ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 851-860 (2021)
©Biomedical Informatics (2021)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

OPEN ACCESS GOLD

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 17(10)

Research Article

Geometrical and electro-static determinants of protein-protein interactions

Vicky Kumar[2], AshitaSood[1], Anjana Munshi[3], Tarkeshwar Gautam[4] & Mahesh Kulharia[1,*]

[1]Centre for Computational Biology and Bioinformatics, School of Life Sciences, Central University of Himachal Pradesh, Kangra, India, 176206; [2]Department of Computational Sciences, School of Basic and Applied Sciences, CentralUniversity of Punjab, Bathinda, India, 151001; [3]Department of Human Genetics and Molecular Medicine, School of Health Sciences, Central University of Punjab, Bathinda, India, 151001; [4]Department of Zoology, Kalindi College, University of Delhi , Delhi, India, 110008; *Corresponding author: Email: kulharia@gmail.com Phone: +91-99884 28856

Received September 10, 2021; Revised October 12, 2021; Accepted October 12, 2021, Published October 31, 2021
DOI: 10.6026/97320630017851

Declaration on official E-mail:
The corresponding author declares that official e-mail from their institution is not available for all authors

Declaration on Publication Ethics:
The authors state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Abstract:
Protein-protein interactions (PPI) are pivotal to the numerous processes in the cell. Therefore, it is of interest to document the analysis of these interactions in terms of binding sites, topology of the interacting structures and physiochemical properties of interacting interfaces and the of forces interactions. The interaction interface of obligatory protein-protein complexes differs from that of the transient interactions. We have created a large database of protein-protein interactions containing over100 thousand interfaces. The structural redundancy was eliminated to obtain a non-redundant database of over 2,265 interaction interfaces. Therefore, it is of interest to document the analysis of these interactions in terms of binding sites, topology of the interacting structures and physiochemical properties of interacting interfaces and the offorces interactions. The residue interaction propensity and all of the rest of the parametric scores converged to a statistical indistinguishable common sub-range and followed the similar distribution trends for all three classes of sequence-based classifications PPInS. This indicates that the principles of molecular recognition are dependent on the preciseness of the fit in the interaction interfaces. Thus, it reinforces the importance of geometrical and electrostatic complementarity as the main determinants for PPIs.

Keywords:
Protein-protein interactions; protein-protein interaction interface; non-redundant database; residue interface propensity; hydrophobicity; solvation free energy; planarity; protrusion; depth.

Abbreviations:
PPInS, Protein-protein interaction sitesbase; NRDB, Non-redundant database; ACP, Atomic contact pair; PPII, Protein-protein interaction interface; PPIP, Protein-protein interacting patch; RIP, Residue interface propensity; LSS, Low sequence similarity; MSS, Moderate sequence similarity; HSS, High sequence similarity; PDB, Protein Data Bank; PPC, Protein-protein complex

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 851-860 (2021)
©Biomedical Informatics (2021)

Background:
The cellular milieu where proteins perform their function is crowded. However, the spatial and temporal preciseness of interactions is rarely violated. This specificity and accuracy of the interacting proteins determines the fate of cells [1-3]. Protein–protein interactions (PPIs) form the very basis for all biological processes, such as signal transduction, material /energy transport, metabolic reactions, regulation of gene expression, cellular growth and proliferation[3-6]. These interactions form the fundamentals of the intracellular / intercellular communications [7]. Proteins therefore act as communication cogs, transferring the information through conformational changes and triggering corresponding transient structural adjustments in other molecules [8-9]. Understanding the protein interactions and enumerating the precise rules that govern these interactions mediate these molecular communication events can provide us with the deeper insights about metabolic and signaling network dynamics [10]. Any deviation from the normal interaction behavior of a protein leads to the gain or loss of a function, often leading to debilitating diseases such as neuro-degenerative diseases, cancer or even auto-immune problems [11].

The PPI manifests in formation of various types of complexes (*viz.* protein-protein complex, protein-DNA complex, protein-RNA complex, protein-membrane complex, protein-lipid complex, protein-carbohydrate complex, and others). In protein-protein complexes, the interaction of two or more proteins is very specific and is usually characterized by only a small subset of their surfaces [12-15]. Only those sites which own the proper binding features participate in the formation of PPCs [16-20] (Supplementary Figure 1).Such understanding may augment the development of computational tools for PPI sites prediction [9], and drug discovery [20-22]. In this direction, various research groups have examined the protein binding sites with respect to their size, shape, evolutionary conservation, chemical and amino acid composition, change in solvent accessibility of amino acids, and other such parameters. For example, Jones and Thornton, derived a parameter to determine the planarity index of the protein-protein interfaces [20]. As per their study, the average value of the planarity is 3.5 ± 1.7 Å for homodimers and 2.8 ± 0.9 Å for heterocomplexes. Bogan and Thorn showed that evolutionary conserved residues (often termed as hot spot residues) are the major contributor to the binding energy of the interactions [26]. Ozdemir *et al. (*2018) showed that a slight disruption in conserved residues more often results in change in the binding affinity and specificity [27]. Lo Conte *et al.* (1999) described the extent of burial of protein surface, during complex formation. He identified thisto be in the range of 1600 (±400) Å^2 of interaction site for majority of heteromeric protein complexes. Bahadur *et al.* (2004) reported the abundance of aliphatic and aromatic residues and deficit of charges residues (except for Arg) in homodimeric interfaces in conformation with previous studies [21, 28, 29, 30]. We have studied protein-protein interaction interfaces (PPIIs) inPPInS and NRDB [31]. On the basis of sequence similarity between the interacting protein chains thePPIIs from both the datasets were classified into low-sequence, moderate-sequence, and high-sequence similarity classes. All three classes of NRDB dataset were examined for six important parameters of: residue interface propensity, hydrophobic content, solvation energy, compactness of interacting residue's neighborhood, planarity, and depth index.

Materials and Methods
Datasets of protein-protein interaction interfaces

(i) Protein-protein interaction sitesbase (PPInS) has the protein-protein complexes (PPCs)as reported in PDB with their structural classification based onSCOPe (version 2.06). It harbors over 32000 X-ray crystallized structures of PPCs with structural resolution better than 2.5 Å. The information about these PPCs is available in the form of atom contact pairs wherein two atoms belonging to two different protein chains of a PPC were considered to be in contact if the intervening distance between them was less than the sum of their van der Waals radii plus 1 Å as a tolerance factor. We utilized the entire PPInS and its non-redundant form (NRDB) database for our study.

(ii) Categorization of PPIIs based on the sequence similarity in the interacting protein chains
Tostudy the influence of homo or heterodimeric nature of the proteins in PPIs, sequence similarity between the protein chains involved in the PPII was calculated using BLAST [32]. Based on sequence similarity observed, PPIIs were categorized into three classes. If the interacting protein chains were similar homologous up to 49%, then the corresponding PPII was marked under low sequence similarity (LSS) class; the PPIIs with protein chains sharing 50-89% sequence similarity were grouped under moderate sequence similarity (MSS) class; and the PPIIs with protein chains sharing 90-100% sequence similarity were grouped under high sequence similarity (HSS) class. All PPIIs from both of the datasets were categorized into three PPII classes.

Calculation of residues' interface propensity:
Not all the amino acid residues favor their occurrence on the protein surface, some prefer to stay in the protein core thereby avoid or does not contribute much in protein complexation [28, 33-34]. The relative contribution of amino acids in promoting a protein site as the binding site is described as residue interface propensity (RIP) and is defined as the ratio of residue's relative contribution to the protein binding site to its relative contribution to the complete protein surface [35]. To calculate the RIP, PPIIs from the NRDB only were taken into consideration. The area contributed by an amino acid *i* to the protein binding site was calculated as the difference between its solvent accessible surface area (SASA) bearing its unbound and bound states. The propensity of a residue *i* to occur on the protein binding site (θ_i^{PBS}) was calculated using Eq. 1

$$\theta_i^{PBS} = \frac{\Delta SASA_i^{PBS} \Big/ \sum_{j=1}^{20} \Delta SASA_j^{PBS}}{SASA_i^{PBS} \Big/ \sum_{j=1}^{20} SASA_j^{PBS}} \qquad \rightarrow \textbf{Eq. 1}$$

where $\Delta SASA_i^{PBS}$ is the SASA of residue i buried in protein bound state, $\sum_{j=1}^{20} \Delta SASA_j^{PBS}$ is the total SASA of all residues buried in protein bound complexes, $SASA_i^{PBS}$ is the SASA contributed by residue i to the protein surface, and $\sum_{j=1}^{20} SASA_j^{PBS}$ is the total solvent accessible surface area of all residues of the protein.

Analysis of protein-protein interaction interfaces/patches with respect to binding site parameters:

The binding nature of proteins is determined from physicochemical, structural, and evolutionary properties of their constituents favouring the non-covalent interactions with partner protein(s). Such properties bring two molecules closer, influence them for biological interactions, and define the destiny of the PPCs. Knowing the implications of protein binding sites analysis in protein engineering, all of the 223,714 PPIPs from PPInS and 4,530 PPIPs from NRDB were examined with respect to various PPI site parameters. While analyzing a PPIP, if an atom from a PPIP was seen interacting with more than one atom of the partner PPIP, its contribution in PPII formation was considered only for once.

Hydrophobicity:

The hydrophobic residues are reported in abundance of PPI sites [36]. The kinetics of interfaces with predominant hydrophobic residues are reported to be different than hydrophilic ones because the tightly bound aquasphere of surface bound water molecules acts as an additional barrier that has to be removed before direct protein-protein interaction can take place[29]. To determine the level of hydrophobicity (Φ) associated with PPIPs, the hydrophobicity scale for amino acids was used from literature[37]. For each interacting atom in the PPIP, its corresponding hydrophobicity score was obtained by dividing the residue hydrophobicity score by the number of atoms of the residue. The hydrophobicity score of all interacting atoms in the protein-protein interaction patch (PPIP) was calculated by linear augmentation to represent the hydrophobicity score of the PPIP (Eq. 2 and 3)

$$\emptyset_i = \frac{\varphi_{aa}}{N_{aa}} \quad \rightarrow \textbf{Eq. 2}$$

$$PPIP_{\Phi} = \sum_{i=1}^{N} \emptyset_i \quad \rightarrow \textbf{Eq. 3}$$

Where, $\emptyset_i$ represents the average hydrophobicity for i^{th} atom of an amino acid. φ_{aa} represents the hydrophobicity value for a particular amino acid and N_{aa} represents the number of atoms in that amino acid, $PPIP_{\Phi}$ represents the total hydrophobicity score for an interacting patch, and N represents total number of atoms on an interacting surface.

Solvation free energy:

The solvation free energy of amino acids from the interacting protein also influences its kinetics, hence we calculated the solvation free energy (ω) of PPIP by linear summation of the average individual contributions of the interacting atoms (Eq. 4 and 5). The solvation free energy scale for amino acids given by Wimley et al., 1996 [38] was used.

$$\omega_i = \frac{\omega_{aa}}{N_{aa}} \quad \rightarrow \textbf{Eq. 4}$$

$$PPIP_{\omega} = \sum_{i=1}^{N} \omega_i \quad \rightarrow \textbf{Eq. 5}$$

Where, ω_i represents the average solvation energy for i^{th} atom of an amino acid. ω_{aa} represents the hydrophobicity value for a particular amino acid and N represents the number of atoms in the interacting surface and $PPIP_{\omega}$ represents total solvation energy score for an interacting patch, and N represents total number of atoms on an interacting surface.

Size of the interacting patch:

The protein-protein interaction patches are generally very small region on the protein surface withvery specificcertain structural / thermodynamical features. We determined the size of such interaction interfaces by summing up the difference of the solvent accessible surface area (SASA) of the atoms in the bound and unbound forms.

$$\boldsymbol{PPIP_s = SASA_a^U + SASA_b^U - SASA_{a::b}^B} \quad \rightarrow \textbf{Eq. 6}$$

Where, $PPIP_s$ represents size of an interacting patch, $SASA_a^U$ and $SASA_b^U$ represents SASA of chains "a" and "b" in unbound states, and $SASA_{a::b}^B$ represents the SASA of the complex (a::b).

Depth index:

Contrary to the ASA,the depth index of an amino acid indicates the extent to which an amino acid is buried in the protein core. The location of amino acids in proteins is determined using their solvent accessibility. An amino acid is said to exist on protein surface if the sum of solvent accessibilities for all of its constituting atoms is a non-zero value. While the amino acids with zero solvent accessibility are considered to be buried in the protein core [39].The depth index of PPIP (ζ) was computed using PSAIA [30]. The depth index of each interacting atom from a PPIP was summed up and represented as a depth index of the PPIP (Eq. 7)

$$\boldsymbol{PPIP_{\zeta} = \sum_{i=1}^{n} \zeta_i} \quad \rightarrow \textbf{Eq. 7}$$

where n represents the total number of interacting atoms in the PPIP, and ζ_i represents the per-atom depth score of the interacting atom.

Protrusion index:

The protrusion index studies the topology of the interface site and gives the measure of how much dense is the neighborhood of an atom on the protein surface [40].The protrusion index of PPIPs (ψ) was also determined by using PSAIA [30] in a manner similar to the depth index for all non-hydrogen atoms (Eq. 8).

$$\boldsymbol{PPIP_{\Psi} = \sum_{i=1}^{n} \psi_i} \quad \rightarrow \textbf{Eq. 8}$$

where n represents the total number of non-hydrogen interacting atoms in the PPIP, n_i represents the non-hydrogen interacting atom, and ψ_i represents the per-atom protrusion score of the non-hydrogen interacting atom.

Planarity index:

The protein binding sites are flat and circular in shape [41]. The calculation of root mean square deviation of all the surface atoms

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 851-860 (2021)
©Biomedical Informatics (2021)

from the least-squares plane (derived from the surface atoms) gives the planarity index of the interacting interfaces. If all atoms correctly fit a plane, the planarity index comes out to be zero.To calculate the planarity index of the PPIPs, *princip* function of the SURFNET [42] was used. Using *princip*, an equation of plane was derived by employing the coordinates of the interacting atoms in the PPIP. Following this, the root-mean-square deviation (RMSD) of interacting atoms from the derived plane was determined and designated as the planarity index of the PPIPs.

Statistical analysis of PPII parametric scores obtained from PPInS and NRDB analysis:

The overall trends of parameteric score distribution wereapparently very similar. Hence, the statistical aspect of the data was explored. The distributions of PPIP parametric scores (after removal of 1% statistical outliers) for each PPI site parameter from all three PPII classes were taken into the consideration. For each PPI site parameter, the mean and standard deviation of parametric scores were calculated with respect to each PPII class separately. Thereafter, for each PPII class, p-value describing the statistical significance between the parametric score of three PPII classes was calculated using two-tailed ANOVA test.

Results and Discussion:

We examined the PPIPs derived from experimentally determined PPCs in terms of various physicochemical and geometrical properties. Two datasets *viz.* PPInS and NRDB where interaction sites were demarcated based on the inter atomic distance between the constituents of two protein chains of the PPCs, were considered for the study. The collection of ACPs between two protein chains of a PPC were referred as the PPII while the collection of atoms involved in PPII from each interacting protein chains were termed as the PPIPs. All the PPIIs from PPInS and NRDB were categorized into three separate classes *viz.* LSS, MSS, and HSS by looking at the sequence similarity between the protein chains involved in formation of PPII under consideration (Table 1). Out of total 111,857 PPIIs in PPInS, around 73% of PPIIs were formed by the protein chains sharing HSS. Around 25% PPIIs were formed by the protein chains sharing LSS, and only 2% PPIIs were results of the interaction between protein chains sharing MSS. For NRDB, the values for PPIIs with HSS, LSS, and MSS were around 62%, 32%, and 6%, respectively. This showed the presence of homodimers in abundance. The possible reasons for this may be the fact that the origin of life started with interactions in absolutely homologous proteins. However, through the course of evolution, perturbation in the genomic code might have caused the formation of heterologous protein complexes. Similar findings have been reported by (Winter el al., 2002) [43]. The factors which might have played a crucial role in bringing two heterologous protein units closer must be their physicochemical, geometrical and other characteristics.

Table 1: Categorization of PPIIs

Dataset	Number of PPIIs		
	LSS	MSS	HSS
PPInS	27,770	2,591	81,496
NRDB	724	130	1,411

Calculation of residue interface propensity (RIP):

The RIP was calculated separately for all three classes of PPIIs (LSS, MSS, and HSS) of NRDB (Figure 1). The propensity scores that we have obtained are quite similar to those which were earlier proposed [21-22]. The higher propensity for aromatic amino acids (Tyr, Phe, and Trp) and aliphatic hydrophobic amino acids (Met, Cys, Ile, Leu, Val) on interacting interface is reported in these studies unanimously. In other studies [26, 39, 44] too, the aromatic residues were reported in abundance on interaction sites. One reason behind this greater occurrence is the predominant contribution of solvation and hydrophobic effect [45, 46]. The small amino acids such as Ala, Gly, Ser, and Thr is comparatively marginal, they have no specific tendency to either avoid or favour the PPIIs in terms of occurrence. Asn, Gln, and Pro too are borderline PPII avoiders while His slightly favours its occurrence on the PPIIs. It is necessary to point out that His exists in multiple protonation states and our data does not differentiate among these. It is possible that some protonation states could decisively dis/favour the PPIIs, and our data is only an average of the overall effect. The charged amino acids (Lys, Glu, Asp), with the exception of Arg (which was relatively neutral), had the least propensity to occur on PPIIs and this was also reported by many groups [21, 22]. The reduced presence of Glu and Asp on PPIIs is perhaps rooted in their inability to form interaction with aromatic hydrophobic amino acids. Although Lys has the ability to form cation-Pi interaction, yet the conformational entropy associated with multiple single bonds would not favour the PPI.

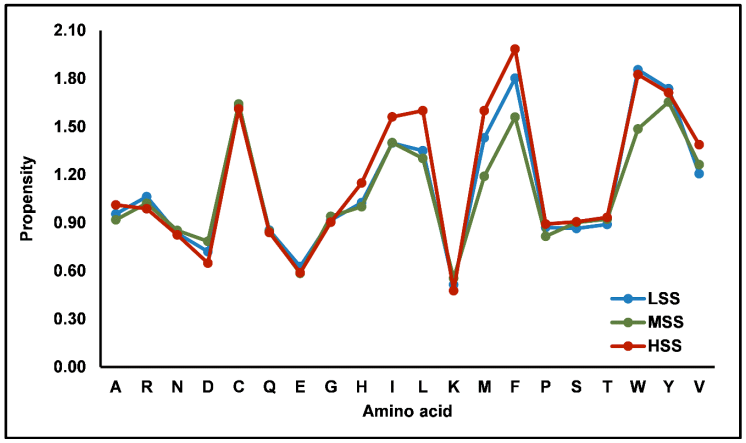

Figure 1: Residue's interface propensity

Table 2: RIP in different PPII classes of NRDB

PPII Class	Relative order of amino acids in terms of their RIP
LSS	K<E<D<N<Q<S<P<T<G<A<H<R<V<L<I<M<C<Y<F<W
MSS	K<E<D<P<Q<N<S<A<T<G<H<R<M<V<L<I<W<F<C<Y
HSS	K<E<D<N<Q<P<G<S<T<R<A<H<V<I<L<M<C<Y<W<F

Analysis of PPIPs from PPInS and NRDB:

The result obtained by analyzing the PPIPs from all the three PPII classes (LSS, MSS, and HSS) of the PPInS and NRDB dataset are given in Table 3 and 4, respectively. The parametric scores for all the PPIPs from LSS, MSS, and HSS are shown graphically in supplementary figures S2-S7*enmasse*. In Table 5, the statistical analysis of PPIP parametric scores is presented in terms of mean,

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 851-860 (2021)
©Biomedical Informatics (2021)

standard deviation, and p-value (calculated using two-tail ANOVA test). The statistical analysis was carried out considering all three PPII classes of both the datasets separately as well as collectively.

Analysis of PPIPs from the PPInS dataset:

The PPIPs were analyzed with respect to hydrophobicity, solvation free energy, depth index, size of the interacting patch, protrusion index, and planarity. It is pertinent to mention that on analyzing PPIPs from PPInS, initially, a significant difference between the parametric scores (obtained with respect to each PPI site parameter) of the three PPII classes was observed (Table 3). However, on removing less than 1% PPIPs (statistical outliers) from each PPII class, the cumulative parametric scores for depth, protrusion, and planarity index of PPIPs from all three classes of PPII reduced down to the identical ranges. Similarly, the cumulative score for solvation free energy and hydrophobic content of PPIPs also were also seen converging to a common sub-range. The possibility that these trends were on account of proportional redundancies of PDB, we looked for the same patterns in the NRDB.

Table 3: Analysis of PPIPs from PPInS

Parameter	Before removal of outliers			After removal of 1% statistical outliers		
	LSS	MSS	HSS	LSS	MSS	HSS
Hydrophobicity	-0.69 to 57.1	-0.39 to 59.12	-1.38 to 114.41	-0.69 to 28	-0.39 to 28	-1.38 to 32
Solvation free energy	-9.37 to 33.71	-5.32 to 32.19	-9.85 to 132.18	-9.37 to 20	-5.32 to 20	-9.85 to 18
Depth	-0.1 to 39.63	0 to 34.15	0 to 127.59	-0.1 to 12	0 to 10	0 to 11
Size of interacting patch	4 to 11532	7 to 11203	3 to 17409	4 to 6000	7 to 5500	3 to 6500
Protrusion	0 to 168.99	0 to 85.2	0 to 182.07	0 to 65	0 to 65	0 to 65
Planarity	0 to 10.85	0 to 9.84	0 to 13.56	0 to 8	0 to 8	0 to 8

Table 4: Analysis of PPIPs from NRDB

Parameter	Before removal of outliers			After removal of 1% statistical outliers		
	LSS	MSS	HSS	LSS	MSS	HSS
Hydrophobicity	-0.58 to 101.11	0.36 to 46.13	-1.38 to 525.05	-0.58 to 36.42	0.36 to 30.34	-1.38 to 93.44
Solvation free energy	-5.67 to 32.65	-5.32 to 32.19	-28.45 to 277.38	-5.67 to 21.08	-5.32 to 22.93	-28.45 to 31.89
Depth	0 to 47.10	0 to 34.15	0 to 172.87	0 to 13.25	0 to 13.04	0 to 27.87
Size of interacting patch	112 to 10876	149 to 5713	143 to 17324	112 to 7048	149 to 5713	143 to 8409
Protrusion	0 to 179.47	0 to 76.28	0 to 1199.49	0 to 79.72	0 to 72.29	0 to 149.55
Planarity	0.18 to 10.76	0.26 to 7.81	0 to 13.45	0.18 to 8.20	0.26 to 7.79	0 to 9.20

Table 5: Statistical analysis of PPIP parametric scores (after removal of 1% statistical outlier)

Parameter	PPII Class	PPInS			NRDB		
		Mean (μ)	S.D. (σ)	p-value	Mean (μ)	S. D. (σ)	p-value
Hydrophobicity	LSS	5.581	5.760	2.8E-96	7.700	7.498	8E-09
	MSS	6.617	6.893		8.306	6.854	
	HSS	6.193	6.389		11.053	21.331	
Solvation free energy	LSS	3.01	4.087	1E-141	3.426	4.129	0.0005
	MSS	3.246	4.721		4.428	5.168	
	HSS	2.537	4.101		4.691	12.218	
Depth	LSS	1.826	2.325	2.6E-07	2.230	2.863	1E-07
	MSS	1.957	2.398		2.687	3.408	
	HSS	1.889	2.664		3.356	7.525	
Size of interacting patch	LSS	1214.913	1144.957	2.3E-52	1538.124	1249.268	4E-14
	MSS	1450.722	1427.384		1811.303	1277.189	
	HSS	1278.987	1225.606		1927.434	1673.538	
Protrusion	LSS	12.231	13.122	1.8E-73	15.487	15.201	2E-05
	MSS	13.712	14.566		17.560	14.690	
	HSS	11.354	12.631		21.155	46.581	
Planarity	LSS	2.091	1.441	9.1E-58	2.642	1.444	0.0002
	MSS	2.368	1.840		3.070	1.720	
	HSS	2.042	1.517		2.790	1.774	

Analysis of PPIPs from NRDB dataset:

The analysis of PPIPs from NRDB considering both the cases, i.e. analysis of all PPIP from each PPII class and analysis of PPIPs after removing less than 1% statistical outlier from each PPII class, also showed the similar trends for the parametric scores for each PPI site parameter (Table 4). Here, the range of parametric scores for each PPIP parameter (except for the planarity) showed the greater variability. After removal of outliers, only the PPIPs from LSS and MSS class found to converge into a common sub-range for parametric scores. However, this could be explained by the methodology of database creation. At the time of NRDB creation, for each pair of SCOP superfamily pair, the PPII selected as a part of NRDB was the one with maximum number of ACPs among all the PPIIs sharing the same SCOP superfamily pair. Therefore, it was obvious for the larger PPIPs to possess relatively higher cumulative hydrophobicity, solvation energy, and other parametric scores. But, as the planarity index of PPIPs is not much dependent on the size of interacting patches (i.e. PPIPs), the cumulative parametric score obtained for planarity for each PPII class, confined to a common sub-range.

The size of PPIPs varied for each of these classes (MSS, LSS, and HSS) and it was least for MSS and highest for the HSS. The maximal tail size of the interacting patch of LSS was almost 1300 Å^2

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 851-860 (2021)
©Biomedical Informatics (2021)

less than HSS and 1300 Å^2 more than MSS. The effect of size can be seen for the rest of the parameters except for planarity index. Planarity score was an average measure of RMSD of PPIP constituent atoms from the best-fit plane, hence this value was almost size independent. The hydrophobicity showed the largest variation, with range for values corresponding to HSS (-1 – 93) was almost 3 times that of LSS (-1 – 36) and MSS (0 – 31). The parametric range for cumulative depth and protrusion index for HSS (Depth: 0 – 28 and Protrusion: 0 – 150) was almost twice that of the other two classes. The upper and lower meniscus for solvation free energy of HSS (-29.0 kJ·mol^{-1}– 32 kJ·mol^{-1}) was very broad as that of LSS (-6.0 kJ·mol^{-1} – 21 kJ·mol^{-1}) and MSS (-5.0 kJ·mol^{-1}– 23 kJ·mol^{-1}). In Supplementary Figures 8-13, the trend of parametric values obtained for around 99% of the PPIPs from each of the three PPII classes of both PPInS and NRDB datasets with respect to six PPI parameters are shown.

Analysis of PPIPs with respect to the hydrophobicity:

The hydrophobic analysis of PPIPs from PPInS and NRDB revealed that even though the range of hydrophobicity values of PPIPs differed among HSS, MSS, and LSS (Figure 2), the distribution of the parametric scores followed almost the same pattern. The statistical analysis of parametric score distribution with respect to mean, standard deviation, and p-values (Table 5) support the hypothesis that the cumulative hydrophobic content of PPIP from all the PPII classes is significantly same. And, this was applicable for PPIIs from both the datasets. The hydrophobicity score for the PPInS is multi-peak, whereas the same obtained from NRDB is single-peak. This apparent difference is perhaps due to the high order of redundancy in PDB for PPI complexes with the relatively larger contribution of hydrophobic effect in the binding energy. As the hydrophobicity is a major contributor to the formation of protein crystals [47, 50], thus, the majority of the PPInS have a very high hydrophobicity component in the interacting energetics.

This redundancy is removed in the NRDB and as a consequence of which the multi-peak distribution converges to a single-peak. Clearly, the hydrophobicity values for PPIPs can vary from as low as 100 to as high as 0 (NRDB part of Figure 3) (See methods section). However, the majority of the PPIPs have the hydrophobic score of 5-8 irrespective of homo/heterodimeric nature of complexes. The analysis made by [49-50] concluded that the hydrophobic effect is predominated contributor in the formation of obligate complexes. The trends of the cumulative hydrophobic index of the PPIPs as seen in Figure 2, clearly shows that the proportionate-contribution of hydrophobic effect (on an average) is independent of the extent of sequence similarity between the interacting interfaces.

Analysis of PPIPs with respect to solvation free energy (in KJ/Mol)

The range of solvation free energy values of PPIP does not differ amongst HSS, MSS, and LSS (Figure 3) as the graphs follow an identical pattern with statistical similarity among the parametric scores of different PPII classes (Table 5). This was true for NRDB too. This similarity is perhaps due to the uniform nature of PPI, which is also one major reason that solvation-energy based prediction tools for PPI and drugability studies are more successful [51-52]. Interestingly the nature of complexes – whether homodimeric or heterodimeric doesn't influence the peak of solvation free energies in NRDB or PPInS.

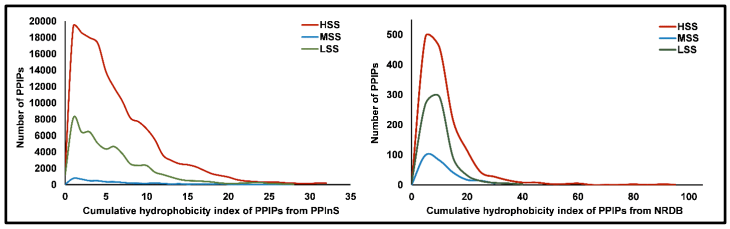

Figure 2: Hydrophobicity index of PPIPs

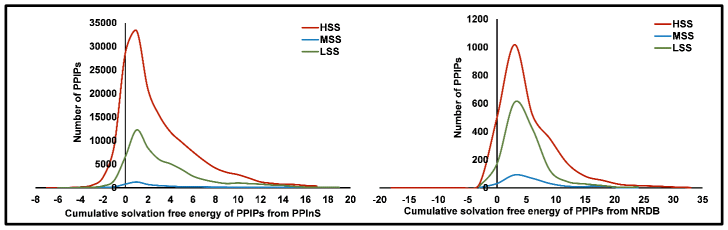

Figure 3: Solvation free energy (in KJ/mol) score of PPIPs

Analysis of PPIPs with respect to the depth index

The distribution range of the depth index of PPIP also does not differ amongst the three classes of PPInS and NRDB both, as seen in Figure 4 where the graphs follow an identical pattern. The depth index of the PPInS is the same as obtained for the PPIPs from NRDB with single-peak with similar shoulders. The statistical analysis of cumulative depth distribution for each PPII class of the both of these datasets was also significantly same (Table 5). The very low value of cumulative depth index shows that the atoms involved in PPIPs are on the surface. This may be due to the effect of interacting surface induction. The majority of PPIPs have identical cumulative depth index and this fact is not influenced by the extent of redundancy in the datasets. Irrespective of the structural/functional class of the protein, the cumulative depth index of the PPIP remains more or less constant. This is another indication that irrespective of the homo/heterodimers, the physicochemical and structural parameters governing PPI occupy the same value-space.

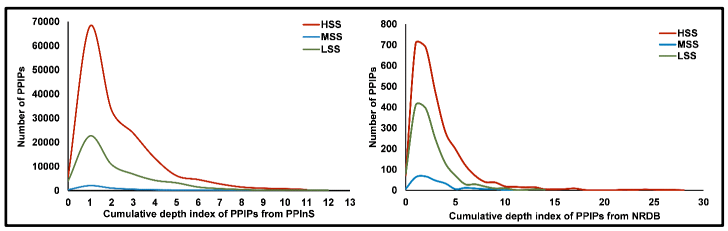

Figure 4: Depth index of PPIPs

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 851-860 (2021)
©Biomedical Informatics (2021)

Analysis of PPIPs with respect to the size of interacting patch (in Å²)

The size of interacting patches has a direct relationship with the van der Waals (vdW) forces of interactions. Greater the size of PPIP, larger is the vdW interaction energy of PPIPs. In earlier studies [18, 53], a limited number of PPCs were studied and it was observed that on an average PPI had a size of 800 ± 400 Å². This is similar to earlier reportswhich reported the size of the interfaces as small as ~800 Å²[19]. In another study [53] the size of interfaces was found in the range of 415 to 3568 Å² for heterodimers, 550 to 4718 Å² for homodimers, and 423 to 2361 Å² for transient complexes. In our study too, the interfaces from PPInS dataset are reported with size upto 6500 Å² (Figure 5-Left) while in NRDB interaction interfaces are reported to be between 112 Å² to 8400 Å² (Figure 5-Right)per interacting partner. However, most of the PPIIs from PPInS and NRDB were seen covering protein surface up to within 800 ± 400 Å² and 1200 Å² ± 400 Å²,respectively in conformity with earlier studies [21, 22]. In Figure 6, the interacting region of protein chains with smallest as well as largest PPIP size from NRDB and PPInS is shown. For this parameter too, the statistical analysis carried out to analyze the distribution of PPIP size, considering all PPII classes of both the datasets, showed the similarity in terms of the PPIP size (Table 5).

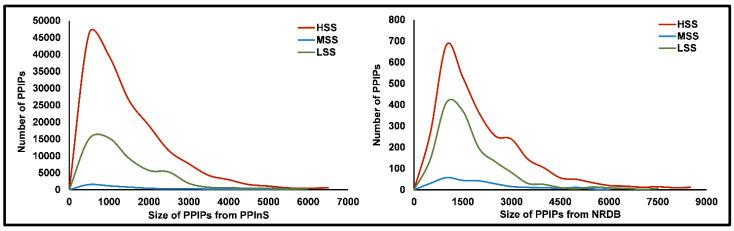

Figure 5: Size of PPIPs

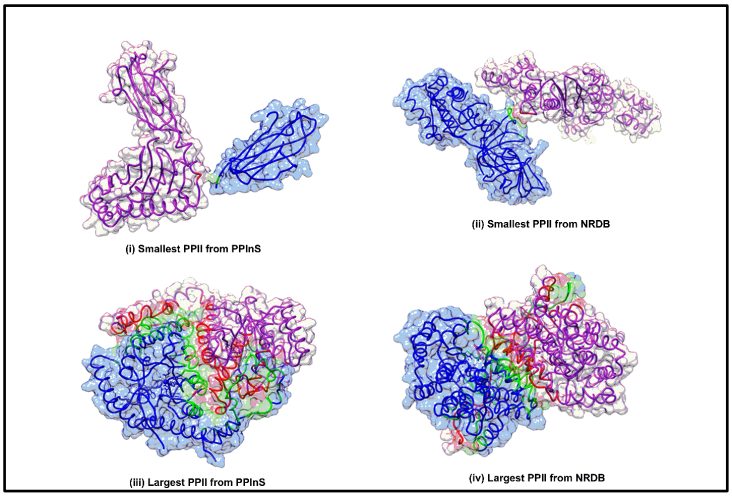

Figure 6: PPIIs with smallest and largest interacting interface from PPInS and NRDB. The interaction interface is shown in red and green. (i) 4Z78::chains A:E (21 Å² and 30 Å²; (ii) 1EFU: chains B:C (201 Å² and 165 Å²); (iii) 5AVN: chains A:B (6363 Å² and 6291 Å²); (iv) IOWC: chains A:B (8491 Å² and 8570 Å²)

Analysis of PPIPs with respect to the protrusion index

The protrusion index or the compactness of neighborhood of interacting residues has been studied by some groups [56] and it has been seen that its average value ranges from 0 to 14 [40]for protein atoms. In our datasets, we studied the cumulative protrusion index (Figure 7). Surprisingly its value (between 7-10) was very low, considering the large number of atoms that contribute to the PPII. This indicates the relatively higher packing (thus increased neighbor density) and perhaps also reduced flexibility as reported earlier [56]. This is an important parameter for prediction of hot-spots residues on PPI sites. Irrespective of sequence similarity between the interacting partners protrusion index followed an identical distribution (Table 5). Interestingly the nature of complexes – whether homodimeric or heterodimeric does not differentially influence the protrusion index (compact packaging of the PPI site).

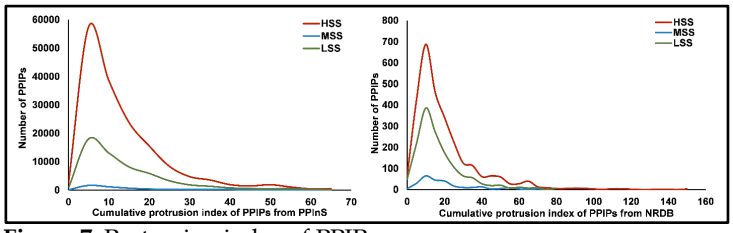

Figure 7: Protrusion index of PPIPs

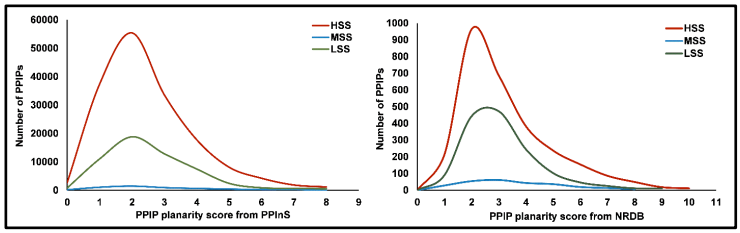

Figure 8: Planarity score of PPIPs

Analysis of PPIPs with respect to the planarity index

The planarity analysis of PPIPs revealed the clear independency of the type and size of protein in terms of their binding site preferences (Figure 8). For both the datasets, PPInS (which is repository of over two hundred thousand PPIPs) and the NRDB (having PPIPs with size up to 8400 Å²), the deviation of interacting atoms from the derived plane was found in the range of 0-8 Å. Maximum number of PPIPs were reported with deviation of 1-5 Å which is analogous to the previous findings [19, 28, 51, 53, 57]. These findings hold true for all three respective PPII classes of both the datasets. The analysis of parametric score distributions with respect to three PPII classes of the both the datasets also showed the similarity in terms of their level of flatness (Table 5).

Conclusion:

The protein-protein interaction interfaces from two datasets, covering the largest collection of experimentally determined protein-protein complexes ever, were examined in terms of their hydrophobic content, associated solvation energy, compactness of interacting residue' neighborhood, planarity and depth index.

Analysis of PPIPs from NRDB pertaining to RIP showed the presence of aliphatic and aromatic residues on interaction sites in abundance and deficit of charged residues (except Arg) as reported in previous studies. On analyzing PPIPs from PPInS, initially, a significant difference between the parametric scores (obtained with respect to each PPI site parameter) of the three PPII classes was observed. However, on removing less than 1% PPIPs (statistical outliers) from each PPII class, the cumulative parametric scores for each PPI site parameter from all three classes of PPII reduced down to the identical ranges. The analysis of PPIPs from NRDB considering each PPII class (with and without outliers) also showed the similar trends for the parametric scores for each PPI site parameter, however, with greater variability (except for the planarity). As the PPIPs in HSS class were relatively larger in size, the resulting cumulative score could not get aligned with the scores from LSS and MSS wherein they were found to converge into a common sub-range after outlier removal. However, overall, the objective analysis of PPIPs (from all three PPII classes of both the datasets) with respect to all PPI sites parameters showed the similar trends. This indicates that the principles of molecular recognition among proteins are not driven by their sequence / structural similarity and reinforces the importance of geometrical and electrostatic complementarity as the main component for PPIs.

Conflict of interest(s): None

Author contribution:
The proposed work was conceptualized by MK and executed by VK and MK. The data was analyzed by VK, AM, TG and MK. The manuscript was written by VK, AS and MK. All authors have read and approved the manuscript prior to submission.

Acknowledgments:
The authors are grateful to the University Grants Commission (UGC), India for providing financial support to VK in the form of UGC-NET JRF award. The authors are also thankful to the, Central Universities of Punjab and Himachal Pradesh, India, for providing academic, administrative, and infrastructural support to carry out this work.

References:

[1] Romero PA & Arnold FH. *Nat. Rev. Mol. Cell Biol.* 2009 10:866. [PMID: 19935669]
[2] Pechmann S *et al*. *Proc. Natl. Acad. Sci.* 2009 106:10159. [PMID : 19502422]
[3] Lulu S *et al*. *J Mol Graph Model*. 2009 28:88. [PMID: 19442545]
[4] Vaishnavi A *et al*. *Bioinformation*. 2010 20 4:310. [PMID: 20978604]
[5] Nilofer C *et al*. *Bioinformation*. 2017 13:164. [PMID: 28729757]
[6] Zhang J & Kurgan, L. *Brief. Bioinform.* 2018 19:821. [PMID: 28334258
[7] Zhang M *et al. Med. Chem.* 2017 13:506. [PMID: 28530547]
[8] Frieden E, *J. Chem. Educ.* 1975 52:754. [PMID: 172524]
[9] Cerny J & Hobza P, *Phys. Chem. Chem. Phys*. 2007 9:5291. [PMID: 17914464]
[10] Berggard T *et al. Proteomics,* 2007 7:2833. [PMID: 17640003]
[11] Young L *et al. Protein Sci.* 1994 3:717. [PMID: 8061602]
[12] Zhanhua C *et al*. *Bioinformation* 2005 11 1:28. [PMID: 17597849]
[13] Sowmya G *et al. Bioinformation* 2011 6:137. [PMID: 21572879]
[14] Karthikraja V *et al*. *Bioinformation* 2009 4:101. [PMID: 20198182]
[15] Li L *et al*. *Bioinformation* 2005 1:42. [PMID: 17597851]
[16] Barry G *et al. Current Opinion in Structural Biology* 2010 20:142. [PMID: 20060708]
[17] Cheng TMK *et al. Briefings in Functional Genomics*, 2012 11:543. [PMID: 22811516]
[18] Nilofer C *et al. J Biomol Struct Dyn.* 2020 38:3260. [PMID: 31495333]
[19] Tanaka T & Rabbitts TH, *Cell Cycle* 2008 7:1569. [PMID: 18469527]
[20] Jones S & Thornton JM, *Prog. Biophys. Molec. Biol*. 1995 63:31. [PMID: 7746868]
[21] Lo Conte L *et al. J. Mol. Biol.* 1999 285:2177. [PMID: 9925793]
[22] Bahadur RP & Zacharias M, *Cell. Mol. Life Sci.* 2008 65:1059. [PMID: 18080088]
[23] Bai F *et al. Proc. Natl. Acad. Sci.* 2016 113:E8051. [PMID: 27911825]
[24] Sudha G *et al*. *Prog. Biophys. Mol. Biol.* 2014 116:141. [PMID: 25077409]
[25] Arkin MR *et al*. *Chem. Biol*. 2014 21:1102. [PMID: 25237857]
[26] Bogan AA & Thorn KS. *J. Mol. Biol*. 1998 280:1. [PMID: 9653027]
[27] Ozdemir ES *et al. Bioinformatics.* 2018 34:i795. [PMID: 30423104]
[28] Jones S & Thornton JM. *Proc. Natl. Acad. Sci.*1996 93:13. [PMID: 8552589]
[29] Kastritis PL & Bonvin AMJJ *J. R. Soc. Interface.*2013 10:20120835. [PMID: 23235262]
[30] Mihel J *et al. BMC Struct. Biol.*2008 8:21. [PMID: 18400099]
[31] Kumar V *et al. Sci. Rep.*2018 8:12453. [PMID: 30127348]
[32] Altschul SF, *Nucleic Acids Res.* 1997 25:3389. [PMID: 9254694]
[33] Jones S & Thornton JM, *J. Mol. Biol.* 1997 272:121. [PMID: 9299342]
[34] Crowley PB & Golovin A. *Proteins Struct. Funct. Bioinforma.* 2005 59:231. [PMID: 15726638]
[35] Jones S & Thornton JM, *Prog. Biophys. Mol. Biol.* 1995 63: 31. [PMID: 7746868]
[36] Tsai CJ *et al. Protein Sci.* 1997 6:53. [PMID: 9007976]
[37] Hessa T *et al. J. Gen Physiol.* 2007 129:363. [PMID: 17438116]
[38] Wimley WC *et al. Biochemistry*. 1996 35:5109. [PMID: 8611495]
[39] Chakrabarti P & Janin J, Proteins Struct. Funct. Bioinforma. 2002 47: 334. [PMID: 11948787]
[40] Pintar A *et al. Bioinformatics.* 2002 18: 980. [PMID: 12117796]
[41] Jones S & Thornton JM, *J. Mol. Biol.* 1997 272: 133. [PMID:

9299343]
[42] Laskowski RA, *J. Mol. Graph.*1995 13:323. [PMID: 8603061]
[43] Winter KU *et al. Mol. Biol. Evol.* 2002 19:587. [PMID: 11961093]
[44] Elkayam T *et al. Proc. Natl. Acad. Sci.* U.S.A. 2003 100:5772. [PMID: 12730379]
[45] ChandlerD. *Nature.*2005 437:640. [PMID: 16193038]
[46] Fuller JC *et al. Drug Discov. Today.* 2009 14:155. [PMID: 19041415]
[47] Regnier FE *Science* 1987 238:319. [PMID: 3310233]
[48] Baldwin EP & Matthews BW, *Curr. Opin. Biotechnol.* 1994 5:396. [PMID: 7765172]
[49] Janin J *et al. J. Mol. Biol.* 1988 204:155. [PMID: 3216390]
[50] Bahadur RP *et al. Proteins Struct. Funct. Bioinforma.*2003 53:708. [PMID: 14579361]
[51] Kulharia M *et al.J. Mol. Graph. Model.*2009 28:297. [PMID: 19762259]
[52] Cheng AC *Nat. Biotechnol.*2007 25:71. [PMID: 17211405]
[53] Chakrabarti P & Janin J, *Proteins Struct. Funct. Genet.*2002 47:334. [PMID: 11948787]
[54] Bahadur RP, *J. Mol. Biol.* 2004 336:943. [PMID: 15095871]
[55] Caffrey DR *et al. Protein Sci.* 2004 13:190. [PMID: 14691234]
[56] Gao M. *Proc. Natl. Acad. Sci. U.S.A.*, 2010 107:22517. [PMID: 21149688]
[57] Chen J *et al. Protein Sci.* 2013 22:510. [PMID: 23389845]

Edited by P Kangueane
Citation: Kumar ***et al.*** Bioinformation 17(10): 851-860 (2021)
License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 17(10): 851-860 (2021)
©Biomedical Informatics (2021)