting domains are not much different in terms of evolution of the interface, and the type of functions they are involved in, when investigated human proteome only. However, we found that these two types of DDI differ in the physicochemical properties of their interface, dynamically correlated motion of their residues, and preference for choosing its interacting partner. This work would shed light on the principles of domain interactions, prediction of domain orientation, and protein functioning by these rules of domain interactions. Structurally, this study would also help to understand the folding of multi-domain proteins correctly in the near future.

2 | MATERIALS AND METHODS

2.1 | Dataset creation: Monomeric two-domain proteins

We used SCOPe[57] (dir.cla.scope 2.07) database for structural domain definitions. Entries or proteins having either a single domain or only a single domain available in the database were removed. Furthermore, some unsuitable SCOPe classes (such as low-resolution protein structures, peptides, designed proteins, and artifacts belonging to classes I, J, K, and L, respectively) were removed. For the analyses to be conducted on a nonredundant protein set, a 40% sequence identity was set for clustering proteins using CD-HIT.[58] The resulting entries were filtered for monomeric proteins solved by x-ray crystallographic method in RCSB[59] filter using parameters like asymmetric unit, biological unit, experimental method, and structures with 3 Å or better resolution. Proteins having only two domains were next alone considered through SCOPe definitions (only continuous domains were taken). Finally, the structure having the best resolution was taken as the representative structure for the RCSB entries of proteins. The various filtering steps for dataset creation are summarized in Figure 1.

2.2 | Identification of domain–domain interactions

Identification was done only for those protein structures whose domains interact with each other. To define interacting domains, 5–5 rule was used, which states that interacting domains have at least five residue contacts within 5 Å.[60] The distance criterion was adopted using our in-house PIC software.[61] Furthermore, at least five interactions (all kinds of interactions provided by PIC were taken into consideration) arising from residues of domains were considered that are at least six residues apart to consider a short linker[62,63] connecting two domains, which would include linkers of varying lengths. The classification of DDIs in monomeric multi-domain proteins to obligate and non-obligate (here permanent and transient) ones were done using NOXclass.[16] This tool is an SVM classifier which is based on the physicochemical properties of the interface. For this study, we used the parameters which showed highest accuracy using multi-stage SVM. Necessary PDB manipulations were done using pdb-tools.[64] To get a cutoff to define the interaction as obligate or non-obligate, this was tested on Block et al.[19] dataset and tried to match the accuracy of prediction of NOXclass with different cutoffs, resulting in a cutoff of 70% to distinguish the interaction as obligate and non-obligate (please see supplementary information and Figure S1). We also used multiple structures of proteins to check large structural deviations (>2 Å) using MUSTANG,[65] and reclassified the interaction, wherever needed, based on literature.

2.3 | Interfacial properties

We used our in-house server, PIC,[61] to obtain interactions arising from two domains by taking care of multiple occupancies of atoms. PIC provides multiple types of interactions between protein entities, and we considered all types of interactions to perform different analyses.

A Python script obtained from Pymol (The PyMOL Molecular Graphics System, Schrödinger, LLC) was used for the recognition of interfacial residues, which is based on the change in solvent accessibility upon complex formation.

Interaction energies of DDIs were calculated using PPCheck,[66] which measures energies as sum of van der Waals, hydrogen bond, and electrostatic interactions. The energy of the proteins was minimized using GROMACS[67] for those proteins which showed unfavorable calculated energies.

We calculated the normalized propensities of amino acids at the interface with the following formula:

$$\frac{\frac{\text{No. of amino acid X in the interface of DDI type 1}}{\text{No. of amino acid X in the interface of both DDI types}}}{\frac{\text{Total no. of amino acids in the interface of DDI 1}}{\text{Total no. of amino acids in the interface of both DDI types}}},$$

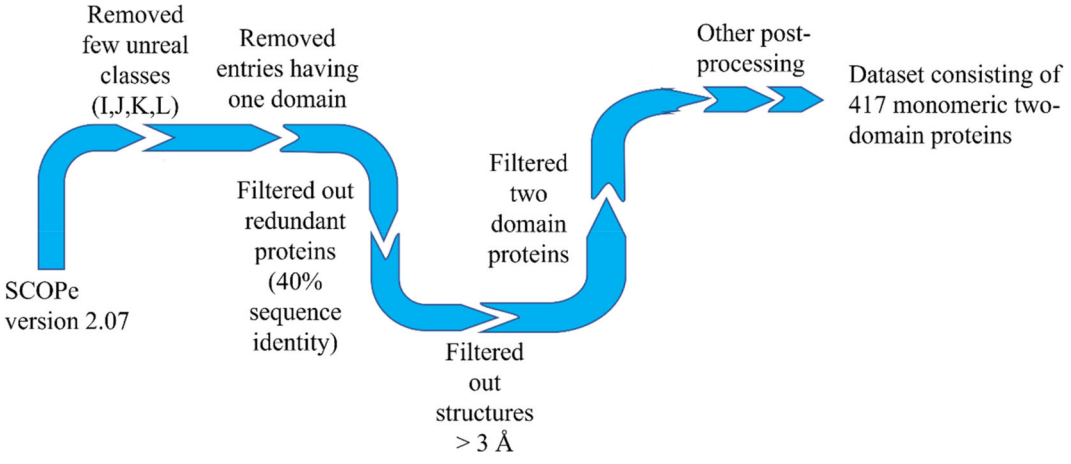

FIGURE 1 Methods and criteria for dataset creation.

where "amino acid X" can be any one of the 20 amino acids and "DDI type 1" can be any one of the two DDI types (permanent or transient) at once. Change in normalized propensity (Δ normalized propensity) was calculated by subtracting the propensity values of the small interface dataset from the propensity values of the whole dataset.

2.4 | Gene ontology studies

We used PANTHER[68] to carry out gene ontology studies on both interaction types. As these studies are difficult for a dataset containing multiple genomes, we only considered the highest occurring genome in the dataset, that is, human genome, consisting of 24 and 25 two-domain proteins containing permanent and transient domains, respectively. We considered major GO terms to evaluate biological processes, molecular functions, and protein classes which could discriminate permanent and transient domain containing proteins.

2.5 | Conserved interfacial residues

ConSurf-DB[69] was employed to identify conserved residues across domains. It is a database for evolutionary rates of residues of a protein of known structure. We used the ConSurf colors greater than 7 to define conserved positions. Common residues to both ConSurf and interface residues were considered as conserved interfacial residues.

2.6 | Correlated residue movement

ProDy[70] was used to perform Anisotropic Network Model (ANM) based Normal Mode Analysis (NMA) calculations. Twenty modes were calculated, and the same were used to calculate cross-correlated motions of C^{α} atoms, keeping a cutoff of 0.7 correlation value to define high correlation. Residue cross-correlation of domain 1 with domain 2 was only considered.

2.7 | Repeats analysis

Uniprot[71] was used to know the presence of sequence repeat-containing proteins in the dataset. RepeatsDB[72] was used to get structural repeats populating at least one domain in proteins in the dataset. We used SCOPe "sccs" id till superfamily level to define homodomain containing proteins and used fold information to get folds of domains.

3 | RESULTS AND DISCUSSIONS

3.1 | Intra-protein domain interactions can be further classified as permanent and transient

Interactions between domains in multi-domain proteins are now possible to study with the availability of larger such entries in structural databanks. Moreover, structures are more conserved than sequences,[73] which makes protein studies more accessible. In order to avoid complications of higher-order domain interactions, we have considered only interactions arising from two structural domains within a single polypeptide chain which would eliminate the interfering effect of another domain in other chain(s). For this study, we created a protein structural dataset of monomeric two-domain proteins (Figure 1).

Next, the dataset was classified into DDIs which permanently or transiently interact. As the domain interface and subunit interface are somewhat similar,[56,74] we searched for a tool to predict PPIs as permanent and transient, which could be used for domain interactions within proteins. We chose NOXclass[16] since this is one of the highly accurate classifiers and is easy to use.

From a dataset of 417 monomeric two-domain proteins, only 263 proteins could be classified due to stringent cutoff (please see supplementary information and Figure S1), and the rest of the proteins were of lower confidence. We observed that around 109 proteins retain permanent inter-domain interactions, while around 154 of them showed to have transient interactions (Figure 2A, Table S1) at a stringent cutoff. Comparatively large number of proteins showed to have transient interactions, which could prove their inherent flexibility to accommodate any function of the protein.

There are few studies which compare inter-chain protein interactions to intra-chain interactions and comment on their resident time of interaction. In one such study, the authors analyzed protein interaction sites by taking 750 transient PPIs and 2000 domain interactions within a chain.[55] The authors assumed such domain interactions within the same protein chain as obligate interactions. In another such study, the authors analyzed six different types of interfaces in protein structures, and domain–domain interface within a single chain was one of the six interface types.[26] They viewed such interaction as permanent interaction between independent folding units and compared these with hetero-obligomers. On the other hand, it is also seen that most of the domain interfacial properties within a chain are intermediate between inter-chain permanent and non-obligate complexes.[56] The observations from our study clearly suggest that intra-chain DDIs can also be classified as permanent and transient interactions. Moreover, we could further diverge domain interfaces which are intermediate between permanent and transient PPI into permanent and transient domain interactions, which are discussed in detail in the following sections.

3.2 | Permanent domains have enhanced interfacial properties

To understand the interfacial features of DDIs within a single chain which could be responsible for their permanent and transient behavior, we first investigated the total number of interactions through which two domains are held. We observed that most of the proteins having transient domains have a comparatively smaller number of interactions (Figure 2B). Also, around 91% of the transient domain containing proteins have interactions less than 75 in number. On the

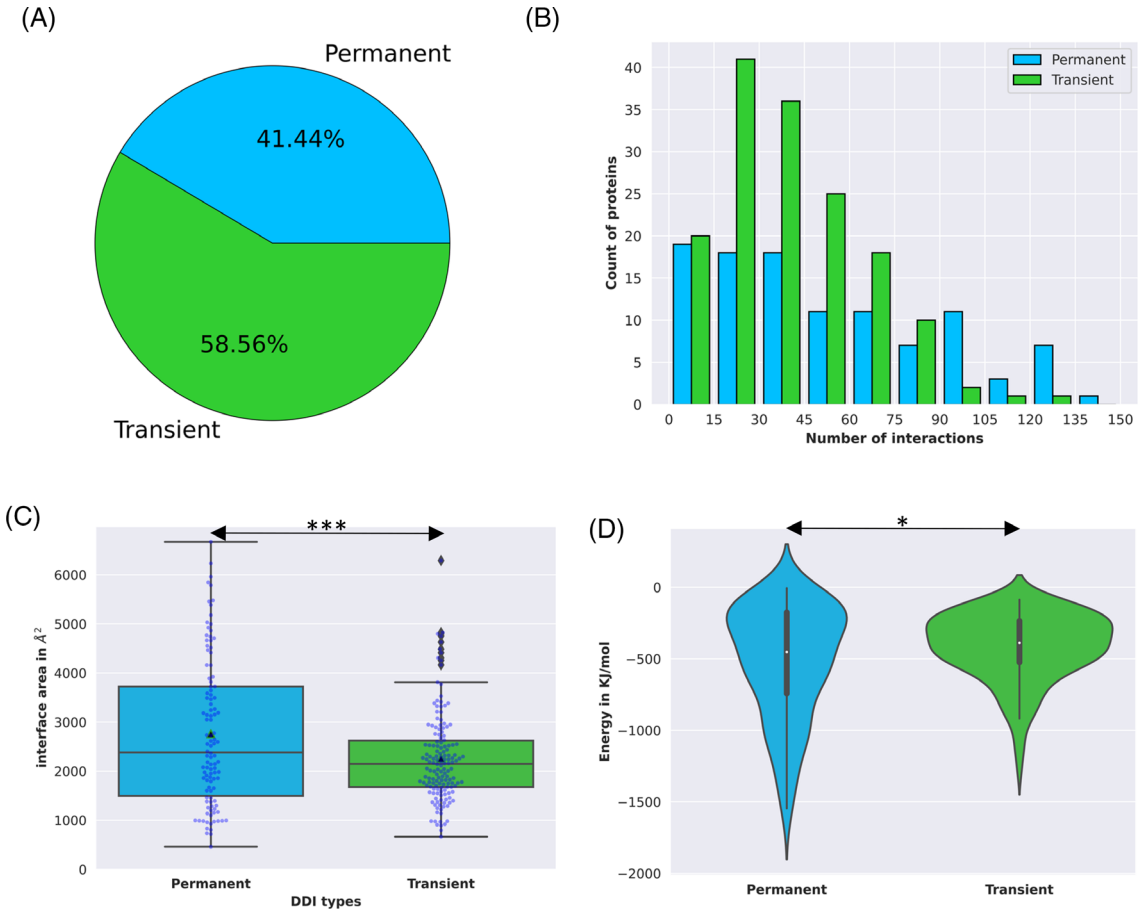

FIGURE 2 Permanent and transient domains and their interfacial properties. (A) Percentage of monomeric proteins having permanent and transient domains. (B) Plot showing number of interactions between permanent and transient domains where *x*-axis shows different ranges of number of interactions and *y*-axis shows the frequency of proteins having such domains. (C) Boxplot showing the interfacial surface area in terms of solvent accessible surface area. The triangle in the box represents the mean and the blue dots represent the data points (here, two-domain proteins). (D) Distribution of interaction energies of permanent and transient domains.

other hand, comparatively more permanent domain proteins have a larger number of interactions (Kolmogorov–Smirnov two-sample test [two-sample KS test]: 0.0009). The total number of interactions between permanent domains follow a near similar uniform distribution throughout different number of interaction ranges. We then computed interfacial areas to find the interface size of different domain–domain interfaces. From Figure 2C, we infer that proteins having transient domain interactions have a comparatively smaller distribution of interface areas than proteins having permanent domain interactions (two-sample KS test: 0.00027). The interfacial areas of transient domain interactions are concentrated around the median of the distribution, which suggests that these domains have smaller interfaces consistently. Instead, permanent domains of proteins have widespread interfacial areas, of which most of the domains have larger interfaces which is evident from comparatively large differences between upper quartiles in the boxplot. The domain pairs which have larger interfaces than 4000 Å^2 are listed in Table S2, and most of the large transient domain interfaces are outliers (Figure 2C). This shows that permanent domains harbor larger elaborate interfaces than transient domains. The better the interaction energy of a complex, the stronger the binding and stability. For this case, we next checked the strength of domain interactions in these two kinds of interacting domains. From Figure 2D, it is observed that permanent domains indeed have better interaction energies than transient domains, as measured through PPCheck.[66] The energies associated with permanent domains are more stabilizing than transient domains (two-sample KS test: 0.011). It is also observed that the energies of transient domains are concentrated to a comparatively lower stabilizing energy, while the energies of permanent domains have a wide range of interaction strengths. Next, we looked at the amino acid preferences at the domain interfaces. From Figure S2A, we observed that permanent domain interfaces are highly populated with nonpolar residues with a few exceptions like Asparagine and Lysine. On the other hand, transient domain interfaces mainly consist of polar charged and uncharged residues except for Proline, which might provide irregularity to the transient interface.

It is noteworthy that the number of permanent and transient domain containing proteins in our dataset is not equal, where we had comparatively a greater number of domain pairs in transient interactions than permanent ones. Therefore, we sampled a random number

of entries from transient domain pair dataset to match permanent domain pair dataset, and the observed trends are very similar to asymmetric dataset. Although permanent domains have higher interfacial physical properties, it is also observed that both permanent and transient have similar average number of interactions per interfacial residue, 0.685 and 0.641 interactions per interfacial residue, respectively, which would mean that the residue interaction networks at the interface are not much different. The average number of interactions per interfacial residue is a proportional value and hence could be the reason for such similarity. The interface of such domain interaction types reveals many discriminatory facts that would help to distinguish permanent and transient domain interactions.

3.3 | A tie between hydrophobic interactions and hydrogen bonds: permanent and transient domains

It is clear from the previous analysis that permanent domains have a larger number of interactions between interacting domains than transient domains. To obtain a clearer perspective on the interactions, we probed atomic interactions of interfacial residues. We observed around 52% of the total number of interactions in the case of proteins having permanent domains are hydrophobic (Figure 3A). On the contrary, as shown in Figure 3B, transient domains have only 37% of hydrophobic interactions. These hydrophobic interactions are known to drive different PPIs[75] and are known to comprise major interactions in the biomolecules which stabilize interacting complexes.[76] Moreover, we found that transient domains have large proportions of side chain associated hydrogen bonds in comparison to permanent domains (Figures 3A, B, and S3), and such polar interactions are known to bring out specificity.[77] Apart from hydrophobic interactions and side chain associated hydrogen bonds, all other interaction types were similarly populated in the interface, which would be required for sustained domain interactions and the functioning of multi-domain proteins.

3.4 | Residues of permanent domains have higher correlated motion

A variety of functions of proteins are achieved by cooperative motions of their constituent atoms. This cooperativity is further achieved by crosstalk among domains of the proteins either by physical contacts or by correlated motions of its atoms. To explore any discrimination of dynamic behavior of residues between permanent and transient domains, we studied residue cross-correlation of motions, which are represented by C^{α} atoms using ANM of NMA. Cross-correlation values range from −1 to 1, and we considered those residue pairs to be highly correlated if their value is greater than 0.7 to keep a balance between the number of correlated residues and their high correlation. Figure 4 shows the differences in inter-residue correlated motions between domains interacting permanently and transiently. Residues from permanent domain pairs showed a wider range of highly correlated motions than residues from transient domains. When the data were plotted in a histogram to better understand the difference, we found a maximum number of transient domain pairs to have extremely low percentages of highly correlated residues, while consistently, a greater number of permanent domain pairs showed a higher percentage of highly correlated residues (Figure S4). The same analysis was tested with different cutoffs to define high residue correlations ranging from 0.5 to 0.9, and the patterns obtained were similar. This observation implies that the domains which interact transiently carry out short-range correlated motions, whereas, in case of permanent domains, extensive interactions are going on across domains and

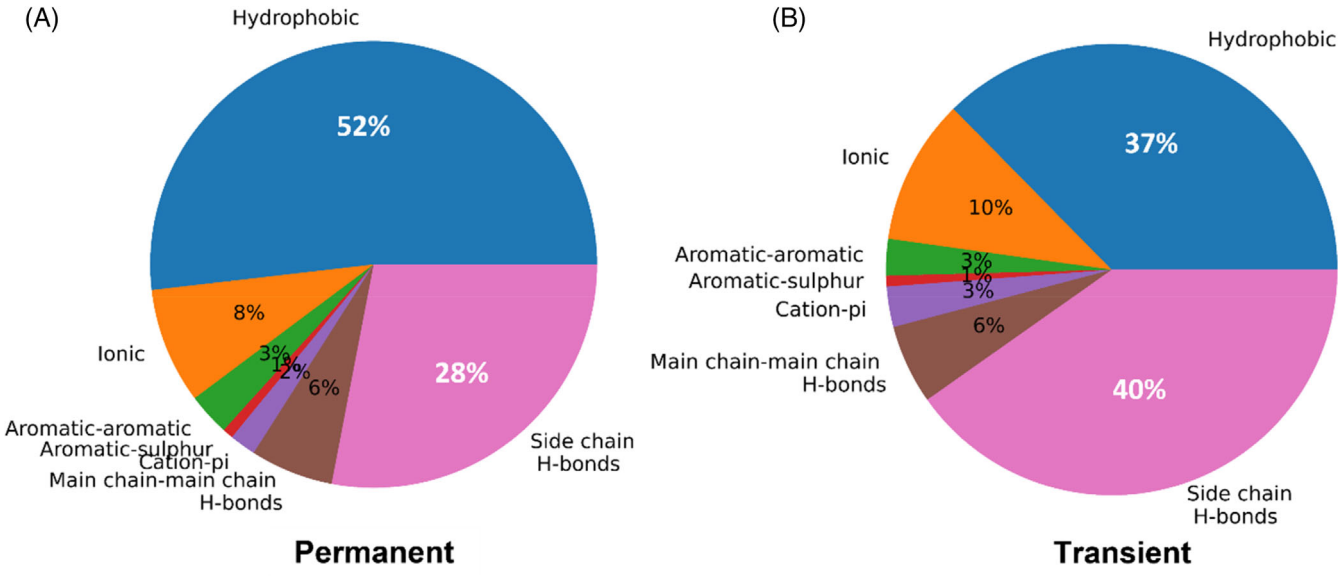

FIGURE 3 Different types of biomolecular interactions that exists between permanent (A) and transient (B) domains. The percentages imply average number of interactions per total number of interactions. Most discriminating interaction types are labeled in white in the pie chart.

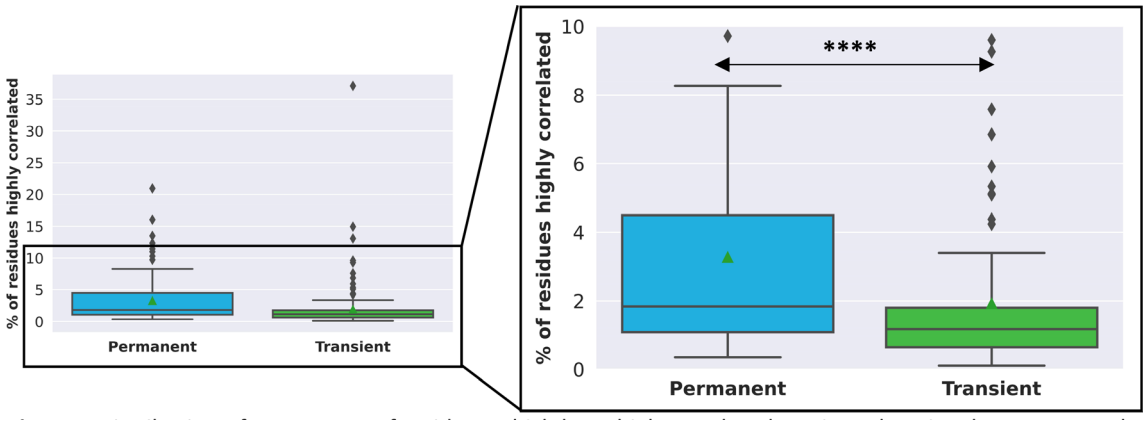

FIGURE 4 Distribution of percentage of residues which have high correlated motion. The triangle represents the mean of the distribution. Permanent vs transient two-sample KS test: 4.57e-05.

can engage in long-range correlated motion. The range here conveys the strength of interaction or force of movement using a large number of residues (long) or a small number of residues (short).

Such behavior of permanent domains could be thought of as due to their lifetime of interactions. These residues are needed to be synchronous to maintain the integrity of the domain interface, and this correlated motion would help the domains to maintain resonance for the stability of the monomeric protein. On the other hand, transient domains would need to associate and dissociate frequently. Comparatively lower percentage of highly correlated residues between domains would be enough to maintain the interface and hence the transient nature. This clearly conveys how dynamics is associated with the long-term interactions within protein interiors.

3.5 | Number of conserved interfacial residues is similar in permanent and transient domains

Residues present in the interface of interacting partners are solely responsible for communication between the partners. There are several studies on PPI, which state higher conservation of interfacial residues than remaining protein surface.[78–80] It is also known that interfaces of permanent protein complexes have a lower evolution rate than transient interactions, which allows better co-evolution with its interacting protein partner.[15] Similarly, in this case, we ought to look into the conservation at the domain interfaces of different DDI types in a protein chain. Due to the requirement of ConSurf-DB[69] to have at least a certain number of homologues to follow the evolutionary rate, some of the proteins in our dataset could not be retrieved from the database. Hence, we analyzed only 101 and 147 permanent and transient domain pairs in proteins, respectively. We computed the absolute number of interfacial residues which are conserved in domain pairs and observed that permanent domain pairs have a little wider distribution of the number of conserved interfacial residues than transient domain pairs, which could be due to a large number of interfacial residues arising from larger interfaces. But, the number

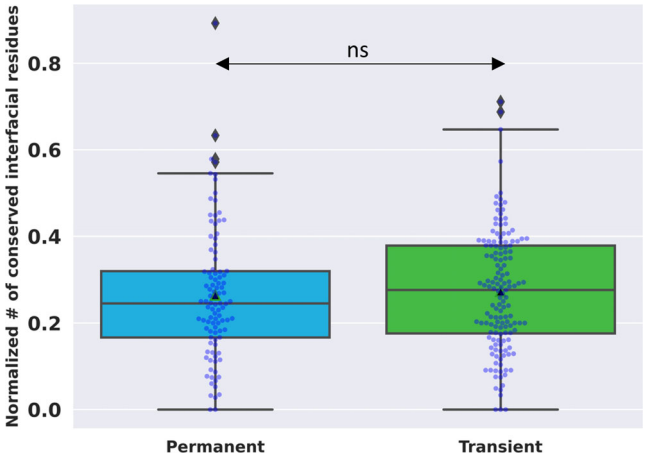

FIGURE 5 Distribution of number of conserved interfacial residues that are normalized to total number of interfacial residues. The black triangle represents the mean and blue dots represent the data points (here, normalized number of conserved domain interfacial residues in proteins).

of such residues in both domain types is not significantly different to account for any dissimilarity (with a Mann–Whitney *p*-value: 0.1341 and two sample two-sample KS test: 0.2874). Next, we also computed the normalized number of conserved interfacial residues, that is, the number of conserved interface residues per total number of interface residues in a domain pair of a protein, and we observed near similar distribution of such residues with respect to their interface residues with near identical mean and median (Figure 5). These observations suggest that both permanent and transient domain pairs have similar proportions of conserved interfacial residues. Unlike PPIs, this similarity in maintaining the conservation at the interface could be due to the fact that the interacting partners, here domains, are referred to as semi-independent and evolutionary units which are thought to be conserved. Like permanent PPI, residues in the permanent domains might be under evolutionary pressure to co-evolve with partner domains. On the contrary, two transient domains might be harboring

functional sites at their interface and hence the obligation to preserve the interfacial residues. Each domain interaction type has its compulsion to maintain the interface geometry, resulting in similar preferences to conserve the domain interfaces in the two-domain protein.

3.6 | Permanent domains structurally prefer similar folds

Different types of domain interactions in a protein chain might have some influence on the anatomy of the protein structure landscape. Hence, we next explored a few of the structural aspects which could be discriminated by permanent and transient domain interactions within a protein. First, we aimed to look at their preferences to have repeats. Repeats can be of two types, viz, sequence repeats and structural repeats. Structural repeats can be further classified into different classes.[81] Using different databases (see methods) to map the proportion of proteins in our dataset to have repeats, we found a few proteins in both repeat types where proteins having permanent domain interactions showed a little more preference for sequence and structural repeats. However, this observation cannot be relied upon due to the sparse number of proteins. Among the structural repeats, the repeating units (domains) of bead-on-string repeats (class-IV) are thought to either interact loosely or not interact,[81] which could have been interesting examples of transient domains in multi-domain proteins. However, from the proteins having at least one structural repeat-containing domain, there was no protein which belonged to this class. This could be due to the limited amount of information in the database or due to the limited number of domains in our study to represent multi-domain proteins. Second, to overcome this limitation, we defined homodomains, where both domains have same class, fold, and superfamily according to SCOPe.[57] Thus, these domains will have similar architecture and are evolutionarily related to each other, which are supposed to be originated by duplication.[82] Using such a definition, we observed a comparatively higher proportion of permanent domain containing proteins to have homodomains, 37.3% in comparison to 28.6% of homodomains in the dataset. Although these homodomains may not be true tandem repeats, such domains can provide functional and structural advantages to the proteins having permanent domains due to evolutionary pressure and topological constraints, respectively. Third, to investigate their structural constraints, we explored their fold distribution in homodomains. We found that proteins having permanent homodomains have a comparatively lower number of unique folds than transient homodomains, which could signify the capability to re-use folds. This suggests that if domains interact permanently in a protein, there is a greater chance of finding another interacting domain of common ancestry and similar structural topology. This observation is similar to the observations of PPI, where obligate PPI tends to have more homo-DDIs.[22,24] When we considered the whole dataset to look into the number of unique folds, both permanent and transient domain pairs showed a similar count of unique folds quantitatively. However, qualitatively, we observed a few biases of folds toward permanent and transient domain interactions (Table 1). Superfolds[83] such as TIM beta/alpha-barrel, OB fold, and beta-grasp showed an inclination toward transient domains. On the other hand, 7-bladed beta-propeller, Ribonuclease H-like motif fold, and a few others showed inclinations toward permanent domains. Apart from that, superfolds[83] like Immunoglobulin-like beta-sandwich, and DNA/RNA-binding 3-helical bundle showed preferences for both permanent and transient domains. Other sparsely occurring folds (frequency: less than 5) showed little or no bias (Tables S3, S4). These observations show the structural preferences of different domain interaction types and also justify how a limited number of folds are re-used to sample various protein structural landscapes in DDI following a power-law.[49,84–86] This will enlighten the basic principles of domain interaction type prediction, given that we

TABLE 1 Some of the highly occurring folds in the dataset.

Number of folds in permanent domains	Fold name (number of superfamilies[a])	Number of folds in transient domains
19	Immunoglobulin-like beta-sandwich (33)	13
10	7-bladed beta-propeller (15)	1
11	DNA/RNA-binding 3-helical bundle (14)	9
2	TIM beta/alpha-barrel (33)	20
6	P-loop containing nucleoside triphosphate hydrolases (1)	15
8	Ribonuclease H-like motif (7)	2
7	ADP-ribosylation (1)	0
6	Concanavalin A-like lectins/glucanases (1)	5
5	OB-fold (17)	10
5	Beta-Trefoil (8)	2
5	Alpha–alpha superhelix (28)	2
5	Ferredoxin-like (62)	2
5	Spectrin repeat-like (16)	0
0	Glycosyl hydrolase domain (1)	10
0	Reductase/isomerase/elongation factor common domain (4)	9
0	Beta-Grasp (ubiquitin-like) (15)	9
0	Ferredoxin reductase-like, C-terminal NADP-linked domain (1)	6
4	SH3-like barrel (21)	5

Note: Rows are marked with colors if the count difference is more than 60%. *Blue*: folds preferring transient domains and *Gray*: folds preferring permanent domains.

[a]Number of superfamilies the fold has according to SCOPe[57] version 2.07.

know the interacting domains in a protein, their topology, and evolutionary information.

3.7 | Functional classification of domain interaction types using gene ontology

Next, we explored if there are functional preferences between permanent and transient domains. As our dataset consists of proteins from various genomes, we only considered the genome which populates the maximum number of proteins in our dataset. Using gene ontology functional classification analysis, we observed that both permanent and transient domain containing proteins in humans are involved in similar kinds of biological processes, molecular functions and belong to similar protein classes, and there is no bias (Figure S5). The variance in functional roles may be seen if the whole dataset is compared, and this needs more sophisticated algorithms, which are out of scope at present.

4 | CONCLUSIONS

Interactions between domains are responsible for the functionality of a protein. Apart from functional advantages, domains in multi-domain proteins provide additional stability to proteins and to the neighboring domains. Studying types of domain interactions in multi-domain proteins, focusing on their resident time becomes crucial to understand the intra-protein interactions. In this current work, we recognized DDI arising from two domains in a monomeric multi-domain protein as permanent and transient using an algorithm used to classify PPIs. These sorts of analyses can become more complex for the large number of discontinuous domains, which we have not considered in our study. We demonstrate that permanently interacting domains have larger interfaces that facilitate a larger number of interactions between the domains, which in turn support stronger interactions. Their interfaces are populated by a larger proportion of hydrophobic interactions, while transient domain interfaces have comparatively lower hydrophobic interactions, which are compensated by a large number of side chain associated hydrogen bonding. A comparatively increased number of residues in permanent domains have highly correlated motions. Domains interacting permanently have a higher chance of interacting with a structurally similar domain, and there are a few topological biases for each interaction type. Furthermore, both permanent and transient domains have equal number of conserved interfacial residues, and the domains in the human genome do not discriminate upon the functions or processes they are associated with. We note that few of these observations are consistent with the way permanent and transient PPIs differ from each other.

This work will be very useful to understand the molecular basis of function and how the functional sites are disposed in 3D structures. This analysis provides objective realization that two-domain monomeric proteins which are permanently interacting are more likely to adorn their interface by hydrophobic residues. This observation is certainly of predictive value to obtain clues on biochemical function and to recognize reasonable poses while performing domain–domain docking and modeling.

AUTHOR CONTRIBUTIONS

Swayam Prakash Das Sidhanta: Methodology; data curation; investigation; formal analysis; visualization; writing – original draft. **Ramanathan Sowdhamini:** Investigation; formal analysis; supervision; funding acquisition; writing – review and editing. **Narayanaswamy Srinivasan:** Conceptualization; investigation; formal analysis; supervision; funding acquisition; project administration; resources; writing – review and editing.

ACKNOWLEDGMENTS

The authors thank lab members for their helpful discussions. Swayam Prakash Das Sidhanta thanks Dr. Ashraya Ravikumar for discussions on the work. Swayam Prakash Das Sidhanta thanks Dr. Hongbo Zhu for providing NOXclass files. Swayam Prakash Das Sidhanta thanks CSIR for its financial support. Ramanathan Sowdhamini acknowledges funding and support provided by JC Bose Fellowship (JBR/2021/000006) from Science and Engineering Research Board, India and Bioinformatics Centre Grant funded by Department of Biotechnology, India (BT/PR40187/BTIS/137/9/2021). Ramanathan Sowdhamini would also like to thank Institute of Bioinformatics and Applied Biotechnology for the funding through her Mazumdar-Shaw Chair in Computational Biology (IBAB/MSCB/182/2022).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/prot.26581.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in RCSB at https://www.rcsb.org/. These data were derived from the following resources available in the public domain: - RCSB, https://www.rcsb.org/.

ORCID

Swayam Prakash Das Sidhanta https://orcid.org/0009-0003-3707-5452
Ramanathan Sowdhamini https://orcid.org/0000-0002-6642-2367
Narayanaswamy Srinivasan https://orcid.org/0000-0002-6566-6239

REFERENCES

1. Levy ED, Pereira-Leal JB. Evolution and dynamics of protein interactions and networks. *Curr Opin Struct Biol*. 2008;18(3):349-357. doi:10.1016/j.sbi.2008.03.003
2. Reichmann D, Rahat O, Cohen M, Neuvirth H, Schreiber G. The molecular architecture of protein–protein binding sites. *Curr Opin Struct Biol*. 2007;17(1):67-76. doi:10.1016/j.sbi.2007.01.004

3. Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease. *Cell*. 2011;144(6):986-998. doi:10.1016/j.cell.2011.02.016
4. Nooren IMA, Thornton JM. Diversity of protein–protein interactions. *EMBO J*. 2003;22(14):3486-3492. doi:10.1093/emboj/cdg359
5. Janin J, Bahadur RP, Chakrabarti P. Protein–protein interaction and quaternary structure. *Q Rev Biophys*. 2008;41(2):133-180. doi:10.1017/S0033583508004708
6. Acuner Ozbabacan SE, Engin HB, Gursoy A, Keskin O. Transient protein–protein interactions. *Protein Eng Des Sel*. 2011;24(9):635-648. doi:10.1093/protein/gzr025
7. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient protein–protein interactions: structural, functional, and network properties. *Structure*. 2010;18(10):1233-1243. doi:10.1016/J.STR.2010.08.007
8. Ohlson S. Designing transient binding drugs: a new concept for drug discovery. *Drug Discov Today*. 2008;13(9–10):433-439. doi:10.1016/J.DRUDIS.2008.02.001
9. Rudolph J. Inhibiting transient protein–protein interactions: lessons from the Cdc25 protein tyrosine phosphatases. *Nat Rev Cancer*. 2007;7(3):202-211. doi:10.1038/nrc2087
10. Nooren IMA, Thornton JM. Structural characterisation and functional significance of transient protein–protein interactions. *J Mol Biol*. 2003;325(5):991-1018. doi:10.1016/S0022-2836(02)01281-0
11. De S, Krishnadev O, Srinivasan N, Rekha N. Interaction preferences across protein–protein interfaces of obligatory and non-obligatory components are different. *BMC Struct Biol*. 2005;5(1):1-16. doi:10.1186/1472-6807-5-15
12. Jones S, Thornton JM. Principles of protein–protein interactions. *Proc Natl Acad Sci USA*. 1996;93(1):13-20. doi:10.1073/pnas.93.1.13
13. Lo CL, Chothia C, Janin J. The atomic structure of protein–protein recognition sites. *J Mol Biol*. 1999;285(5):2177-2198. doi:10.1006/JMBI.1998.2439
14. Bahadur RP, Chakrabarti P, Rodier F, Janin J. A dissection of specific and non-specific protein–protein interfaces. *J Mol Biol*. 2004;336(4):943-955. doi:10.1016/J.JMB.2003.12.073
15. Mintseris J, Weng Z. Structure, function, and evolution of transient and obligate protein–protein interactions. *Proc Natl Acad Sci USA*. 2005;102(31):10930-10935. doi:10.1073/PNAS.0502667102/SUPPL_FILE/INDEX.HTML
16. Zhu H, Domingues FS, Sommer I, Lengauer T. NOXclass: prediction of protein–protein interaction types. *BMC Bioinformatics*. 2006;7(1):27. doi:10.1186/1471-2105-7-27
17. Kottha S, Schroeder Biotec M, Dresden T. Classifying Permanent and Transient Protein Interactions. 2006.
18. Vennmann E, Schneider N, Rarey M. Discrimination of permanent and transient heterodimers: The real challenge in the classification of protein–protein interactions. Published online, 1–9. 2012.
19. Block P, Paern J, Hüllermeier E, Sanschagrin P, Sotriffer CA, Klebe G. Physicochemical descriptors to discriminate protein–protein interactions in permanent and transient complexes selected by means of machine learning algorithms. *Proteins: Struct Funct Genet*. 2006;65(3):607-622. doi:10.1002/prot.21104
20. Liu Q, Li J. Propensity vectors of low-ASA residue pairs in the distinction of protein interactions. *Proteins: Struct Funct Bioinformatics*. 2009;78(3):589-602. doi:10.1002/prot.22583
21. Mintseris J, Weng Z. Atomic contact vectors in protein–protein recognition. *Proteins: Struct Funct Genet*. 2003;53(3):629-639. doi:10.1002/prot.10432
22. Maleki M, Hall M, Rueda L, Using structural domains to predict obligate and non-obligate protein–protein interactions. *IEEE symposium on computational intelligence and computational biology, CIBCB 2012*. Published online. Vol 2012; 2012:9-15. doi:10.1109/CIBCB.2012.6217204
23. Maleki M, Aziz MM, Rueda L. Analysis of relevant physicochemical properties in obligate and non-obligate protein–protein interactions. *IEEE international conference on bioinformatics and biomedicine workshops, BIBMW 2011*. Published online. Vol 2011; 2011:345-351. doi:10.1109/BIBMW.2011.6112397
24. Maleki M, Mominul Aziz M, Rueda L. Analysis of obligate and non-obligate complexes using Desolvation energies in domain–domain interactions. Published online 2011.
25. Rueda L, Banerjee S, Aziz MM, Raza M. Protein–protein interaction prediction using desolvation energies and interface properties. *Proceedings—2010 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2010*. Published online. Vol 2010:17-22. doi:10.1109/BIBM.2010.5706528
26. Ofran Y, Rost B. Analysing six types of protein–protein interfaces. *J Mol Biol*. 2003;325(2):377-387. doi:10.1016/S0022-2836(02)01223-8
27. Goebels F, Frishman D. Prediction of protein interaction types based on sequence and network features. *BMC Syst Biol*. 2013;7(6):16-18. doi:10.1186/1752-0509-7-S6-S5
28. La D, Kong M, Hoffman W, Choi YI, Kihara D. Predicting permanent and transient protein–protein interfaces. *Proteins: Struct Funct Bioinformatics*. 2013;81(5):805-818. doi:10.1002/PROT.24235
29. Phillips DC. The three-dimensional structure of an enzyme molecule. *Sci Am*. 1966;215(5):78-90. doi:10.1038/SCIENTIFICAMERICAN1166-78
30. Wetlaufer DB. Nucleation, rapid folding, and globular Intrachain regions in proteins. *Proc Natl Acad Sci USA*. 1973;70(3):697-701. doi:10.1073/PNAS.70.3.697
31. Orengo CA, Thornton JM. Protein families and their evolution—a structural perspective. *Annu Rev Biochem*. 2005;74(1):867-900. doi:10.1146/annurev.biochem.74.082803.133029
32. Vinayagam A, Shi J, Pugalenthi G, Meenakshi B, Blundell TL, Sowdhamini R. DDBASE2.0: updated domain database with improved identification of structural domains. *Bioinformatics*. 2003;19(14):1760-1764. doi:10.1093/BIOINFORMATICS/BTG233
33. Sowdhamini R, Blundell TL. An automatic method involving cluster analysis of secondary structures for the identification of domains in proteins. *Protein Sci*. 1995;4(3):506-520. doi:10.1002/PRO.5560040317
34. Holm L, Sander C. Parser for protein folding units. *Proteins: Struct Funct Bioinformatics*. 1994;19(3):256-268. doi:10.1002/PROT.340190309
35. Swindells MB. A procedure for detecting structural domains in proteins. *Protein Sci*. 1995;4(1):103-112. doi:10.1002/PRO.5560040113
36. Siddiqui AS, Barton GJ. Continuous and discontinuous domains: an algorithm for the automatic generation of reliable protein domain definitions. *Protein Sci*. 1995;4(5):872-884. doi:10.1002/PRO.5560040507
37. Gelly JC, de Brevern AG, Hazout S. 'Protein peeling': an approach for splitting a 3D protein structure into compact fragments. *Bioinformatics*. 2006;22(2):129-133. doi:10.1093/BIOINFORMATICS/BTI773
38. Taheri-Ledari M, Zandieh A, Shariatpanahi SP, Eslahchi C. Assignment of structural domains in proteins using diffusion kernels on graphs. *BMC Bioinformatics*. 2022;23(1):369. doi:10.1186/S12859-022-04902-9
39. Islam SA, Luo J, Sternberg MJE. Identification and analysis of domains in proteins. *Protein Eng Des Sel*. 1995;8(6):513-526. doi:10.1093/PROTEIN/8.6.513
40. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol*. 2004;14(2):208-216. doi:10.1016/J.SBI.2004.03.011
41. Itzhaki Z, Akiva E, Altuvia Y, Margalit H. Evolutionary conservation of domain–domain interactions. *Genome Biol*. 2006;7(12):1-12. doi:10.1186/GB-2006-7-12-R125/TABLES/4
42. Memišević V, Wallqvist A, Reifman J. Reconstituting protein interaction networks using parameter-dependent domain–domain interactions. *BMC Bioinformatics*. 2013;14(1):1-15. doi:10.1186/1471-2105-14-154/FIGURES/5
43. Aloy P, Russell RB. InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics*. 2003;19(1):161-162. doi:10.1093/BIOINFORMATICS/19.1.161

44. Li X, Yang L, Zhang X, Jiao X. Prediction of protein–protein interactions based on domain. *Comput Math Methods Med*. 2019;2019:1-7. doi:10.1155/2019/5238406
45. Deng M, Mehta S, Sun F, Chen T. Inferring domain–domain interactions from protein–protein interactions. *Genome Res*. 2002;12(10):1540-1548. doi:10.1101/GR.153002
46. González AJ, Liao L. Predicting domain–domain interaction based on domain profiles with feature selection and support vector machines. *BMC Bioinformatics*. 2010;11(1):1-14. doi:10.1186/1471-2105-11-537/TABLES/3
47. Zhao XM, Chen L. *Domain–Domain Interaction Identification With A Feature Selection Approach*. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, Heidelberg, Vol 5265; 2008:178-186. doi:10.1007/978-3-540-88436-1_16/COVER
48. Sprinzak E, Margalit H. Correlated sequence-signatures as markers of protein–protein interaction. *J Mol Biol*. 2001;311(4):681-692. doi:10.1006/JMBI.2001.4920
49. Basu MK, Poliakov E, Rogozin IB. Domain mobility in proteins: functional and evolutionary implications. *Brief Bioinform*. 2009;10(3):205-216. doi:10.1093/BIB/BBN057
50. Basu MK, Carmel L, Rogozin IB, Koonin EV. Evolution of protein domain promiscuity in eukaryotes. *Genome Res*. 2008;18(3):449-461. doi:10.1101/GR.6943508
51. Yu L, Tanwar DK, Penha EDS, Wolf YI, Koonin EV, Basu MK. Grammar of protein domain architectures. *Proc Natl Acad Sci USA*. 2019;116(9):3636-3645. doi:10.1073/pnas.1814684116
52. Bhaskara RM, Srinivasan N. Stability of domain structures in multi-domain proteins. *Sci Rep*. 2011;1:40. doi:10.1038/SREP00040
53. Vishwanath S, de Brevern AG, Srinivasan N. Same but not alike: structure, flexibility and energetics of domains in multi-domain proteins are influenced by the presence of other domains. *PLoS Comput Biol*. 2018;14(2):e1006008. doi:10.1371/JOURNAL.PCBI.1006008
54. Winter C, Henschel A, Kim WK, Schroeder M. SCOPPI: a structural classification of protein–protein interfaces. *Nucleic Acids Res*. 2006;34(suppl_1):D310-D314. doi:10.1093/NAR/GKJ099
55. von Eichborn J, Günther S, Preissner R. Structural features and evolution of protein–protein interactions. *Genome Inform*. 2010;22:1-10. doi:10.1142/9781848165786_0001
56. Jones S, Marin A, Thornton JM. Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Eng*. 2000;13(2):77-82. doi:10.1093/protein/13.2.77
57. Fox NK, Brenner SE, Chandonia JM. SCOPe: structural classification of proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res*. 2014;42(D1):D304-D309. doi:10.1093/nar/gkt1240
58. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-1659. doi:10.1093/BIOINFORMATICS/BTL158
59. Berman HM, Westbrook J, Feng Z, et al. The protein data Bank. *Nucleic Acids Res*. 2000;28(1):235-242. doi:10.1093/nar/28.1.235
60. Gong S, Yoon G, Jang I, et al. PSIbase: a database of protein structural Interactome map (PSIMAP). *Bioinformatics*. 2005;21(10):2541-2543. doi:10.1093/BIOINFORMATICS/BTI366
61. Tina KG, Bhadra R, Srinivasan N. PIC: protein interactions calculator. *Nucleic Acids Res*. 2007;35(SUPPL.2):W473-W476. doi:10.1093/nar/gkm423
62. Bhaskara RM, de Brevern AG, Srinivasan N. Understanding the role of domain–domain linkers in the spatial orientation of domains in multi-domain proteins. *J Biomol Struct Dyn*. 2013;31(12):1467-1480. doi:10.1080/07391102.2012.743438
63. George RA, Heringa J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng*. 2002;15(11):871-879. doi:10.1093/protein/15.11.871
64. Rodrigues JPGLM, Teixeira JMC, Trellet M, Bonvin AMJJ. pdb-tools: a swiss army knife for molecular structures. *F1000Research*. 2018;7:1961. doi:10.12688/f1000research.17456.1
65. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. MUSTANG: A multiple structural alignment algorithm. *Proteins: Struct Funct Genet*. 2006;64(3):559-574. doi:10.1002/prot.20921
66. Sukhwal A, Sowdhamini R. Oligomerisation status and evolutionary conservation of interfaces of protein structural domain superfamilies. *Mol Biosyst*. 2013;9(7):1652-1661. doi:10.1039/C3MB25484D
67. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19-25. doi:10.1016/J.SOFTX.2015.06.001
68. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, Mi H. PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci*. 2022;31(1):8-22. doi:10.1002/PRO.4218
69. Ben Chorin A, Masrati G, Kessel A, et al. ConSurf-DB: an accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci*. 2020;29(1):258-267. doi:10.1002/PRO.3779
70. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*. 2011;27(11):1575-1577. doi:10.1093/BIOINFORMATICS/BTR168
71. Bateman A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506-D515. doi:10.1093/NAR/GKY1049
72. di Domenico T, Potenza E, Walsh I, et al. RepeatsDB: a database of tandem repeat protein structures. *Nucleic Acids Res*. 2014;42(D1):D352-D357. doi:10.1093/NAR/GKT1175
73. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J*. 1986;5(4):823-826. doi:10.1002/J.1460-2075.1986.TB04288.X
74. Argos P. An investigation of protein subunit and domain interfaces. *Protein Eng Des Sel*. 1988;2(2):101-113. doi:10.1093/PROTEIN/2.2.101
75. Keskin O, Gursoy A, Ma B, Nussinov R. Principles of protein–protein interactions: what are the preferred ways for proteins to interact? *Chem Rev*. 2008;108(4):1225-1244. doi:10.1021/CR040409X/ASSET/IMAGES/LARGE/CR040409XF00013.JPEG
76. Chothia C, Janin J. Principles of protein–protein recognition. *Nature*. 1975;256(5520):705-708. doi:10.1038/256705a0
77. Fersht AR. Basis of biological specificity. *Trends Biochem Sci*. 1984;9(4):145-147. doi:10.1016/0968-0004(84)90122-1
78. Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES. Are protein–protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci*. 2004;13(1):190-202. doi:10.1110/PS.03323604
79. Bordner AJ, Abagyan R. Statistical analysis and prediction of protein–protein interfaces. *Proteins: Struct Funct Bioinformatics*. 2005;60(3):353-366. doi:10.1002/PROT.20433
80. Choi YS, Yang JS, Choi Y, Ryu SH, Kim S. Evolutionary conservation in multiple faces of protein interaction. *Proteins: Struct Funct Bioinformatics*. 2009;77(1):14-25. doi:10.1002/PROT.22410
81. Kajava AV. Tandem repeats in proteins: from sequence to structure. *J Struct Biol*. 2012;179(3):279-288. doi:10.1016/J.JSB.2011.08.009
82. Doolittle RF. The multiplicity of domains in proteins. *Annu Rev Biochem*. 1995;64:287-314. doi:10.1146/annurev.bi.64.070195.001443
83. Orengo CA, Jones DT, Thornton JM. Protein superfamilles and domain superfolds. *Nature*. 1994;372(6507):631-634. doi:10.1038/372631a0
84. Koonin EV, Wolf YI, Karev GP. The structure of the protein universe and genome evolution. *Nature*. 2002;420(6912):218-223. doi:10.1038/nature01256
85. Karev GP, Wolf YI, Rzhetsky AY, Berezovskaya FS, Koonin EV. Birth and death of protein domains: a simple model of evolution explains power law behavior. *BMC Evol Biol*. 2002;2(1):1-26. doi:10.1186/1471-2148-2-18

86. Naveenkumar N, Kumar G, Sowdhamini R, Srinivasan N, Vishwanath S. Fold combinations in multi-domain proteins. *Bioinformation*. 2019;15(5):342-350. doi:10.6026/97320630015342

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: Sidhanta SPD, Sowdhamini R, Srinivasan N. Comparative analysis of permanent and transient domain–domain interactions in multi-domain proteins. *Proteins*. 2023;1-12. doi:10.1002/prot.26581