

Interaction modes at protein hetero-dimer interfaces

Ayyappan Vaishnavi^{1,3}, Gopichandran Sowmya^{1,3}, Jayaseelan Kalaivani^{1,3}, Selvarajan Ilakya^{1,3}, Uma Kanguane¹, Pandjassaram Kanguane^{1,2*}

¹Biomedical Informatics, Pondicherry 607 402, India; ²AIMST, Kedah 08100, Malaysia; ³Equal contributions; Pandjassaram Kanguane – E-mail: kanguane@bioinformatics.net; Phone: +91 413 2633 589; +91 413 2633 722; *corresponding author

Received November 15, 2009; Revised December 07, 2009; Accepted December 16, 2009; Published January 20, 2010

Abstract:

Hetero dimer (different monomers) interfaces are involved in catalysis and regulation through the formation of interface active sites. This is critical in cell and molecular biology events. The physical and chemical factors determining the formation of the interface active sites is often large in numbers. The combined role of interacting features is frequently combinatorial and additive in nature. Therefore, it is important to determine the physical and chemical features of such interactions. A number of such features have been documented in literature since 1975. However, the use of such interaction features in the prediction of interaction partners and sites given their sequences is still a challenge. In a non-redundant dataset of 156 hetero-dimer structures determined by X-ray crystallography, the interacting partners are often varying in size and thus, size variation between subunits is an important factor in determining the mode of interface formation. The size of protein subunits interacting are either small-small, large-large, medium-medium, large-small, large-medium and small-medium. It should also be noted that the interface formed between subunits have physical interactions at N terminal (N), C terminal (C) and middle (M) region of the protein with reference to their sequences in one dimension. These features are believed to have application in the prediction of interaction partners and sites from sequences. However, the use of such features for interaction prediction from sequence is not currently clear.

Keywords: protein-protein interaction, mode of interaction, protein size, interface

Background:

Protein hetero-dimer subunit interaction is important in regulation and catalysis in living cells. The modes and types of protein-protein interactions using hetero-dimer protein complexes are numerically large in both prokaryotic and eukaryotic cells. However, documented data on such interactions are inadequate in the literature [1-17]. The formation of the interface between the two subunits is governed by both biophysical and chemical features as described in a number of studies elsewhere [1-17]. The documented features available thus far are based on structural datasets of protein-protein complexes of limited size. In these studies, geometrical (interface size, planarity, sphericity and complementarity) and chemical properties (the types of amino acid chemical groups, hydrophobicity, electrostatic interactions and H-bonds) are frequently analyzed.

Janin and colleagues have described the principles of protein-protein interaction since 1975 using a modest dataset of 3 protein complexes to more than 75 complexes in recent years [1-11]. Some of the parameters described by them for understanding protein subunit interactions include, close atomic packing [1, 2, 8], hydrophobicity and its free energy [1, 2, 10, 16], structural mobility and interface conformation changes [4, 11], Interface area based crude energy function [5], surface complementarity [8], interface size and chemistry [3, 6, 13, 15, 16, 17], statistically derived mean-field potential [7], polar interactions [8], atomic nature of recognition sites [8, 9, 12], interface residue propensity score [10], hydrophobic score indexes for atomic packing [10] and interface hydrophobic patches [17]. Thornton and colleagues simultaneously studied size and shape, surface complementarity; interface propensity, hydrophobicity, H-bonds, segmentation & secondary structures and conformational changes in protein subunit complexes [12, 13, 14]. Nussinov and colleagues described hydrogen bonds and their geometry, salt bridges and their distributions, interface hydrophobicity and, charge distributions at the interface [15, 16]. This data in later years lead to the development and benchmarking of interaction functions for protein docking predictions by Weng and colleagues [18-19]. The size of the test cases used for the identification of structural features in the development of the docking scoring functions is often the challenge in protein-protein interaction prediction. Despite these progresses, the use of such interaction features for the prediction of interaction partners and sites given their sequences is yet a challenge. Our interest is to describe hetero-dimer protein-protein interactions using size and mode of interactions in a non-redundant

dataset of 156 hetero-dimer structures determined by X-ray crystallography.

Methodology:

Dataset:

We used a non-redundant dataset of 156 hetero-dimers created by Zhanhua and colleagues (Table 1 in supplementary material) [20]. This non-redundant dataset was created from an initial redundant set of 2,488 hetero-dimer structures downloaded from PDB (Protein Databank) and PQS (Protein Quaternary Structure Server). The redundant entries were removed using a 30% sequence similarity cut-off. These structures are solved by X-ray crystallography with resolution ≤ 2.5 Å. Each entry in the dataset is a complex of two different monomers (different proteins) of varying lengths (Table 2 in supplementary material). The minimum, maximum lengths for subunit A is 68 and 535, respectively. The same for subunit B is 41 and 456, respectively. The dataset mean for subunit A and B are about 140 and 252, respectively. However, the standard deviation about the mean for subunit A and B are about 109 and 78, respectively. The categorization of the dataset based on the source type for chains A and B is given in Table 3 (see supplementary material). The dataset is most represented with complexes having chains A and B from the same source and these have regulatory/catalytic functions. The distribution of the protein-protein complexes with different A & B source types are given in Table 4 (see supplementary material).

Size based grouping of complexes

Size variation between the interacting partners is an important factor in determining the mode of interface formation. The size of protein subunits interacting are either small-small, large-large, medium-medium, large-small, large-medium and small-medium (Figure 1). Proteins of sizes 100-300 residues are most represented for subunit A and 50-200 residues for subunit B in the dataset (Table 5 in supplementary material).

Modes of interface interaction

The interface formed between subunits A and B have physical interactions at N terminal (N), C terminal (C) and middle (M) region of the protein with reference to its sequence in one dimension (Figure 2). A representation of the protein – protein interface in 3D, 2D and 1D is shown for chemo-taxis proteins chey and chea (PDB ID: 1FFG) in Figure 2. In this dataset, nearly 33% of interfaces have NMC - NMC mode of interactions (Table 6). However, the dataset contains a wide

type of interaction modes like N-N, C-C, NMC-NMC, M-M, NC-NC, NM-NM, and MC-MC (**Table 4** in supplementary material).

Interface residues

Interface residues in a protein-protein complex are identified by calculating the change in accessible surface area (delta ASA) upon complex formation from individual monomers using the software Surface Racer 5.0 with probe radius 1.4 Å in Lee and Richard implementation [21]. The interface residues are represented as a function of residue number using delta ASA in an X-Y plot (**Figure 2b**).

Results:

Figure 1 illustrates the role of protein size in protein-protein interactions using examples. Examples of protein-protein complexes are used to show interactions between (a) small-small protein; (b) large-large proteins; and (c) large-small proteins. Thus, protein-protein complexes are formed between small, large and medium sized proteins in different combinations. The protein-protein complexes in the dataset of 156 non-redundant structures are subsequently classified based on protein sizes forming the interface (**Table 5** in supplementary material). Proteins of sizes 100-300 residues are most represented for subunit A and 50-200 residues for subunit B in the dataset.

Figure 2 illustrates the representation of protein-protein interfaces in 3-dimensional (structure complex in 2a), 2-dimensional (X-Y plots in 2b) and 1-dimensional (sequence in 2c) using the example structural

complex of chey - chea with the protein databank (PDB) entry 1FFG. A qualitative representation of the interface between the interaction proteins chey and chea is given in **Figure 2a** using 3-dimensional structural features. However, a quantitative understanding of the interacting residues forming the interface is not possible using this representation. Thus, we represented the interface residues as a function of residue number using delta ASA in **Figure 2b**. The measure of delta ASA calculated using the Lee and Richards implementation (1977) for the complex determine the extent of residues involved in the formation of the interface. This shows the region (N or C terminal) in the sequence (1-dimensional) involved in the formation of the interface. **Figure 2b** shows that residues towards the C terminal in chain A and residues in the middle regions of chain B are involved in the formation of the interface. **Figure 2c** highlights (bold letters) interface residues in sequence (1-dimension). Thus, **Figure 2** shows the representation of the protein-protein interface using structure (3D), X-Y plots (2D) and sequence (1D) for visual comparison in all the three dimensions. We extrapolated the principle to all the 156 protein-protein complex structures in the dataset. As shown in **Figure 2**, the interface formed between subunits A and B have physical interactions at N terminal (N), C terminal (C) and middle (M) region of the protein with reference to its sequence in one dimension for all the structures (**Table 6** in supplementary material). In this dataset, nearly 33% of interfaces have NMC - NMC mode of interactions. However, the dataset contains a wide type of interaction modes like N-N, C-C, NMC-NMC, M-M, NC-NC, NM-NM, and MC-MC.

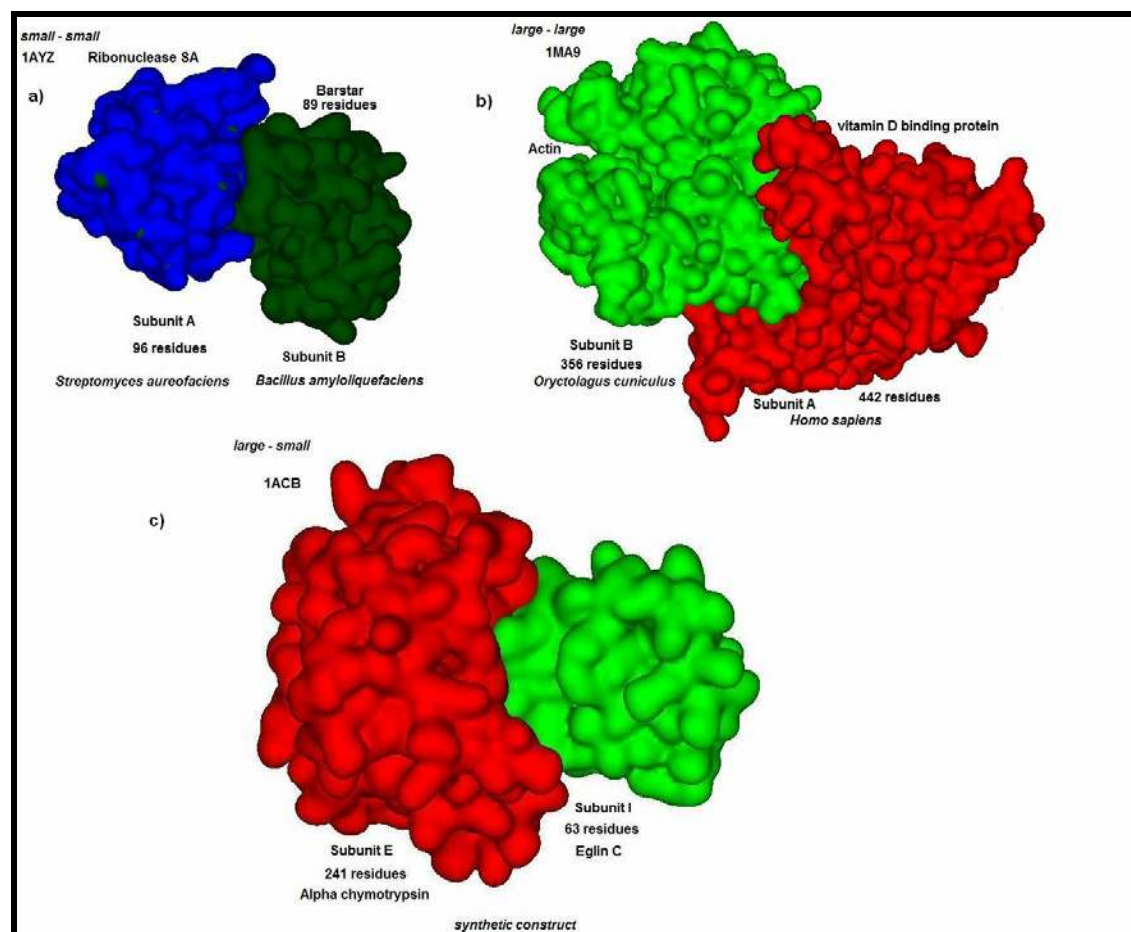


Figure 1: Protein-protein interactions and protein size is illustrated using examples; (a) small protein – small protein; (b) large protein – large protein; (c) large protein – small protein.

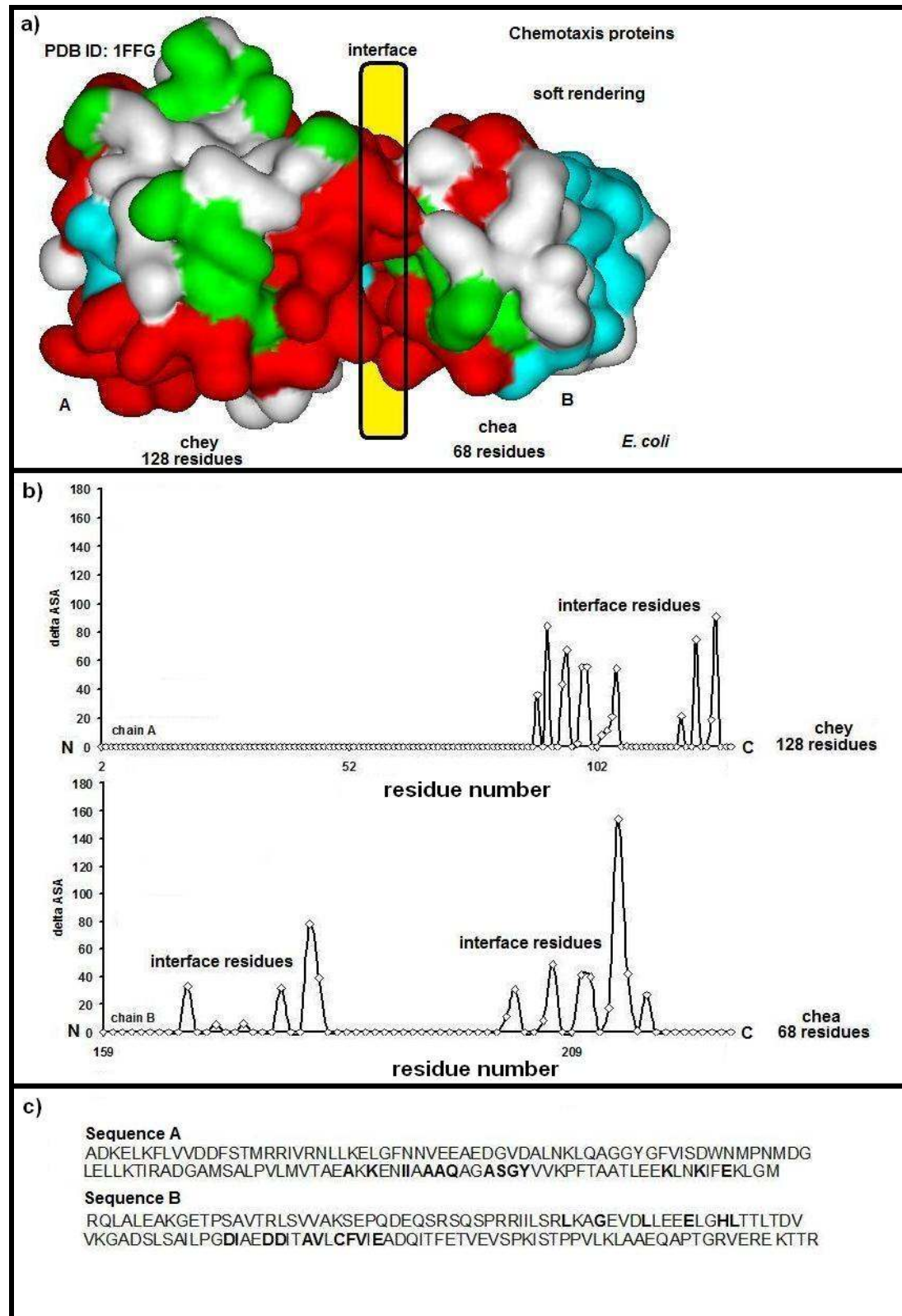


Figure 2: Representation of protein-protein interfaces in 3D (structure complex in panel a), 2D (X-Y plots using X-Y plots in panel b) and 1D (sequence in panel c).

Discussion:

The principle of protein-protein interactions in living cells is a subject of debate and discussion for about four decades now [1-18]. The importance of a protein-protein interface for catalysis and regulation has been frequently described in recent years (Table 3 in supplementary material). It is also known that many such interactions are also associated with those of enzyme -inhibitors and antigen-antibody complexes (Table 4 in supplementary material). The formation of an interface in cellular milieu is often dictated through bio-physical features followed by chemical properties to activate a cascade of subsequent reactions. The biophysical aspects of the protein-protein interaction principles have been extensively studied using protein complexes determined by X-ray crystallography [1-16]. Janin & colleagues [1-11], Thornton & colleagues [12-14] and Nussinov & colleagues [15-16] studied protein complexes in protein databank (PDB) and documented several structural features that are found to be significant. The features described thus far include atom packing, hydrophobicity, polarity, shape complementarity, conformational effect, interface size and residue propensity and interface hydrogen bonds. The information documented has significantly improved the understanding of protein-protein interactions using biophysical data. The use of these data to develop interaction scoring functions for application in protein docking has been shown [18-19]. Thus, it is now fairly possible to dock protein structures given the structures for the interacting partners through the identification of recognition sites in known structural partners. Nonetheless, the prediction of interaction partners and sites given their sequences is still a grand challenge. Hence, we studied a non-redundant dataset of 156 hetero-dimers to understand their mode of interactions in structure to relate to their corresponding sequence. The dataset consists of 89 structures with regulatory role and 61 with enzyme-inhibitor function (Table 3 and Table 4 in supplementary material). The protein sizes involved in the formation of the complexes are variable in the dataset with significant standard deviation about the mean (Table 2 in supplementary material). Protein-protein complexes in the dataset are formed between small, large and medium sized proteins in different combinations as shown in Figure 1. The protein-protein complexes in the dataset of 156 structures are subsequently classified based on protein sizes forming the interface (Table 5 in supplementary material). Proteins of sizes 100-300 residues are most represented for subunit A and 50-200 residues for subunit B in the dataset. Thus, size variation between the interacting partners is an important factor in determining the mode of interface formation. The size of protein subunits interacting are either small-small, large-large, medium-medium, large-small, large-medium and small-medium. Nonetheless, it should be emphasized that conceptualization of size variations between and across interacting partners is a challenge. This is an important factor in hetero-dimer interaction as proteins of different sizes are capable of interacting in specific and sensitive ways. We believe that formulation of a method to incorporate the protein size components for potential interaction and selection will provide insight to detect protein partners and sites from sequence. It is also important to categorize interacting proteins based on their size to relate and link their potential partners.

The future challenge is to detect interaction partners and sites from their primary sequence. This requires relating interface features in structures to sequences. Hence, we attempted to represent interface features in 3-dimension using 2-dimensional X-Y plots as shown in Figure 2b. In this example, the chey-chea protein complex (PDB ID 1FFG) is shown to interact with each other structurally in 3-dimension. The interface is represented qualitatively and realized visually by the reader. Thus, a qualitative representation of the interface between the interaction proteins chey and chea is established in Figure 2a using 3-dimensional structural features. However, this representation does not provide information on the interacting residue numbers at the interface. Thus, we represented the interface residues as a function of residue number using delta ASA in Figure 2b. The measure of delta ASA calculated using the Lee and Richards implementation (1977) for the complex

determine the extent of residues involved in the formation of the interface. This shows the region (N or C terminal) in the sequence (1-dimensional) involved in the formation of the interface. Figure 2b shows that residues towards the C terminal in chain A and residues in the middle regions of chain B are involved in the formation of the interface. Figure 2c highlights (bold letters) interface residues in sequence (1-dimension). Thus, Figure 2 shows the representation of the protein-protein interface using structure (3D), X-Y plots (2D) and sequence (1D) for visual comparison in all the three dimensions. We then used the same principle to extrapolate to all the 156 protein-protein complex structures in the dataset. As shown in Figure 2, the interface formed between subunits A and B have physical interactions at N terminal (N), C terminal (C) and middle (M) region of the protein with reference to its sequence in one dimension for all the structures (Table 6 in supplementary material). Nearly 33% of interfaces in the dataset have NMC - NMC mode of interactions. However, the dataset contains a wide type of interaction modes like N-N, C-C, NMC-NMC, M-M, NC-NC, NM-NM, and MC-MC. Thus, the analysis provides a method to relate interface features in structure to its occurrence in sequence using delta ASA represented in X-Y plots. This method of relating interface features from structure to sequence is interesting. We believe the procedure when incorporated in a formulation (development of a method) will help to detect interaction partners and sites from sequences.

Conclusion:

Protein hetero-dimer (different monomers in sequence composition) subunits interact in cellular systems to create interface active sites for catalysis and regulation. The formation of the interface active site is generally determined by the interface shape complementarity and chemical properties between the interacting subunits. These interactions are usually specific and sensitive between the interacting partners. The possible numbers of such interacting partners are often huge and their magnitude is frequently beyond our realization for both prokaryotic and eukaryotic cells. The physical and chemical properties determining subunit interactions have been shown in a relatively small dataset of hetero-dimer structures determined by X-ray crystallography. However, the information thus far gleaned from such datasets is typically limited for the prediction of interaction partners and sites given their sequences. Hence, we analyzed a non-redundant dataset of 156 hetero-dimer structures determined by X-ray crystallography to assemble information to understand interface types using size, shape and interaction mode for potential application in protein-protein interaction prediction from sequence. In this analysis, the dataset mean size of subunit A (140 residues) is smaller than subunit B (250 residues) with a large standard deviation (109.4). Hence, the interacting partners are often varying in size. Thus, size variation between the interacting partners is an important factor in determining the mode of interface formation. The size of protein subunits interacting are either small-small, large-large, medium-medium, large-small, large-medium and small-medium. Proteins of sizes 100-300 residues are most represented for subunit A and 50-200 residues for subunit B in the dataset. Nonetheless, it should be emphasized that conceptualization of size variations between and across interacting partners is a challenge. The interface formed between subunits A and B have physical interactions at N terminal (N), C terminal (C) and middle (M) region of the protein with reference to its sequence in one dimension. In this dataset, nearly 33% of interfaces have NMC - NMC mode of interactions. However, the dataset contains a wide type of interaction modes like N-N, C-C, NMC-NMC, M-M, NC-NC, NM-NM, and MC-MC. These features are implied to have application in the prediction of interaction partners and sites. Nonetheless, the comprehension of such features for protein-protein interaction prediction from sequence is not explicit at the moment.

Acknowledgement

We express our sincere appreciation to all members of Biomedical Informatics for many discussions on the subject of this article.

References:

- [1] C Chothia, J Janin, *Nature* 256:705-8 (1975) [PMID: 1153006]
[2] C Chothia et al, *Proc Natl Acad Sci* 73:3793-7 (1976) [PMID:1069263]
[3] S Miller et al, *Nature* 328:834-6 (1987) [PMID:3627230]
[4] J Janin, C Chothia, *J Biol Chem* 265:16027-30 (1990) [PMID:2204619]
[5] S Duquerroy et al, *Proteins* 11:271-80 (1991) [PMID:1758882]
[6] J Janin, *Biochimie* 77: 497-505 (1995) [PMID:8589061]
[7] C H Robert, J Janin, *Mol Biol* 283: 1037-47 (1998) [PMID:9799642]
[8] L Lo Conte et al, *J Mol Biol* 285:2177-98 (1999) [PMID:9925793]
[9] P Chakrabarti, J Janin, *Proteins* 47:334-43 (2002) [PMID:11948787]
[10] RP Bahadur et al, *Mol Biol* 336:943-55 (2004) [PMID:15095871]
[11] J Bernauer et al, *Phys Biol* 2:S17-23 (2005) [PMID:16204845]
[12] S Jones, JM Thornton, *Prog Biophys Mol Biol* 63:31 (1995) [PMID:7746868]
[13] S Jones, JM Thornton, *Proc Natl Acad Sci* 93:13 (1996) [PMID:8552589]
[14] IM Nooren, JM Thornton, *J Mol Biol* 325:991 (2003) [PMID:]
[15] D Xu et al, *Protein Engineering* 10:999 (1997) [PMID:9464564]
[16] CJ Tsai et al, *Protein Science* 6:53 (1997) [PMID:9007976]
[17] P Lijnzaad, P Argos, *Proteins* 28:333 (1997) [PMID:9223180]
[18] J Mintseris et al, *Proteins* 60:214-6 (2005) [PMID:15981264]
[19] H Hwang et al, *Proteins* 73:705-9 (2008) [PMID:18491384]
[20] C Zhanhua et al, *Bioinformatics* 1(2): 28-39 (2005) [PMID:17597849]
[21] O V Tsodikov et al, *J. Comput. Chem* 23, 600-609 (2002) [PMID:11939594]

Edited by V. S. Mathura

Citation: Vaishnavi et al., *Bioinformatics*, 4 (7) 310-319, (2010)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material

Table 1: Hetero-dimer dataset

PDB ID	Chain 1	ML	Chain 2	ML
1FFG	B	68	A	128
1E44	B	96	A	84
1AY7	A	96	B	89
3YGS	P	97	C	95
1KTZ	B	106	A	82
1V74	A	107	B	87
3FAP	A	107	B	94
1BND	A	109	B	108
1B27	A	110	D	90
1QAV	B	115	A	90
1HL6	A	119	B	137
1JLT	B	122	A	122
1DFJ	E	124	I	456
1UJZ	B	127	A	87
1EAY	A	128	C	67
1UUZ	D	129	A	130
1OP9	B	130	A	121
1FR2	B	131	A	83
1JTT	A	133	L	129
1UZX	A	135	B	75
1JTP	A	135	L	129
1S1Q	A	137	B	71
1OHZ	A	140	B	56
1MG9	B	143	A	84
1OO0	A	144	B	92
1F2T	A	145	B	143
2HBE	B	146	A	141
1G4Y	R	147	B	81
1EM8	A	147	B	110
1J7D	B	149	A	140
1D4V	B	163	A	117
1M9E	A	164	D	135
1KSH	A	164	B	141
1OXB	A	166	B	124
1GUA	A	167	B	76
1C1Y	A	167	B	77
1LFD	B	167	A	87
1HE1	C	176	A	135
1JKG	B	180	A	139
1Q40	B	180	A	163
1SHW	B	181	A	138
1DS6	A	181	B	179
1CXZ	A	182	B	86
1CT4	E	185	I	51
4SGB	E	185	I	51

1SGD	E	185	I	51
1E96	B	185	A	178
1R8S	E	187	A	160
1IAR	B	188	A	129
1TE1	B	190	A	274
1P5V	A	191	B	136
1NRJ	B	191	A	147
1AXI	B	191	A	175
1YCS	B	193	A	191
1NF3	A	194	C	123
1US7	B	194	A	207
1RE0	B	195	A	162
1TX4	A	196	B	174
2NGR	B	196	A	191
1M4U	A	199	L	112
1PVH	A	201	B	169
1STF	E	212	I	98
1N0L	A	212	B	116
1F5R	A	216	I	57
1SLU	B	216	A	131
1PPF	E	218	I	56
1P2J	A	220	I	56
1EUV	A	221	B	79
1BRB	E	223	I	51
1D6R	A	223	I	58
1UGH	E	223	I	82
1AVW	A	223	B	171
1OPH	B	223	A	375
1K9O	E	223	I	376
1DZB	A	224	X	129
1H2S	A	225	B	60
1HCG	A	236	B	51
1CHO	E	238	I	53
1P2M	A	238	B	58
1NW9	B	238	A	91
1PQZ	A	238	B	99
2KIN	A	238	B	100
1EAI	A	240	C	61
1JW9	B	240	D	81
1ACB	E	241	I	63
1FYH	A	242	B	201
1CGI	E	245	I	56
1USU	A	246	B	132
1P57	B	247	A	110
1L4Z	A	248	B	125
1KA9	F	251	H	195
1GPW	A	253	B	200
1H31	A	260	B	138
1RKE	A	262	B	176

1JTD	A	262	B	273
1NPE	A	263	B	164
1S0W	A	263	C	165
1SPB	S	264	P	71
1OKK	D	265	A	290
1ABR	B	267	A	251
1NQI	B	272	A	123
1GL4	A	273	B	89
1R0R	E	274	I	51
1CSE	E	274	I	63
1ONQ	A	274	B	99
1MEE	A	275	I	64
2SNI	E	275	I	64
1SCJ	A	275	B	71
2SIC	E	275	I	107
1RJ9	A	277	B	282
2TEC	E	279	I	63
1LW6	E	281	I	63
1BPL	B	290	A	179
1S6V	A	294	B	108
1F5Q	A	296	B	247
1H1S	A	296	B	258
1PDK	A	296	B	258
1FIN	A	298	B	260
1I1R	A	301	B	167
1TA3	B	301	A	274
1DTD	A	303	B	61
1BLX	A	305	B	160
1ITB	B	310	A	153
1EFV	A	312	B	252
1WQ1	G	320	R	166
1F34	A	325	B	138
1BRL	A	340	B	319
1K1I	B	342	A	178
1G4U	S	360	R	180
1NLV	A	364	G	123
1JQL	A	366	B	140
1D4X	A	368	G	124
1SVX	B	369	A	157
1YVN	A	372	G	125
1HX1	A	377	B	112
1EUC	B	393	A	306
1FCD	A	401	C	174
1AVA	A	403	C	181
1FT1	B	416	A	315
1KXP	D	438	A	349
1F60	A	440	B	90
1MA9	A	442	B	356
1IBR	B	458	A	169

1O6S	A	461	B	105
1SMP	A	468	I	100
1JIW	P	470	I	105
1TMQ	A	470	B	117
1VG0	A	481	B	182
1DHK	A	495	B	195
1BVN	P	496	T	71
1KXV	A	496	C	119
1LUJ	A	501	B	71
1OR0	B	510	A	152
1H2A	L	534	S	267
1UBK	L	534	S	267
1KU6	A	535	B	61

ML = monomer length

Table 2: Hetero-dimer data statistics

Chain	A	B
Dataset size	156	156
Monomer length		
Minimum	68	51
Maximum	535	456
Dataset mean	139.8	252.20
SD	109.24	77.6

Table 3: Categorization of the dataset based on source type

Category	Source type	Count
1	A && B same	89
2	A && B different	61
3	A && B synthetic	3
4	A B synthetic	3
	Total	156

Table 4: Distribution of the protein-protein complexes with different A & B source types

Category	Protein function	Total
1	Enzyme inhibitor	24
2	Antibody-antigen	7
3	Signal transduction protein	5
4	Protease inhibitor	4
5	Others complex	19
	Total	61

Table 5: Length-wise distribution of 156 protein-protein complexes

Chain B	Chain A											Total
	<=50	51-100	101-150	151-200	201-250	251-300	301-350	351-400	401-450	451-500	>500	
<=50	0	0	0	0	0	0	0	0	0	0	0	0
51-100	0	3	14	7	19	10	1	0	1	2	2	59

101-150	0	1	11	12	6	4	1	5	0	4	0	44
151-200	0	0	0	9	2	6	5	2	2	3	1	30
201-250	0	0	0	1	1	1	0	0	0	0	0	3
251-300	0	0	0	1	0	7	2	0	0	0	2	12
301-350	0	0	0	0	0	0	1	1	2	0	0	4
351-400	0	0	0	0	2	0	0	0	1	0	0	3
401-450	0	0	0	0	0	0	0	0	0	0	0	0
451-500	0	0	1	0	0	0	0	0	0	0	0	1
>500	0	0	0	0	0	0	0	0	0	0	0	0
Total	0	4	26	30	30	28	10	8	6	9	5	156

Table 6: Clustering of protein-protein interfaces

Chain B	Chain A							Total
	N	C	M	NMC	NC	NM	MC	
N	1	0	1	1	0	1	0	4
C	0	0	0	0	0	0	0	0
M	0	0	4	2	0	0	1	7
NMC	2	2	6	37	6	11	6	70
NC	0	3	3	7	4	1	3	21
NM	4	0	3	13	4	3	0	27
MC	0	1	4	11	3	3	5	27
Total	7	6	21	71	17	19	15	156

N = N terminal; C = C terminal; M = Middle