BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 6(4)

Insights from the structural analysis of protein heterodimer interfaces

Gopichandran Sowmya[1, 2]*, Sathyanarayanan Anita[2], Pandjassarame Kangueane[1, 2]

[1]Department of Biotechnology, Faculty of applied Science, AIMST University, 08100 Semeling, Malaysia; [2]Biomedical Informatics, Pondicherry 607402, India; E-mail: sowmyagopichandran@gmail.com; Phone: +914132633589; Fax: +914132633722; *Corresponding author

Received April 15, 2011; Accepted May 07, 2011; Published May 07, 2011

Abstract:
Protein heterodimer complexes are often involved in catalysis, regulation, assembly, immunity and inhibition. This involves the formation of stable interfaces between the interacting partners. Hence, it is of interest to describe heterodimer interfaces using known structural complexes. We use a non-redundant dataset of 192 heterodimer complex structures from the protein databank (PDB) to identify interface residues and describe their interfaces using amino-acids residue property preference. Analysis of the dataset shows that the heterodimer interfaces are often abundant in polar residues. The analysis also shows the presence of two classes of interfaces in heterodimer complexes. The first class of interfaces (class A) with more polar residues than core but less than surface is known. These interfaces are more hydrophobic than surfaces, where protein-protein binding is largely hydrophobic. The second class of interfaces (class B) with more polar residues than core and surface is shown. These interfaces are more polar than surfaces, where binding is mainly polar. Thus, these findings provide insights to the understanding of protein-protein interactions.

Keywords: protein-protein interaction (PPI); heterodimer; interface; surface; core; polar abundance.

Background:
The formation of protein complexes by two different proteins (heterodimers) involves a stable interface. The driving force deterministic of their interface features (chemical and physical) is essential for its molecular function. However, our current knowledge on the molecular principles of protein-protein binding is limited. Hence, the identification of a binding partner from sequence alone is still a great challenge. Therefore, it is of importance to document interface residue types in heterodimers using an updated yet non-redundant dataset of structures determined by X-ray crystallography. The description of interfaces using amino acid residues and their types help understand protein-protein interaction (PPI). The principles of PPI gleaned from the analysis of protein complexes determined by X-ray crystallography have been documented in the literature **[1-19]**. PPI was described using various structural (e.g. interface area, interface size, gap index, volume, planarity, hydrogen bonds, etc.) and sequence (e.g. protein size, residue type, residue frequency, conserved interface patterns, etc.) property parameters in these studies. These studies provide mean statistics on interface features for large datasets. This provided valuable insights to the understanding of protein-protein interactions. However, protein-protein interaction is specific and every interface is unique. Hence, it is important to classify known protein complexes based on interfaces.

The classical work by Chothia & Janin (1975) showed that protein interfaces are dominantly hydrophobic **[1]**. It was later detailed by Jones & Thornton (1995) that interfaces have more hydrophobic residues than surface but less than core **[2]**. The role of interface hydrophobic residues in binding was also later acknowledged by Tsai *et al.* (1997) **[3]**. It was found that large and strong hydrophobic patches are dominating features at the interface **[4]**. The use of a hydrophobic mean-field potential for protein subunit docking was also subsequently demonstrated **[5]**. Hydrophobic interfaces with few charged groups have been described **[6]**. This study also documented that interface residues are either "abundantly polar" or "abundantly hydrophobic". The presence of distinctly clustered yet conserved residues at the interface was known **[7]**. Interfaces have also been described using features (e.g. protein size, interface size, interface area, gap volume, gap index, planarity, hydrogen bonds, salt bridges, residue propensity, etc.) based on mean statistics for large datasets **[8, 9, 10, 11, 12, 13]**. Online web servers are also available for studying PPI using these features **[14, 15, 16]**. Thus, the progress on the understanding of the molecular principles of protein-protein binding is prominent. It should be stated that these studies use datasets consisting of both heterodimers and homodimers. The formation of homodimers and their folding through 2-state (2S - without intermediate) and 3-state (3S - with stable intermediate) mechanisms is distinct from that of heterodimers **[20]**. Therefore, it is our interest here, to study and understand heterodimer complexes only, using interface residue types. Moreover, it is known that non-specific interfaces are less pronounced in heterodimer complexes and hence, the need to distinguish true and false complexes is not compelling **[9]**. We use percentage polar residues to describe interface in comparison with core and surface for 209 heterodimer complexes to classify them into distinct classes.

Materials & Methodology:
Heterodimers dataset:
We created an updated yet non-redundant heterodimer dataset from protein databank (PDB) **[21]**. The availability of precompiled datasets are described in ProtorP **[16]** and PQS **[22]** online servers. ProtorP provides no option for download and PQS has not been updated since 1999. Therefore, it is essential to create an updated yet non-redundant heterodimer dataset from PDB **(Table 1 see Supplementary material)** using the procedure outlined in **Figure 1**. In this procedure, we downloaded 5,387 entries from PDBelite web interface using the predefined keywords "hetero AND dimer" **[23]**. However, this dataset was redundant corresponding to about 28,525 sequence chains. This is more than

the expected 10,774 (5,387*2) due to the presence of multiple sequence chains (>2 chains) in several entries. Therefore, we extracted the PDB entries (984) with just two sequence chains. Thus, a sequence set of 1,968 sequences corresponding to 984 PDB entries was created. This dataset was redundant at sequence level and hence, the dataset was subjected to CD-HIT (sequence redundancy removal program) **[24]** at 40% sequence similarity cut-off (with step size n = 2). This resulted in 680 unique sequences corresponding to 457 PDB entries. It should be noted that the number of complexes is more than half of the number of chains. This is because the interface is a combination of two chains and thus, the interfaces are non-redundant. This set contained about 60 RNA/DNA, homodimer and HETATOM structures and these entries were removed. The 397 protein complexes produced were further refined to remove short peptides of chain length <=50 residues and resolution > 3.5 Å. This resulted in a non-redundant dataset of 192 heterodimer protein complexes **(Table 1)**. The dataset was subsequently characterized for protein size distribution **(Figure 2).**

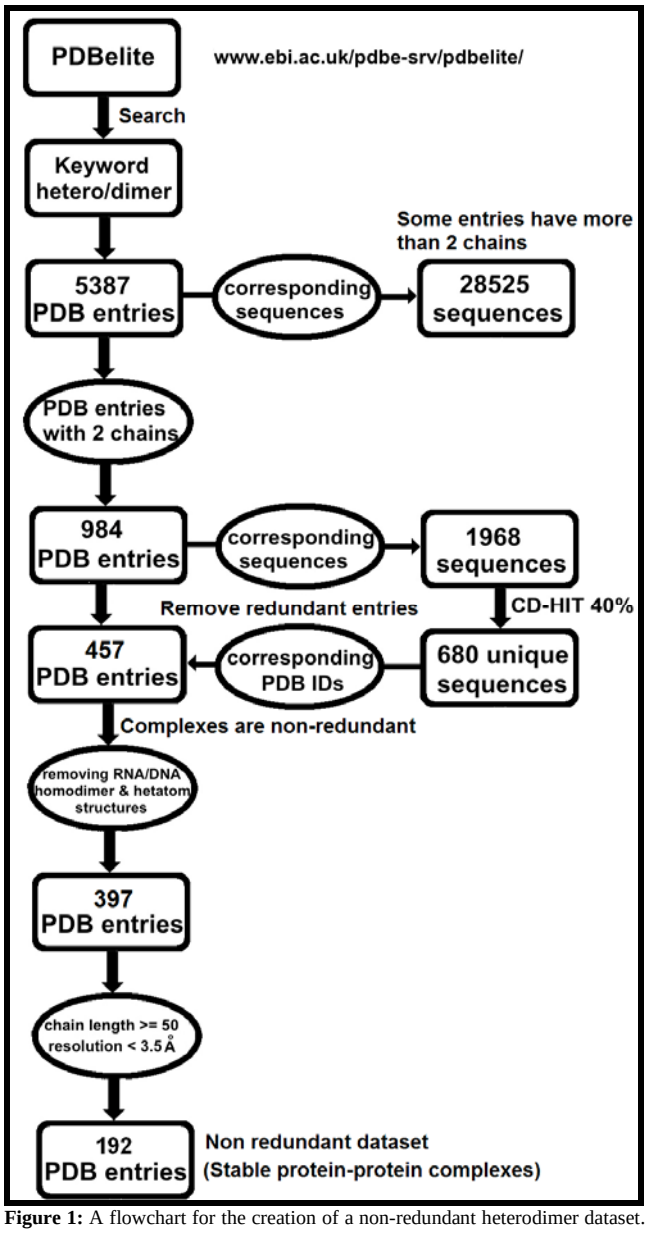

Figure 1: A flowchart for the creation of a non-redundant heterodimer dataset. PDB = Protein databank.

Source organism based grouping:
Each heterodimer complex is made up of two protein monomer subunits. The source for each protein subunit is either different (different organism (DO)) or same (same organism (SO)) **(Table 1)**. The formation of a protein complex with interacting partners from DO is possible, often for a non-essential (non-obligatory, e.g. inhibitory) role, only in heterodimers. Thus, the dataset is divided based on organism source of interacting partners. The dataset also consists of 5 (FIVE) complexes with at least one synthetic partner (SP).

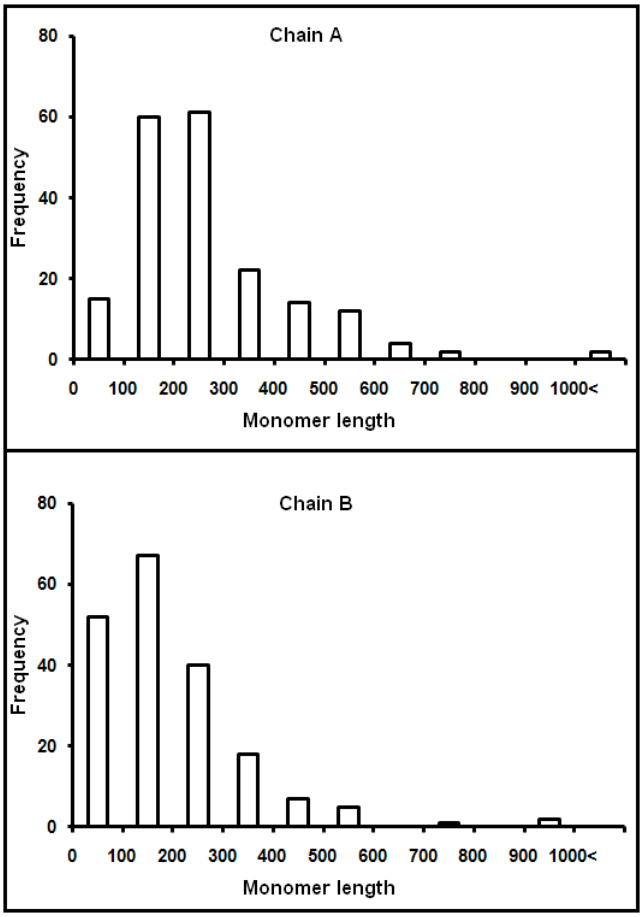

Figure 2: Characterization of the dataset based on protein size.

Functional grouping of complexes:
We extracted "descriptive" functional data (usually semantic) for each complex from the PDB header annotation records. This data was manually curated ("by domain expert decision") through visual inspection using available literature information. Thus, complexes were generally grouped based on function into catalysis (enzymes), regulatory (cellular), assembly (structural), immunity and inhibitory **(Table 1)**. It should be noted that this exercise is not comprehensive. However, we have taken reasonable effort on a case by case basis to classify complexes into their respective functional groups. Manual inspection of PDB description records suggests that DO complexes are often inhibitory (e.g. PDB code: 1K9O) or immune (e.g. PDB code: 1GH6) related **(Table 3 see Supplementary material).** However, SO complexes are associated with catalysis, regulatory, assembly and immunity. The SP group consists of a synthetic partner for *in vitro* inhibitory or regulatory studies. It is often possible that a complex may align with two different functional groups, where such complexes are grouped based an "expert decision" using known information.

Accessible surface area (ASA):
ASA was calculated using the WINDOWS software Surface Racer **[25]** with Lee and Richard (1971) **[26]** implementation. A probe radius of 1.4 Å was used for ASA calculation.

Interface residues:
Interface (I) residues in heterodimers are identified using change in accessible surface area (ΔASA) from a "monomer-state" to a "dimer-state". Residues with ΔASA > 0 Å are considered to be at the interface. Thus, interface residues contributed by subunits A and B were identified.

Interface size and Interface area:
The distribution of complexes with interface size (number of interface residues) is given in **Figure 3.** The relationship between interface size and interface area is given in **Figure 4.**

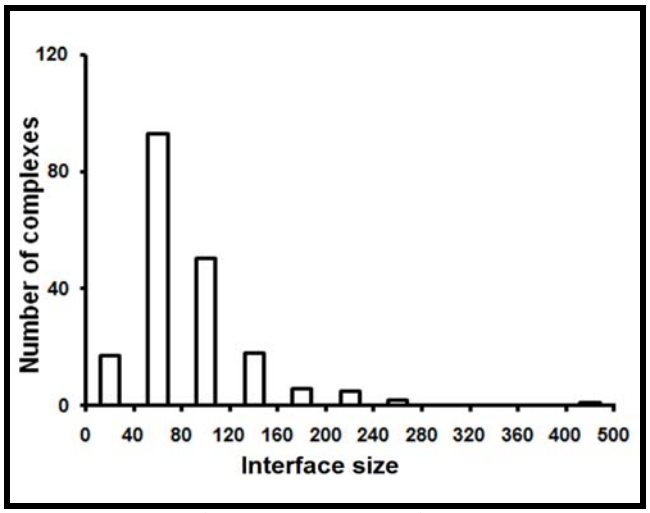

Figure 3: Distribution of complexes based on interface size.

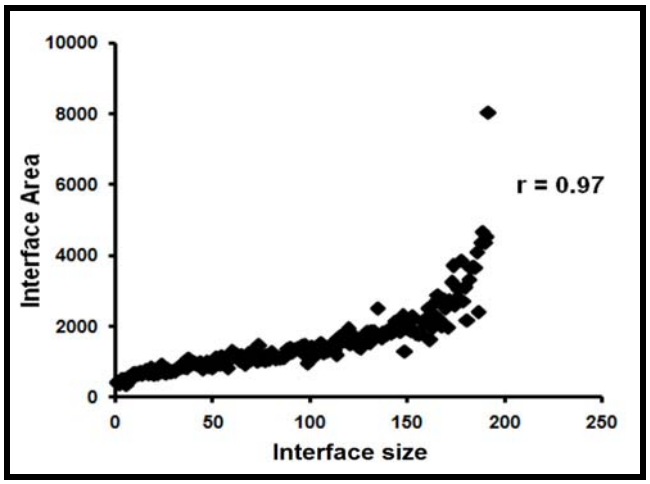

Figure 4: Relationship between interface size and interface area among complexes.

Interface property abundance:
The interface between two interacting subunits is made of both polar and hydrophobic residues. The number of polar and hydrophobic residues at the interface varies from complex to complex. Some interfaces are rich in polar residues, while some others are rich in hydrophobic residues. Therefore, we calculated the percentage of polar and hydrophobic residues at the interface for each complex. The difference in the percentages of polar (P) and hydrophobic (H) residues at the interface is measured **(Figure 5).** Thus, interface residues have "polar abundance" when %P - %H > 0 and "hydrophobic abundance" when it is < 0**.** This help to classify complexes with interfaces based on "abundant polar" and "abundant hydrophobic" residues.

Surface residues:
Surface (S) residues in heterodimers are identified using residue ASA values in a "dimer state". Residues with ASA > 0 Å are considered as surface residues. Thus, surface residues in the subunits A and B of the complex were identified.

Core residues:
Core (C) residues in heterodimers are identified using residue ASA values in a "monomer state". Residues with ASA = 0 Å are considered as core residues. Thus, core residues in the subunits A and B were identified.

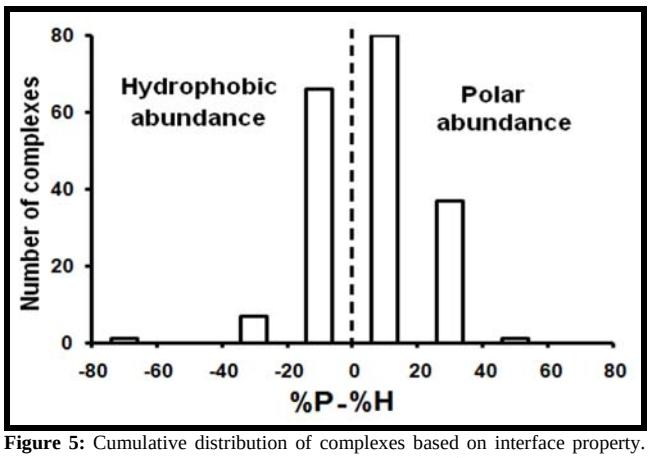

Figure 5: Cumulative distribution of complexes based on interface property. Complexes distributed in the positive X-axis have interfaces with polar residue abundance and those distributed in the negative X-axis have interfaces with hydrophobic residue abundance.

Interface, surface and core polarity:
A protein heterodimer complex consists of three distinct regions (core (C), interface (I) and surface (S)) as shown in **Figure 6**. Interface, surface, core residues in a complex thus documented are further classified into polar and hydrophobic residues. Thus, interface, surface and core residues are grouped as polar {R, N, D, Q, H, K, S, T, Y, E} and hydrophobic {A, C, G, I, L, M, F, P, V, W} based on residue type. We then estimated the percentage of polar residues at interface (I), surface (S) and core (C) for each complex.

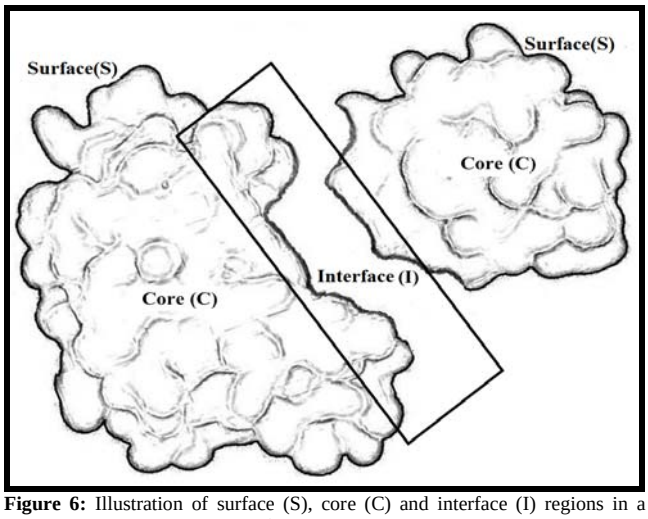

Figure 6: Illustration of surface (S), core (C) and interface (I) regions in a heterodimer complex. The interface is the interacting region between the two protein partners. The core is the buried region in the individual monomers. The surface is the solvent exposed region in the complex state.

Classification of complexes:
Complexes were grouped into four distinct classes based on the relative difference in percentage polar residues (referred thereafter as polarity) between interface and core **(Figure 7; Table 2 see Supplementary material)**. Complexes with interface polarity greater than core but less than surface, such that [S>I>C] are "class A"**.** Complexes with interface polarity greater than core and surface, such that [S<I>C] are "class B". Complexes with interface polarity less than core and surface, such that [S>I<C] are "class C". It should be stated that "class D" are such that [S<I<C].

Statistical analysis:
The statistical significance analysis was calculated using the GraphPad Prism (version 5) software [http://www.graphpad.com/]. The F test for variance comparison was used for calculating the significance of functional preference between DO and SO group of complexes.

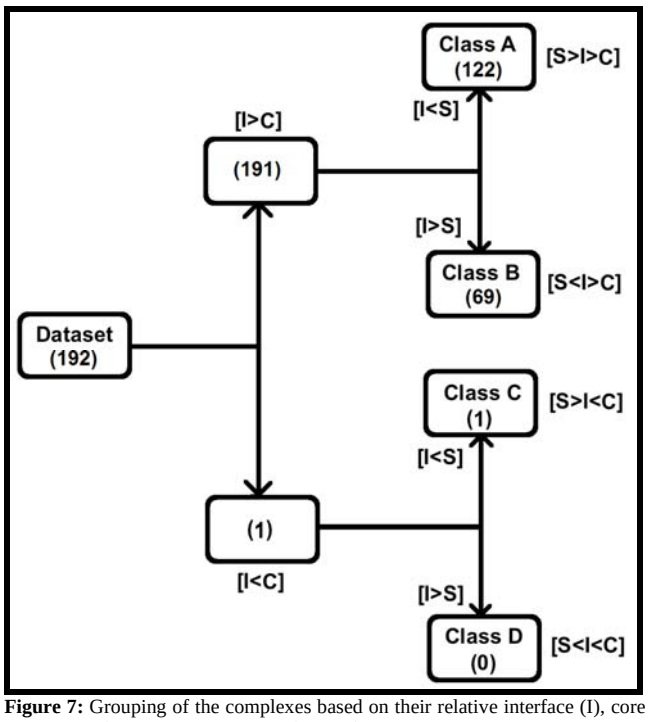

Figure 7: Grouping of the complexes based on their relative interface (I), core (C) and surface (S) polarity. Interfaces often have more polar residues than core in [I>C] groups. The hierarchical grouping shows the abundance of class A [S>I>C] and class B [S<I>C] complexes in the dataset. Class C [S>I<C] is rare and class D [S<I<C] is absent in the dataset.

Results:

The principles of PPI were studied using a dataset of 192 heterodimer complexes **(Table 1)** created using a procedure described in **Figure 1**. The dataset is divided based on the organism source of the interacting partners. Thus, SO, DO, and SP group of complexes were identified **(Table 1)**. The distribution of complexes based on interacting protein size is given in **Figure 2.** This describes the size of interacting protein partners forming the complex. These partners interact through interface residues. The distribution of interface size among heterodimer complexes is given in **Figure 3.** The interfaces have interface areas which correlate with interface size **(Figure 4).** The chemical nature of interface residues in complexes is given in **Figure 5.** This shows that interface residues in complexes are either "abundantly polar" or "abundantly hydrophobic". However, majority of interfaces (121/192 – 63%) have abundantly polar residues.

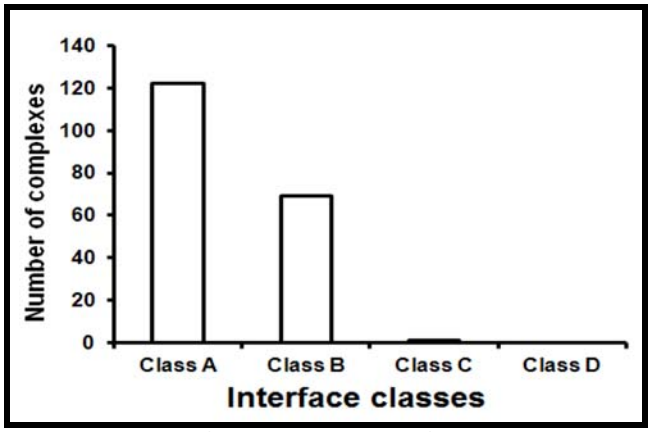

Figure 8: Distribution of complexes based on interface class. The distribution shows that 64% of complexes have "class A" interface and 36% of complexes have "class B" interface.

The classification of complexes using relative polarity between interface, core and surface into classes A-D was shown **(Table 2; Figure 8)**. This grouping shows that majority (191/192 - 99%) of interfaces have polarity greater than core [I>C] as shown in **Figure 7.** However, interfaces in two complexes (1/192 – <1%) have polarity less than core [I<C]. We further found that 64% (122/192) of complexes are grouped under "class A" having interface polarity greater than core but less than surface**.** It was also noted that 36% (69/192) of complexes are "class B" with interface polarity greater than core and surface. Complexes having interface polarity less than core and surface (class C) are rare (1/192 – <1%) in the dataset. It should be stated that "class D" type of complexes are absent in the dataset. Grouping of complexes based on source organism of interacting partners shows that DO complexes are mostly inhibitory and SO are usually associated with catalysis, regulation and assembly **(Table 1; Table 3)**. Thus, DO and SO group of complexes show functional preference (p = 0.019). However, this is not true for classes (A–D) as shown in **Table 4** (p = 0.12). **Table 2** shows that complexes grouped in classes A, B, C and D does not show significant difference for function preference.

Discussion:

Protein-protein interactions are vital for cellular function. Two different proteins associate with one another for function (catalysis, regulatory and assembly) that are often obligatory (essential for cellular activity). However, this is not always true. They also interact for inhibitory and immune related role, where their association is frequently non-obligatory (not essential for cellular activity). The dataset shows that obligatory role is usually observed among SO complexes and non-obligatory functions are common among DO complexes. Thus, the functional role exhibited by complexes based on organism source is significantly distinct (p value = 0.019). However, the molecular principles for such associations are not clearly known. The molecular forces for protein interactions are gathered through analysis of known structural complexes. Hence, we describe the analysis of a dataset of 192 heterodimer complexes using polarity of the interface, surface and core for classifying them into classes A - D.

Analysis of protein structural complexes showed that interfaces are either "dominantly polar" **[6]** or "dominantly hydrophobic" **[1, 2, 6]**. It is also known that the interface hydrophobic residues are more than surface but less than core **[2]**. Hydrophobic interfaces are similar to surface with few charged groups **[6]**. Our analysis shows that class A complexes have interface polarity greater than core but less than surface as reported elsewhere **[2]**. Thus, this observation is acknowledged in this study using an extended dataset. Interfaces are part of the surfaces in the monomers, where the interface hydrophobic residues are more than the rest of the surface and the partners interact through relative hydrophobic forces. It should be noted that we identified an unusual complex (PDB code: 2F95) under class C describing rhodopsin II/transducer interaction. The core is made of more polar residues than the interface in this complex. Thus, protein binding is hydrophobic, although, folding of the individual monomers are driven by polar residues, as in several non-globular proteins. We also identified class B complexes with interface polarity greater than both core and surface. In this class, interface polar residues are more than the rest of the surface and partners interact through polar interactions. Thus, relative polarity is the driving force in class B complexes. This class of interfaces has not been described in the literature and it is novel. The driving force for protein binding is hydrophobic in class A and polar in class B complexes. These observations using interface residue properties are imminent to the understanding of protein binding in heterodimer complexes. This study should be extended using a combined formulation of residue types and atomic features in future investigation. It should also be noted that interfaces between partners are part of surfaces in interacting monomers. These interfaces are clearly defined in known structural complexes. However, there are often several binding sites in an interacting monomer under *in vivo* conditions and these have not yet been characterized. Therefore, experiments should be formulated to capture these combined features in future studies.

Conclusion:

Proteins associate with one another as a resultant effect of both polar and hydrophobic residues at the interface. The unresolved challenge here is to quantify their combined effect at the interface. Inter-subunit scoring functions for polar and hydrophobic effects are available based on a limited set of structural complexes and are always inadequate to describe new classes of interfaces. It is known that interface residues are either "abundantly polar" or "abundantly hydrophobic". It is also known that interfaces are less hydrophobic than core but more than surface in a class of complexes. We document a new class of complexes with more interface residues than core and surface. Thus, the driving force for protein-protein interaction is selectively either hydrophobic or polar for different classes of interfaces.

References:
[1] Chothia C & Janin J. *Nature.* 1975 **256**: 705 [PMID: 1153006]
[2] Jones S & Thronton JM. *Prog Biophys Mol Biol.* 1995 **63**: 31 [PMID: 7746868]
[3] Tsai CJ *et al. Protein Sci.* 1997 **6**: 53 [PMID: 9007976]
[4] Lijnzaad P & Argos P. *Proteins.* 1997 **28**: 333 [PMID: 9223180]
[5] Robert CH & Janin J. *J Mol Biol.* 1998 **283**: 1037 [PMID: 9799642]
[6] Lo Conte L *et al. J Mol Biol.* 1999 **285**: 2177 [PMID: 9925793]
[7] Guharoy M & Chakrabarti P. *BMC Bioinformatics.* 2010 **11**: 286 [PMID: 20507585]
[8] Chakrabarti P & Janin J. *Proteins.* 2002 **47**: 334 [PMID: 11948787]
[9] Bahadur RP *et al. J Mol Biol.* 2004 **336**: 943 [PMID: 15095871]
[10] Zhanhua C *et al. Bioinformation.* 2005 **1**: 28 [PMID: 17597849]
[11] Pal A *et al. J Biosci.* 2007 **32**: 101 [PMID: 17426384]
[12] Guharoy M & Chakrabarti P. *Bioinformatics.* 2007 **23**: 1909 [PMID: 17510165]
[13] Vaishnavi A *et al. Bioinformation.* 2010 **4**: 310 [PMID: 20978604]
[14] Murakami Y & Jones S. *Bioinformatics.* 2006 **22**: 1794 [PMID: 16672257]
[15] Saha RP *et al. BMC Struct Biol.* 2006 **6**: 11 [PMID: 16759379]
[16] Reynolds C *et al. Bioinformatics.* 2009 **25**: 413 [PMID: 19001476]
[17] Chothia C *et al. Proc Natl Acad Sci U S A.* 1976 **73**: 3793 [PMID: 1069263]
[18] Miller S *et al. Nature.* 1987 **328**: 834 [PMID: 3627230]
[19] Janin J & Chothia C. *J Biol Chem.* 1990 **265**: 16027 [PMID: 2204619]
[20] Lulu S *et al. J Mol Graph Model.* 2009 **28**: 88 [PMID: 19442545]
[21] Berman HM *et al. Nucleic Acids Res.* 2000 **28**: 235 [PMID: 10592235]
[22] Henrick K & Thronton JM. *Trends Biochem Sci.* 1998 **23**: 358 [PMID: 9787643]
[23] Velankar S *et al. Nucleic Acids Res.* 2011 **39**: D402 [PMID: 21045060]
[24] Li W & Godzik A. *Bioinformatics.* 2006 **22**: 1658 [PMID: 16731699]
[25] Tsodikov OV *et al. J Comput Chem.* 2002 **23**: 600 [PMID: 11939594]
[26] Lee B & Richards FM. *J Mol Biol.* 1971 **55**: 379 [PMID: 5551392]

Edited by VS Mathura
Citation: Sowmya *et al.* Bioinformation 6(4): 137-143(2011)

BIOINFORMATION

open access

Supplementary material:

Table 1: Heterodimer dataset (192)

	R			E			I	SA	EI	Un
	1A6D	1PDK	2OVP	1A50	1LDJ	2F9Y	1BKD	1B8M	1BI7	1J7D
	1B34	1R1K	2P7V	1B7Y	1NVI	2GAF	1H3P	1CAU	1OC0	
	1BP3	1RKE	2PJW	1BMQ	1PG5	2GGV	1HDM	1P5V	1XT9	
	1C3A	1RY7	2PM9	1BPL	1QDL	2HLE	1MCW	1XOU	1ZJD	
	1CE7	1SC5	2PQN	1BUH	1QS0	2JGZ	1NY7	2F95	2HSN	
	1F60	1XB2	2QSF	1DF0	1SPP	2NXN	1P4I	2VT1	2IO0	
	1G4Y	1XEW	2RAW	1DTW	1SVD	2O8V	1PGW		2RD0	
	1GGP	1YHN	2W99	1EUV	1V29	2Q5W	1SB2		2YSU	
	1H59	1YKH	2Z5B	1FS0	1W98	2QVS	1V1P		2Z6B	
SO (144)	1IAR	2B3T	2ZNL	1FX0	1XRS	2RF2	2AAB			
	1JEQ	2BW3	2ZSH	1GHD	1Y8N	2VH6	2EYR			
	1JKG	2C0J	3BX7	1GLC	1Y8X	3D6N	2FHZ			
	1KTZ	2C0L	3BZV	1I1Q	1YGC	3EQQ	2H7Z			
	1LEM	2CDF	3C98	1I72	2ATC	3EXI	2HLA			
	1LZW	2CO6	3CQC	1JK0	2AYO	3FN1	2HR0			
	1M2V	2D74	3DXR	1JQL	2BKR		2P24			
	1M4U	2DYO	3F1P	1JWB	2BUQ		2Z35			
	1N1J	2FEP	3GB8	1KA9	2CYZ		2ZS6			
	1NW9	2OT3		1L5H	2DFX		3EYJ			

	RI		EI				I	SI
	1CI6	3D7T	1D4X	1ML0	1SKO	3F62	1GH6	1KU6
DO (43)	1NT2	3EVS	1EJA	1OPH	1SLU	3BES	1U58	1STF
	1UDI		1FLE	1ORY	2ARP	3CBK	2FH5	2VRW
	1XDT		1I1R	1PQZ	2E31	3DDC	2Q97	2ZIU
	2B42		1I4E	1QAV	2G2U		3FXE	
	2J12		1JOW	1QZ7	2OUL			
	2Z0D		1K9O	1CGI	2OZA			

SP (5)	1ACB	1BRB	1TTW	2SIC	3SGB

SO – Same Organism; DO – Different Organism; SP – Synthetic Partner; R – Regulatory complexes; E – Enzyme complexes; I – Immune; SA – Structural Assembly; EI – Enzyme-inhibitor; RI – Regulatory-Inhibitor; SI – Structure-Inhibitor; Un – Unknown function. Numbers in parenthesis represent the number of protein complexes.

Table 2: Grouping of complexes into distinct classes

		R			E		I	EI	SA
		1B34	1YHN	2W99	1A50	1YGC	1H3P	1OC0	1B8M
		1C34	1YKH	2ZNL	1B7Y	2BKR	1HDM	1XT9	1CAU
		1CE7	2BW3	2ZSH	1BMQ	2BUQ	1NY7	1ZJD	1P5V
		1F60	2C0J	3BX7	1BPL	2CYZ	1P4I	2YSU	1XOU
		1G4Y	2CDF	3BZV	1DF0	2F9Y	1PGW	2Z6B	2VT1
		1GGP	2CO6	3CQC	1DTW	2GAF	1V1P		
		1H59	2D74	3DXR	1FS0	2HLE	2EYR		
		1JEQ	2DYO		1GHD	2O8V	2HR0		
Class A (122)	**SO (92)**	1JKG	2FEP		1I1Q	2Q5W	2P24		
		1LEM	2OT3		1JK0	2QVS	2Z35		
		1M2V	2OVP		1JWB	2RF2	2ZS6		
		1M4U	2PJW		1L5H	2VH6	3EYJ		
		1N1J	2PM9		1LDJ	3D6N			
		1NW9	2PQN		1NVI	3EXI			
		1SCS	2QSF		1QDL	3FN1			
		1XEW	2RAW		1V29				

		RI		EI			I	SI	Un
		1CI6		1D4X	1OPH	2OUL	1I1R	2FH5	3FXE
		1NT2		1CGI	1SLU	2Z0D	1ML0		
	DO (26)	1ORY		1EJA	1STF	2ZIU	3BES		
		1SKO		1FLE	1UDI	3CBK	3F62		
		3EVS		1K9O	2E31	3DDC			
	SP(4)	1ACB	1BRB	1TTW	3SGB				

		R			E		I	EI	Un
		1A6D	1XB2	1BUH	1SPP	2GGV	1BKD	1BI7	1J7D
		1BP3	2B3T	1EUV	1SVD	2JGZ	1MCW	2HSN	
		1IAR	2C0L	1FX0	1W98	2NXN	1SB2	2IO0	
		1KTZ	2P7V	1GLC	1XRS	3EQQ	2AAB	2RD0	
Class B (69)	**SO (51)**	1LZW	2Z5B	1I72	1Y8N		2FHZ		
		1PDK	3C98	1JQL	1Y8X		2H7Z		
		1R1K	3F1P	1KA9	2ATC		2HLA		
		1RKE	3GB8	1PG5	2AYO				
		1RY7		1QS0	2DFX				

	RI		**EI**			**I**	**SI**
DO	1XDT	3D7T	1I4E	1QAV	2G2U	1GH6	1KU6
(17)	2B42		1JOW	1QZ7	2OZA	1U58	2VRW
	2J12		1PQZ	2ARP		2Q97	
SP(1)	2SIC						

Class C (1)	**SA**
SO (1)	2F95

SO – Same Organism; DO – Different Organism; SYPC – Synthetic Partner; R – Regulatory complexes; E – Enzyme complexes; I – Immune; SA – Structural Assembly; EI – Enzyme-inhibitor; RI – Regulatory-Inhibitor; SI – Structure-Inhibitor; Un – Unknown function. Numbers in parenthesis represent the number of protein complexes.

Table 3: Distribution of complexes based on organism source and function

		SO	DO	SP
NO	R	56	0	0
NO	E	53	0	0
NO	SA	6	0	0
NO & O	EI	9	23	0
NO & O	Un	1	1	0
NO & O	I	19	8	0
O	RI	0	9	0
O	SI	0	2	5
	Other	0	0	5

SO – Same Organism; DO – Different Organism; SP – Synthetic Partner; R – Regulatory complexes; E – Enzyme complexes; I – Immune; SA – Structural Assembly; EI – Enzyme-inhibitor; RI – Regulatory-Inhibitor; SI – Structure-Inhibitor; Un – Unknown function. Numbers in parenthesis represent the number of protein complexes. NO = non-obligatory and O = obligatory. Numbers indicate the number of complexes under each category.

Table 4: Grouping of complexes based on class and function

Function/Class	Class A	Class B	Class C	Class D
Regulatory	39	17	0	0
Enzyme complexes	31	22	0	0
Structural assembly	5	0	1	0
Enzyme Inhibitor	20	12	0	0
Unknown function	1	1	0	0
Immune	16	10	0	0
Regulatory Inhibitor	5	4	0	0
Structural Inhibitor	1	2	0	0
Synthetic	4	1	0	0

Numbers indicate the number of complexes under each category.