Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Protein subunit interfaces: heterodimers versus homodimers

Cui Zhanhua, Jacob Gah-Kok Gan, Li lei, Meena Kishore Sakharkar, Pandjassarame Kangueane*
School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798;
Pandjassarame Kangueane* - Email: mpandjassarame@ntu.edu.sg; Phone: +65 6790 5836; Fax: +65 6774 4340;
* Corresponding author
received July 09, 2005; revised August 05, 2005; accepted August 10, 2005; published online August 11, 2005

Abstract:
Protein dimers are either homodimers (complexation of identical monomers) or heterodimers (complexation of non-identical monomers). These dimers are common in catalysis and regulation. However, the molecular principles of protein dimer interactions are difficult to understand mainly due to the geometrical and chemical characteristics of proteins. Nonetheless, the principles of protein dimer interactions are often studied using a dataset of 3D structural complexes determined by X-ray crystallography. A number of physical and chemical properties govern protein dimer interactions. Yet, a handful of such properties are known to dominate protein dimer interfaces. Here, we discuss the differences between homodimer and heterodimer interfaces using a selected set of interface properties.

Keywords: dimer; heterodimer; homodimer; interface; interaction; molecular recognition; interface properties; interface area; hydrogen bonds; hydrophobicity; interface residues

Background:
Protein subunit interaction (either homodimer or heterodimer) is an important phenomenon in regulation and catalysis. Thousands of such interactions are theoretically possible in a combinatorial manner. The task of documenting each of these interactions is laborious. Therefore, prediction of subunit interaction sites either from folded structures or from primary sequences is required. However, this objective is currently ambitious due to the limited knowledge on the principles of protein subunit interactions using structural data. Therefore, it is our interest to study the nature of subunit interactions. Several studies report on these interactions. Jones & Thornton (used 59 protein complexes) **[1]**, Xu & colleagues (used 319 protein-protein interfaces) **[2]**, Tsai & colleagues (used 362 protein-protein interfaces) **[3]**, Lo Conte & colleagues (used 75 hetero-complexes) **[4]**, Chakrabart & Janin (used 70 hetero-complexes) **[5],** Brinda & colleagues (used 20 homodimers) **[6]**, Bahadur & colleagues (used 122 homodimers) **[7]**, Nooren & Thornton (used 39 protein dimers) **[8]**, Caffrey & colleagues (used 64 protein-protein interfaces) **[9]** and Zhanhua & colleagues (used 65 heterodimers) **[10]**, utilized a dataset of protein complexes determined by X-ray crystallography to examine the properties of subunit interaction. Protein subunit interfaces in these studies have been characterized using geometrical properties (interface size, planarity, sphericity and complementarity) and chemical properties (the types of amino acid chemical groups, hydrophobicity, electrostatic interactions and H-bonds). These studies are influenced by dataset size and their characteristics. However, the analyses are based on limited datasets consisting of heterogeneous (disproportionate mixture of homodimers and heterodimers) data.

The analyses report on the role of inter-subunit H-bonds in protein subunit association. The numbers of H-bonds vary in different studies. **[2, 4, 7, 8, 11]** On average, Bahadur & colleagues show 9.0 H-bonds per homodimer interface with an r value of 0.75 (Pearson correlation coefficient) between H-bonds and interface area. **[7]** Jones & Thornton (used 32 homodimers) shows 0.88 H-bonds per 100 $Å^2$ interface area with an r value of 0.77 between H-bonds and interface area. **[11]** Lo Conte *et al.,* show an average of 10.1 H-bonds with one H-bond per 170 $Å^2$ interface area and an r value of 0.84 between H-bonds and interface area. **[4]** Xu & colleagues also show 11 H-bonds per subunit with an r value of 0.89 between H-bonds and interface area. **[2]** The r value between H-bonds and interface area in these studies varies from 0.75 to 0.89. This variation is influenced primarily by dataset size and nature of data.

Previous studies also show that hydrophobic effect plays an important role in protein association **[3, 7, 12]**, yet not as much as in protein folding. **[3]** There studies showed that protein interfaces are more hydrophobic than surfaces, but less than interior. Hydrophobic effect was measured by the buried non-polar surface area (or percent burial) of residue types. **[3]** The study showed that the ratio between buried hydrophobic and buried hydrophilic residues is approximately 1.5. **[3]** Hydrophobic residues (except ALA) and the charged residue ARG are predominantly present at protein-protein interfaces with TYR and TRP having highest propensity. **[4, 6, 7, 12, 13]**

Interface size is yet another important property widely used to describe protein-protein interfaces and it is usually characterized by interface area. The number of interface residues is linearly correlated to interface area ($r \geq 0.96$) in several studies. **[5,7]** However, the mean number of

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

interface residues varies between these studies. It is shown that the mean is 52 **[7]**, 57 **[5]**, 53.7 **[14]**, 44.4 (for homodimers) and 42.2 (for heterodimers). **[9]** Thus, the number of interface residues vary within a narrow rang of 42 and 57 in these studies.

Here, we created two extended datasets of mutually exclusive homodimers and heterodimers. We believe that these exclusive datasets can reduce data bias to differentiate heterodimer and homodimer interfaces.

Methodology:

Creation of heterodimer and homodimer dataset:

A total of 2488 heterodimer candidates and 1324 homodimer candidates were downloaded from PDB (Protein Databank) and PQS (Protein Quaternary Structure Server). We then created a non-redundant dataset of 156 heterodimers and 170 homodimers (Table 1) such that they satisfy the following conditions. These include: (1) each chain $\geq$ 50 residues; (2) structures determined by x-ray crystallography; (3) resolution $\leq$2.5 Å; (4) the structure with the highest resolution was selected where more than one structure was available; (5) redundant entries were removed at a sequence similarity cut-off of $\geq$30%. **[15]**

Calculation of interface parameters:

Interface area

ASA (accessible surface area) was calculated using NACCESS **[16]** with a probe radius of 1.4 Å and interface area is defined by ΔASA (change in ASA upon complexation from monomer to dimer state) as described elsewhere. **[10]**

Inter-subunit H-bonds

A hydrogen bond is a polar interaction between two electronegative atoms, where a donor and an acceptor participate. The number of H-bonds formed between subunits was calculated using the program HBPLUS. **[17]**

Hydrophobicity

Interface hydrophobicity was estimated using the equation $\sum_{i=1}^{N}(HV)/N$ **[18]**, where *N* is the number of interface residues, and *HV* is the hydrophobicity scale for each residue. **[18]**

Interface residues propensity

Interface residues show an Δ ASA (change in accessibility) of $\geq$ 5% upon complexation. Interface residue propensities were calculated using the percentage frequencies of 20 residues using the following functions:

$$P_{IS}(i) = f_{interface}(i) / f_{surface}(i)$$

$$P_{II}(i) = f_{interface}(i) / f_{interior}(i)$$

where $P_{IS}(i)$ is residue interface propensity compared to protein surface, $P_{II}(i)$ is residue interface propensity compared to protein interior, $f_{interface}(i)$ is residue frequency at the protein interface, $f_{surface}(i)$ is residue frequency at the protein surface, $f_{interior}(i)$ is residue frequency at the protein interior.

Results and Discussion:

Dimer interactions are characterized by a large combination of physical-chemical parameters. Analysis of dimer structures can provide insight into the principles of protein-protein complexation and help develop models to predict interaction sites. The multi dimensional scaling method applied in a recent study reduced a large pool of interface parameters to a small set of six critical properties for heterodimers. **[10]** Zhanhua *et al.*, 2005, showed that the six selected parameters were sufficient to describe subunit interfaces instead of the complete parameter space. Here, we use these selected set of properties to discuss the interface differences between 156 heterodimers and 170 homodimers. The properties used in this study are (1) interface residues, (2) interface H-bonds, (3) interface hydrophobicity, (4) interface residue-composition.

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

Hetero-dimers

PDB code	Resolution (Å)	Chain one	Name of chain one	Length	Chain two	Name of chain two	Length
1YCS	2.2	B	53BP2	193	A	P53	191
1ABR	2.1	B	Abrin-A	267	A	Carbohydrate	251
1KU6	2.5	A	Acetylcholinesterase	535	B	Fasciculin 2	61
1LFD	2.1	B	Active ras protein	167	A	Ras-interacting domain of ralgds	87
1JIW	1.7	P	Alkaline metalloproteinase	470	I	Proteinase inhibitor	105
1BPL	2.2	B	Alpha-amylase	290	A	Alpha-amylase	179
1KXV	1.6	A	Alpha-amylase	496	C	Camelid VHH domain cab10	119
1TMQ	2.5	A	Alpha-amylase	470	B	Ragi bifunctional inhibitor	117
1BVN	2.5	P	Alpha-amylase	496	T	Tendamistat	71
1ACB	2.0	E	Alpha-chymotrypsin	241	I	Eglin C	63
1CHO	1.8	E	Alpha-chymotrypsin	238	I	Turkey ovomucoid third domain	53
1CGI	2.3	E	Alpha-chymotrypsinogen	245	I	Trypsin inhibitor	56
1SLU	1.8	B	Anionic trypsin	216	A	Ecotin	131
1RE0	2.4	B	ARF guanine-nucleotide exchange factor 1	195	A	ADP-ribosylation factor 1	162
1KSH	1.8	A	ARF-like protein 2	164	B	Cyclic phosphodiesterase delta-subunit	141
1MG9	2.3	B	ATP dependent CLP protease	143	A	Protein YLJA	84
1BRL	2.4	A	Bacterial luciferase	340	B	Bacterial luciferase	319
1AVA	1.9	A	Barley alpha-amylase 2	403	C	Barley alpha-amylase/subtilisin inhibitor	181
1B27	2.1	A	Barnase	110	D	Barstar	90
1LUJ	2.5	A	Beta-catenin	501	B	Beta-catenin-interacting protein ICAT	71
1S0W	2.3	A	Beta-lactamase tem	263	C	Beta-lactamase inhibitory protein	165
1BND	2.3	A	Brain derived neurotrophic factor	109	B	Neurotrophin 3	108
1D4X	1.8	A	C. Elegans actin 1/3	368	G	Gelsolin	124
1G4Y	1.6	R	Calmodulin	147	B	Calcium-activated potassium channel RSK2	81
1DTD	1.7	A	Carboxypeptidase A2	303	B	Metallocarboxypeptidase inhibitor	61
1NW9	2.4	B	Catalytic domain of caspase-9	238	A	Inhibitor of apoptosis protein 3	91
1OKK	2.1	D	Cell division protein	265	A	Signal recognition particle protein	290
1H1S	2.0	A	Cell division protein kinase 2	296	B	Cyclin A2	258
1OHZ	2.2	A	Cellulosomal scaffolding protein A	140	B	Endo-1,4-beta-xylanase Y	56
1HL6	2.5	A	CG8781 protein	119	B	Mago nashi protein	137
1P5V	1.7	A	Chaperone protein CAF1M	191	B	F1 capsule antigen	136
1PDK	2.4	A	Chaperone protein PAPD	296	B	Protein PAPK	258
1N0L	2.3	A	Chaperone protein PAPD	212	B	Mature fimbrial protein PAPE	116
1FFG	2.1	B	Chemotaxis protein chea	68	A	Chemotaxis protein chey	128
1EAY	2	A	Chey	128	C	Chea	67
1P2M	1.8	A	Chymotrypsinogen A	238	B	Pancreatic trypsin inhibitor	58
1HCG	2.2	A	Coagulation factor	236	B	Coagulation factor	51
1V74	2.0	A	Colicin D	107	B	Colicin D immunity protein	87
1E44	2.4	B	Colicin E3	96	A	Immunity protein	84
1FR2	1.6	B	Colicin E9	131	A	Colicin E9 immunity protein	83
1F5Q	2.5	A	Cyclin dependent kinase 2	296	B	Gamma herpesvirus cyclin	247
1FIN	2.3	A	Cyclin-dependent kinase	298	B	Cyclin A	260
1BLX	1.9	A	Cyclin-dependent kinase 6	305	B	P19ink4D	160
1M9E	1.7	A	Cyclophilin A	164	D	HIV-1 capsid	135
1S6V	1.9	A	Cytochrome C peroxidase	294	B	Cytochrome C	108

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

1R8S	1.5	E	Cytohesin 2	187	A	ADP-ribosylation factor 1	160
1UJZ	2.1	B	Designed colicin E7 dnase	127	A	Designed colicin E7 immunity protein	87
1NLV	1.8	A	Dictyostelium discoideum actin	364	G	Gelsolin	123
1H31	1.5	A	Diheme cytochrome C	260	B	Cytochrome C	138
1EM8	2.1	A	DNA polymerase III CHI subunit	147	B	DNA polymerase III PSI subunit	110
1JQL	2.5	A	DNA polymerase III, beta chain	366	B	DNA polymerase III delta subunit	140
1EAI	2.4	A	Elastase	240	C	Chymotrypsin isoinhibitor 1	61
1EFV	2.1	A	Electron transfer flavoprotein alpha chain	312	B	Electron transfer flavoprotein beta chain	252
1F60	1.7	A	Elongation factor EEF1A	440	B	Elongation factor EEF1BA	90
1TA3	1.7	B	Endo-1,4-beta-xylanase	301	A	Xylanase inhibitor protein I	274
1TE1	2.5	B	Endo-1,4-xylanase	190	A	Xylanase inhibitor protein I	274
3FAP	1.9	A	FK506-binding protein	107	B	FKBP12-rapamycin associated protein	94
1FCD	2.5	A	Flavocytochrome C sulfide dehydrogenase	401	C	Flavocytochrome c sulfide dehydrogenase	174
1NF3	2.1	A	G25k GTP-binding protein	194	C	PAR-6B	123
1NQI	2	B	Galactosyltransferase	272	A	Alpha lactalbumin	123
1WQ1	2.5	G	Gapette	320	R	Harvey-RAS	166
1OR0	2.0	B	Glutaryl acylase beat subunit	510	A	Glutaryl acylase alpha subunit	152
1AXI	2.1	B	Growth hormone receptor	191	A	Growth hormone	175
2NGR	1.9	B	Gtpase activating protein	196	A	GTP binding protein	191
1TX4	1.7	A	Gtpase-activating protein rhogap	196	B	Transforming protein RHOA	174
1AY7	1.7	A	Guanyl-specific ribonuclease SA	96	B	Barstar	89
1HX1	1.9	A	Heat shock cognate 71 KDA	377	B	Bag-family molecular chaperone regulator-1	112
1USU	2.2	A	Heat shock protein HSP82	246	B	AHA1	132
2HBE	2.0	B	Hemoglobin	146	A	Hemoglobin	141
1GPW	2.4	A	Hisf protein	253	B	Amidotransferase HISF	200
1CXZ	2.2	A	His-tagged transforming protein RHOA	182	B	PKN	86
1US7	2.3	B	HSP90 chaperone protein kinase	194	A	Heat shock protein HSP82	207
1KXP	2.1	D	Human vitamin D-binding protein	438	A	Actin, alpha skeletal muscle	349
1H2A	1.8	L	Hydrogenase	534	S	Hydrogenase	267
1KA9	2.3	F	Imidazole glycerol phosphtate synthase	251	H	Imidazole glycerol phosphtate synthase	195
1IBR	2.3	B	Importin beta-1 subunit	458	A	GTP-binding nuclear protein ran	169
1PVH	2.5	A	Interleukin 6 signal transducer	201	B	Leukemia inhibitory factor	169
1IAR	2.3	B	Interleukin-4 receptor alpha chain	188	A	Interleukin	129
1I1R	2.4	A	Interleukin-6 receptor beta chain	301	B	Viral IL-6	167
1O6S	1.8	A	Internalin A	461	B	E-cadherin	105
1KI1	2.3	B	Intersectin long form	342	A	G25k GTP-binding protein	178
2KIN	1.9	A	Kinesin	238	B	Kinesin	100
1PPF	1.8	E	Leukocyte elastase	218	I	Ovomucoid inhibitor	56
1OP9	1.9	B	Lysozyme C	130	A	Hl6 camel VHH fragment	121
1UUZ	1.8	D	Lysozyme C	129	A	Inhibitor of vertebrate lysozyme	130
1OO0	1.9	A	Mago nashi protein	144	B	Drosophila Y14	92
1SVX	2.2	B	Maltose-binding periplasmic protein	369	A	Ankyrin repeat protein OFF7	157
1PQZ	2.1	A	MCMV M144	238	B	Beta-2-microglobulin	99
1MEE	2.0	A	Mesentericopeptidase	275	I	Eglin-C	64
1JW9	1.7	B	Molybdopterin biosynthesis moeb protein	240	D	Molybdopterin converting factor	81
1Q40	2.0	B	Mrna export factor MEX67	180	A	Mrna transport regulator MTR2	163
1SHW	2.2	B	Neural kinase	181	A	Ephrin-A5	138
1QAV	1.9	B	Neuronal nitric oxide synthase	115	A	Alpha-1 syntrophin	90
1E96	2.4	B	Neutrophil cytosol factor 2	185	A	Ras-related C3 botulinum toxin substrate 1	178

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

1NPE	2.3	A	Nidogen	263	B	Laminin gamma-1 chain	164
1GL4	2.0	A	Nidogen-1	273	B	Proteoglycan core protein	89
1M4U	2.4	A	Noggin	199	L	Osteogenic protein 1	112
1FYH	2.0	A	Nterferon-gamma	242	B	Interferon-gamma receptor alpha chain	201
1STF	2.4	E	Papain	212	I	Stefin B	98
1F34	2.5	A	Pepsin A	325	B	Major pepsin inhibitor PI-3	138
1UBK	1.2	L	Periplasmic hydrogenase large subunit	534	S	Periplasmic hydrogenase small subunit	267
1JLT	1.4	B	Phospholipase A2	122	A	Phospholipase A2 inhibitor	122
1L4Z	2.3	A	Plasminogen	248	B	Streptokinase	125
1DHK	1.9	A	Porcine pancreatic alpha-amylase	495	B	Bean lectin-like inhibitor	195
3YGS	2.5	P	Procaspase 9	97	C	Apoptotic protease activating factor 1	95
1FT1	2.3	B	Protein farnesyltransferase	416	A	Protein farnesyltransferase	315
1G4U	2.3	S	Protein tyrosine phosphatase SPTP	360	R	Ras-related C3 botulinum toxin substrate 1	180
1CT4	1.6	E	Proteinase	185	I	Ovomucoid inhibitor	51
1VG0	2.2	A	Rab escort protein 1	481	B	Ras-related protein rab-7	182
1F2T	1.6	A	Rad50 abc-atpase N-terminal domain	145	B	Rad50 abc-atpase C-terminal domain	143
1GUA	2.0	A	Rap1A	167	B	C-raf1	76
1HE1	2.0	C	Ras-related C3 botulinum toxin substrate 1	176	A	Exoenzyme S	135
1DS6	2.4	A	Ras-related C3 botulinum toxin substrate 2	181	B	RHO GDP-dissociation inhibitor 2	179
1C1Y	1.9	A	Ras-related protein	167	B	Proto-onkogene serine	77
1DFJ	2.5	E	Ribonuclease A	124	I	Ribonuclease inhibitor	456
1DZB	2.0	A	SCFV fragment 1F9	224	X	Turkey egg-white lysozyme C	129
1H2S	1.9	A	Sensory rhodopsin II	225	B	Sensory rhodopsin II transducer	60
1P57	1.8	B	Serine protease hepsin heavy chain	247	A	Serine protease hepsin light chain	110
4SGB	2.1	E	Serine proteinase B	185	I	Potato inhibitor	51
1SMP	2.3	A	Serratia metallo proteinase	468	I	Erwinia chrysanthemi inhibitor	100
1NRJ	1.7	B	Signal recognition particle receptor	191	A	Docking protein	147
1RJ9	1.9	A	Signal recognition protein	277	B	Signal recognition particle protein	282
1JTP	1.9	A	Single-domain antibody	135	L	Lysozyme C	129
1SGD	1.8	E	Streptogrisin B	185	I	Ovomucoid	51
1LW6	1.5	E	Subtilisin BPN	281	I	Ubtilisin-chymotrypsin inhibitor-2A	63
2SIC	1.8	E	Subtilisin BPN	275	I	Streptomyces subtilisin inhibitor	107
1SPB	2.0	S	Subtilisin BPN	264	P	Subtilisin BPN prosegment	71
1R0R	1.1	E	Subtilisin carlsberg	274	I	Ovomucoid	51
1CSE	1.2	E	Subtilisin carlsberg	274	I	Eglin-C	63
1SCJ	2.0	A	Subtilisin E	275	B	Subtilisin E	71
2SNI	2.1	E	Subtilisin novo	275	I	Chymotrypsin inhibitor 2	64
1EUC	2.1	B	Succinyl-coa synthetase, beta chain	393	A	Succinyl-coa synthetase, alpha chain	306
1ONQ	2.2	A	T-cell surface glycoprotein CD1A	274	B	Beta-2-microglobulin	99
1JTD	2.3	A	Tem-1 beta-lactamase	262	B	Beta-lactamase inhibitor protein II	273
1KTZ	2.2	B	TGF-beta type II receptor	106	A	Transforming growth factor beta 3	82
2TEC	2.0	E	Thermitase	279	I	Eglin-C	63
1JKG	1.9	B	Tip associating protein	180	A	NTF2-related export protein 1	139
1D4V	2.2	B	TNF-related apoptosis inducing ligand	163	A	Death receptor 5	117
1AVW	1.8	A	Trypsin	223	B	Trypsin inhibitor	171
1BRB	2.1	E	Trypsin	223	I	BPTI	51
1F5R	1.7	A	Trypsin II	216	I	Pancreatic trypsin inhibitor	57
1K9O	2.3	E	Trypsin II anionic	223	I	Alaserpin	376
1D6R	2.3	A	Trypsinogen	223	I	Bowman-birk proteinase inhibitor	58

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

1OPH	2.3	B	Trypsinogen	223	A	Alpha-1 protease inhibitor	375
1P2J	1.4	A	Trypsinogen	220	I	Pancreatic trypsin inhibitor	56
1S1Q	2.0	A	Tumor susceptibility gene 101 protein	137	B	Ubiquitin	71
1ITB	2.5	B	Type 1 interleukin-1 receptor	310	A	Interleukin-1 beta	153
1J7D	1.9	B	Ubiquitin-conjugating enzyme E2-17 KDA	149	A	MMS2	140
1EUV	1.3	A	ULP1 protease	221	B	Ubitqutin-like protein SMT3	79
1UGH	1.9	E	Uracil-dna glycosylase	223	I	Uracil-DNA glycosylase inhibitor	82
1UZX	1.9	A	Vacuolar protein sorting-associated protein	135	B	Ubiquitin	75
1JTT	2.1	A	VH single-domain antibody	133	L	Lysozyme	129
1RKE	2.4	A	Vinculin	262	B	VCL protein	176
1MA9	2.4	A	Vitamin D-binding protein	442	B	Actin, alpha skeletal muscle	356
1YVN	2.1	A	Yeast actin	372	G	Gelsolin	125
1OXB	2.3	A	YDP1P	166	B	Osmolarity two-component system protein	124

Homodimers

PDB	Resolution (Å)	Name of homodimer	Scientific source	Chain one	Length	Chain two	Length
1CNZ	1.8	3-isopropylmalate dehydrogenase	Salmonella typhimurium	A	363	B	363
1AFW	1.8	3-ketoacetyl-coa thiolase	Saccharomyces cerevisiae	A	390	B	393
1M4I	1.5	Acetyltransferase	Escherichia coli	A	181	B	176
1LQ9	1.3	Actva-orf6 monooxygenase	Streptomyces coelicolor	A	112	B	112
1ADE	2	Adenylosucinate synthetase	Escherichia coli	A	431	B	431
1M7H	2	Adenylylsulfate kinase	Penicillium chrysogenum	A	203	B	200
1NA8	2.3	ADP-ribosylation binding protein	Homo sapiens	A	151	B	145
1OR4	2.2	Aerotactic transducer hemat	Bacillus subtilis	A	169	B	158
1BD0	1.6	Alanine racemase	Bacillus stearothermophilus	A	381	B	380
1A4U	1.9	Alcohol dehydrogenase	Drosophila lebanonensis	A	254	B	254
1ALK	2	Alkaline phosphatase	Escherichia coli	A	449	B	449
1LK9	1.5	Alliin lyase	Allium sativum	A	425	B	427
1HSS	2.1	Alpha-amylase inhibitor	Triticum aestivum	A	111	B	111
1S2Q	2.1	Amine oxidase B	Homo sapiens	A	499	B	494
1EKP	2.5	Amino acid aminotransferase	Homo sapiens	A	365	B	365
2GSA	2.4	Aminotransferase	Synechococcus SP	A	427	B	427
1DQT	2	Antigen	Mus musculus	A	117	B	117
1BJW	1.8	Aspartate aminotransferase	Thermus thermophilus	A	381	B	381
1JFL	1.9	Aspartate racemase	Escherichia coli	A	228	B	228
1MJH	1.7	Atp-binding protein	Methanococcus jannaschii	A	143	B	144
1IRI	2.4	Autocrine motility factor	Homo sapiens	A	557	B	557
1LR5	1.9	Auxin binding protein	Zea mays	A	160	B	160
1N80	2.5	Baseplate structural protein	Bacteriophage T4	A	328	B	328
1EWZ	2.4	Beta lactamase oxa-10	Pseudomonas aeruginosa	A	243	C	243
1EBL	1.8	Beta-ketoacyl-acp Synthase III	Escherichia coli	A	309	B	309
1N1B	2	Bornyl diphosphate synthase	Salvia officinalis	A	516	B	519
1KSO	1.7	Calcium-binding protein A3	Homo sapiens	A	93	B	93
1JD0	1.5	Carbonic anhydrase	Homo sapiens	A	260	B	259
1AUO	1.8	Carboxylesterase	Pseudomonas fluorescens	A	218	B	218
1CDC	2	CD2	Rattus norvegicus	A	96	B	96
1F13	2.1	Cellular coagulation factor	Homo sapiens	A	722	B	719
1NW1	2	Choline kinase	Caenorhabditis elegans	A	365	B	357

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

1R5P	2.2	Circadian oscillation regulator	Anabaena SP	A	90	B	93
1G64	2.1	Cob(I) alamin adenosyltransferase	Salmonella typhimurium	A	169	B	190
1OTV	2.1	Coenzyme pqq synthesis protein C	Klebsiella pneumoniae	A	254	B	254
1I0R	1.5	Conserved hypothetical protein	Archaeoglobus fulgidus	A	161	B	168
1OAC	2	Copper amine oxidase	Escherichia coli	A	719	B	722
1EAJ	1.4	Coxsackie virus	Homo sapiens	A	124	B	120
1CHM	1.9	Creatinase	Pseudomonas putida	A	401	B	401
1S44	1.6	Crustacyanin A1 subunit	Homarus gammarus	A	180	B	180
1GD7	2	CSAA protein	Thermus thermophilus	A	109	B	109
1L5B	2	Cyanovirin-N	Nostoc ellipsosporum	A	101	B	101
1SO2	2.4	Cyclic Phosphodiesterase B	Homo sapiens	A	363	B	363
1P3W	2.1	Cysteine desulfurase	Escherichia coli	A	385	B	385
1COZ	2	Cytidylyltransferase	Bacillus subtilis	A	126	B	126
1P6O	1.1	Cytosine deaminase	Saccharomyces cerevisiae	A	156	B	161
2DAB	2	D-amino acid aminotransferase	Thermophilic bacillus	A	280	B	282
1F17	2.3	Dehydrogenase	Homo sapiens	A	293	B	291
2NAC	1.8	Dehydrogenase	Methylotrophic bacterium pseudomonas	A	374	B	374
1NFZ	2	Delta-isomerase	Escherichia coli	A	176	B	180
1D1G	2.1	Dihydrofolate reductase	Thermotoga maritima	A	164	B	164
1DOR	2	Dihydroorotate dehydrogenase A	Lactococcus lactis	A	311	B	311
1AD1	2.2	Dihydropteroate synthetase	Staphylococcus aureus	A	264	B	251
1NU6	2.1	Dipeptidyl peptidase	Homo sapiens	A	728	B	728
1PE0	1.7	DJ-1	Homo sapiens	A	187	B	187
1G1A	2.5	DTDP-D-glucose 4,6-Dehydratase	Salmonella enterica	A	352	B	352
1BBH	1.8	Electron transport	Chromatium vinosum	A	131	B	131
1Q8R	1.9	Endodeoxyribonuclease rusa	Escherichia coli	A	118	B	109
1RVE	2.5	Endonuclease	Escherichia coli	A	244	B	244
1M9K	2	Endothelial nitric-oxide synthase	Homo sapiens	A	400	B	401
1P43	1.8	Enolase 1	Saccharomyces cerevisiae	A	436	B	436
1JR8	1.5	Erv2 protein mitochondrial	Saccharomyces cerevisiae	A	105	B	105
1V26	2.5	Fatty-acid-coa synthetase	Thermus thermophilus	A	489	B	510
1LBQ	2.4	Ferrochelatase	Saccharomyces cerevisiae	A	356	B	354
1RYA	1.3	Gdp-mannose mannosyl hydrolase	Escherichia coli	A	160	B	160
1QFH	2.2	Gelation factor	Dictyostelium discoideum	A	212	B	212
1JV3	2.2	Glcnac1p uridyltransferase	Homo sapiens	A	490	B	484
1DPG	2	Glucose 6-phosphate dehydrogenase	Leuconostoc mesenteroides	A	485	B	485
1QXR	1.7	Glucose-6-phosphate isomerase	Pyrococcus furiosus	A	187	B	187
1EOG	2.1	Glutathione S-transferase	Escherichia coli	A	208	B	208
1N2A	1.9	Glutathione S-transferase	Escherichia coli	A	201	B	187
1M0W	1.8	Glutathione synthetase	Saccharomyces cerevisiae	A	481	B	479
1R9C	1.8	Glutathione transferase	Mesorhizobium loti	A	125	B	118
1F4Q	1.9	Grancalcin	Homo sapiens	A	161	B	165
1DQP	1.8	Guanine phosphoribosyltransferase	Giardia lamblia	A	230	B	230
3SDH	1.4	Hemoglobin	Scapharca inaequivalvis	A	145	B	145
1IPI	2.2	Holliday junction resolvase	Pyrococcus furiosus	A	114	B	114
1FWL	2.3	Homoserine kinase	Methanococcus jannaschii	A	296	B	296
2HHM	2.1	Hydrolase	Homo sapiens	A	272	B	272
1PP2	2.5	Hydrolase	Crotalus atrox	R	122	L	122
1FJH	1.7	Hydroxysteroid dehydrogenase	Comamonas testosteroni	A	236	B	236

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

1G0S	1.9	Hypothetical Protein	Escherichia coli	A	201	B	202
1JOG	2.4	Hypothetical protein	Haemophilus influenzae	A	129	B	129
1PT5	2	Hypothetical protein	Escherichia coli	A	415	B	415
1QYA	2	Hypothetical Protein	Escherichia coli	A	293	B	307
1FUX	1.8	Hypothetical protein	Escherichia coli	A	164	B	163
1J30	1.7	Hypothetical rubrerythrin	Sulfolobus tokodaii	A	141	B	137
1LHZ	2.3	Immunoglobulin lambda	Homo sapiens	A	213	B	213
1AA7	2.1	Influenza virus matrix mrotein	Influenza virus	A	158	B	157
8PRK	1.9	Inorganic pyrophosphatase	Saccharomyces cerevisiae	A	282	B	282
1R8J	2	Kaia	Synechococcus elongatus	A	272	B	264
1CQS	1.9	Ketosteroid isomerase	Pseudomonas putida	A	124	B	124
1AQ6	2	L-2-haloacid dehalogenase	Xanthobacter autotrophicus	A	245	B	245
1I2W	1.7	Lactamase	Bacillus licheniformis	A	255	B	256
1BH5	2.2	Lactoylglutathione lyase	Homo sapiens	A	177	B	182
1QMJ	2.2	Lectin	Gallus gallus	A	132	B	132
1K75	1.8	L-histidinol dehydrogenase	Escherichia coli	A	425	B	425
1EHI	2.4	Ligase	Leuconostoc mesenteroides	A	360	B	347
1NWW	1.2	Limonene-1,2-epoxide hydrolase	Rhodococcus erythropolis	A	145	B	146
1UC8	2	Lysine biosynthesis enzyme	Thermus thermophilus	A	240	B	239
1EN5	2.3	Manganese superoxide dismutase	Escherichia coli	A	205	B	205
1A4I	1.5	Methylenetetrahydrofolate	Homo sapiens	A	285	B	295
1FC5	2.2	Molybdopterin biosynthesis	Escherichia coli	A	397	B	396
1JYS	1.9	Mta/sah nucleosidase	Escherichia coli	A	226	B	226
1LNW	2.1	Multidrug resistance operon repressor	Pseudomonas aeruginosa	A	137	B	135
1FP3	2	N-acyl-d-glucosamine	Sus scrofa	A	402	B	402
1FYD	2.3	NAD(+) Synthetase	Bacillus subtilis	A	271	B	246
1HJ3	1.6	Nitrite reductase	Paracoccus pantotrophus	A	544	B	542
1G1M	2.3	Nitrogenase iron protein	Azotobacter vinelandii	A	287	B	289
1G8T	1.1	Nuclease SM2 isoform	Seratia marcencsens	A	241	B	241
1EYV	1.6	N-utilizing substance protein	Mycobacterium tuberculosis	A	131	B	133
1M98	2.1	Orange carotenoid protein	Arthrospira maxima	A	316	B	314
1ORO	2.4	Orotate phosphoribosyltransferase	Escherichia coli	A	213	B	206
1DVJ	1.5	Orotidine 5'-phosphate decarboxylase	Methanobacterium thermoautotrophicum	A	239	B	211
1GGQ	2.5	Outer surface protein C	Borrelia burgdorferi	A	162	B	162
1AOR	2.3	Oxidoreductase	Pyrococcus furiosis	A	605	B	605
1BMD	1.9	Oxidoreductase	Thermus flavus	A	327	B	327
1HDY	2.5	Oxidoreductase	Homo sapiens	A	374	B	374
1N2O	2.1	Pantothenate synthetase	Mycobacterium tuberculosis	A	279	B	279
1RN5	2.2	Peptide deformylase	Leptospira interrogans	A	177	B	177
1PN2	2	Peroxisomal hydratase	Candida tropicalis	A	269	B	267
1PN0	1.7	Phenol 2-monooxygenase	Trichosporon cutaneum	A	652	C	656
1BXG	2.3	Phenylalanine dehydrogenase	Rhodococcus SP	A	349	B	347
1M6P	1.8	Phosphate receptor	Bos Taurus	A	146	B	146
1RQL	2.4	Phosphonoacetaldehyde hydrolase	Bacillus cereus	A	257	B	257
1O4U	2.5	Phosphoribosyltransferase	Thermotoga maritima	A	265	B	266
1EZ2	1.9	Phosphotriesterase	Pseudomonas diminuta	A	328	B	328
1EXQ	1.6	Pol polyprotein	Escherichia coli	A	147	B	145
1MNA	1.8	Polyketide synthase	Streptomyces venezuelae	A	276	B	278
1C6X	2.5	Protease	Escherichia coli	A	99	B	99

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

1FL1	2.2	Protease	Escherichia coli	A	192	B	207
1F89	2.4	Protein YLC351C	Saccharomyces cerevisiae	A	271	B	271
1LHP	2.1	Pyridoxal kinase	Ovis aries	A	306	B	309
1CBK	2	Pyrophosphokinase	Haemophilus influenzae	A	160	B	160
1QR2	2.1	Quinone reductase type 2	Homo sapiens	A	230	B	230
1EN7	2.4	Recombination endonuclease	Bacteriophage T4	A	157	B	157
1EV7	2.4	Restriction endonuclease naei	Nocardia aerocolonigenes	A	295	B	293
1H8X	2	Ribonuclease	Homo sapiens	A	125	B	125
1I4S	2.2	Ribonuclease III	Aquifex aeolicus	A	147	B	147
1KGN	1.9	Ribonucleotide reductase protein	Corynebacterium ammoniagenes	A	296	B	296
1TLU	1.6	S-adenosylmethionine decarboxylase	Thermotoga maritima	A	117	B	117
1K6Z	2	Secretion chaperone syce	Yersinia pestis	A	120	B	119
1K3S	1.9	Sige	Salmonella enterica	A	106	B	104
1PJQ	2.2	Siroheme synthase	Salmonella typhimurium	A	447	B	454
1HJR	2.5	Site-specific recombinase	Escherichia coli	A	158	C	158
3LYN	1.7	Sperm lysine	Haliotis fulgens	A	122	B	124
2SQC	2	Squalene-hopene Cyclase	Alicyclobacillus acidocaldarius	A	623	B	623
1SCF	2.2	Stem cell factor	Homo sapiens	A	116	B	118
1OX8	2.2	Stringent starvation protein B	Escherichia coli	A	105	B	105
1M3E	2.5	Succinyl-coa	Sus scrofa	A	459	B	460
1R7A	1.8	Sucrose phosphorylase	Bifidobacterium adolescentis	A	503	B	503
1SOX	1.9	Sulfite oxidase	Gallus gallus	A	463	B	458
1L5X	2	Survival protein E	Pyrobaculum aerophilum	A	270	B	272
1REG	1.9	T4 rega	Bacteriophage T4	X	122	Y	120
1MKB	2	Thiol ester dehydrase	Escherichia coli	A	171	B	171
1QHI	1.9	Thymidine kinase	Herpes simplex virus	A	304	B	308
1HSJ	2.3	Transcription/sugar binding protein	Escherichia coli	A	487	B	487
1NY5	2.4	Transcriptional regulator	Aquifex aeolicus	A	384	B	385
1ON2	1.6	Transcriptional regulator	Bacillus subtilis	A	135	B	135
1SMT	2.2	Transcriptional repressor	Synechococcus	A	98	B	101
1TRK	2	Transferase	Saccharomyces cerevisiae	A	678	B	678
7AAT	1.9	Transferase	Gallus gallus	A	401	B	401
1KIY	2.4	Trichodiene synthase	Fusarium sporotrichioides	A	354	B	354
1I8T	2.4	Udp-galactopyranose mutase	Escherichia coli	A	367	B	367
1F6D	2.5	Udp-n-acetylglucosamine	Escherichia coli	A	366	B	363
1JP3	1.8	Undecaprenyl pyrophosphate synthase	Escherichia coli	A	210	B	207
1JMV	1.9	Universal stress protein A	Haemophilus influenzae	A	140	B	137
1HQO	2.3	URE2 protein	Saccharomyces cerevisiae	A	221	B	217
9WGA	1.8	Wheat germ agglutinin	Triticum vulgaris	A	170	B	170
1MI3	1.8	Xylose reductase	Candida tenuis	A	319	B	319

Table 1: Dataset Creation

Interface H-bonds:

Intermolecular hydrogen bonds between subunits are important in the association and stability of protein-protein interfaces. **[3, 4]** H-bonds in homodimers (range 0 - 51) and heterodimers (range 0 - 98) are different. The mean H-bonds are larger for homodimers (mean = 18) than heterodimers (mean = 12). Figure 1 A and B show that there is a high correlation between H-bonds and interface residues. The correlation coefficient is 0.83 in heterodimers and 0.85 in homodimers. This is similar to the previous reports in the range of 0.75 and 0.89. **[2, 4, 7, 8, 11]** However, there is a subtle difference with the previous studies and the variation is affected by structure resolution, dataset size and data type. The dataset used in this study contains structures with resolution $\leq$ 2.5Å and the data is either exclusively homodimer or heterodimer. However, previous datasets

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

contain structures with resolution $\leq$ 3.0Å and the data is a mixture of heterodimers, homodimers and other oligomers. At low resolution there are fewer H-bonds and the correlation with interface area decreases. **[4]** Here, we show that the relation between H-bonds and interface residues is highly correlated for both heterodimers and homodimers. This is useful to evaluate inter-subunit H-bonds prediction and their involvement in interface stability. On average there are 0.24 H-bonds per interface residue in heterodimers and 0.22 H-bonds per interface residue in homodimer. The maximum number of H-bonds per interface residue is 0.65 in heterodimers and 0.44 in homodimers. Although there are more intermolecular H-bonds in homodimers, the density of H-bonds per interface residue is lower in homodimers than in heterodimers. **[7]**

Interface residues:

The number of interface residues is proportional to interface area. **[5,7]** Stronger protein subunit associations were generally associated with larger interface areas. **[11]** In our study, the range of heterodimer interface residues varies from 18 to 162 with a mean value of 51. While, the range of homodimer interface residues extends from 15 to 308 with a mean value of 81. Like H-bonds, interface residues also varied with different studies and are affected by dataset size and data type. **[5, 7, 9, 14]** Hence, we created mutually exclusive datasets of homodimers and heterodimers for this analysis to reduce bias due to data type heterogeneity. Thus, we show that the amount of interface residues is significantly different for homodimers and heterodimers. The results also suggest that the previous studies are based on datasets biased with heterodimers. The relation between number of interface residues and monomer length is shown in Figure 1 E and F. They show that interface residues increase with both heterodimer and homodimer monomer length. However, the relation is causal. Figure 1 C and D show a causal relationship between interface area and monomer length for both homodimers and heterodimers. The mean interface residues are larger in homodimers than heterodimers. This is consistent with previous studies. **[7, 9]**

Interface residue composition:

Several studies show the prevalence of certain types of residues at the dimer interfaces. **[4,6,7,12,13]** However, the significance of hydrophobic, hydrophilic, and charged residues at the interface of homodimers and heterodimers is not well documented. Figure 1G show the fractional distribution of hydrophobic, hydrophilic and charged residues in homodimer and heterodimer interfaces. Hydrophobic residues (M, F, P, A, B, L), except for I and G are dominant in homodimer interfaces. However, hydrophilic residues (W, C, H, Q, N, Y, S), except for T, are dominant in heterodimer interfaces. This observation is interesting and not surprising because homodimers being made of identical monomer subunits tend to associate by hydrophobic interactions. This is in contrast to the observation in heterodimer interfaces being made of non-identical monomer subunits, associating generally by hydrophilic interactions.

Figure 1H, shows the ration of interface/surface and interface/interior residue propensity difference between heterodimers and homodimers. Interestingly, the ratio of interface to interior charged residues (D, E, K, R) is significantly larger in heterodimers compared to homodimers (Figures 1H, 1I, 1J). On the other hand, the ratio of interface to interior hydrophobic residues (A, V, L, M, I, F) are prevalent in homodimers than in heterodimers (Figures 1H, 1I, 1J). Similarly, hydrophilic residues (N, Q, H, Y, S, T) are prevalent in heterodimer interfaces (Figures 1H, 1I, 1J). However, the propensity difference in the ratio of interface to surface hydrophobic/hydrophilic/charged for homodimers and heterodimers is almost zero (Figure 1H).

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

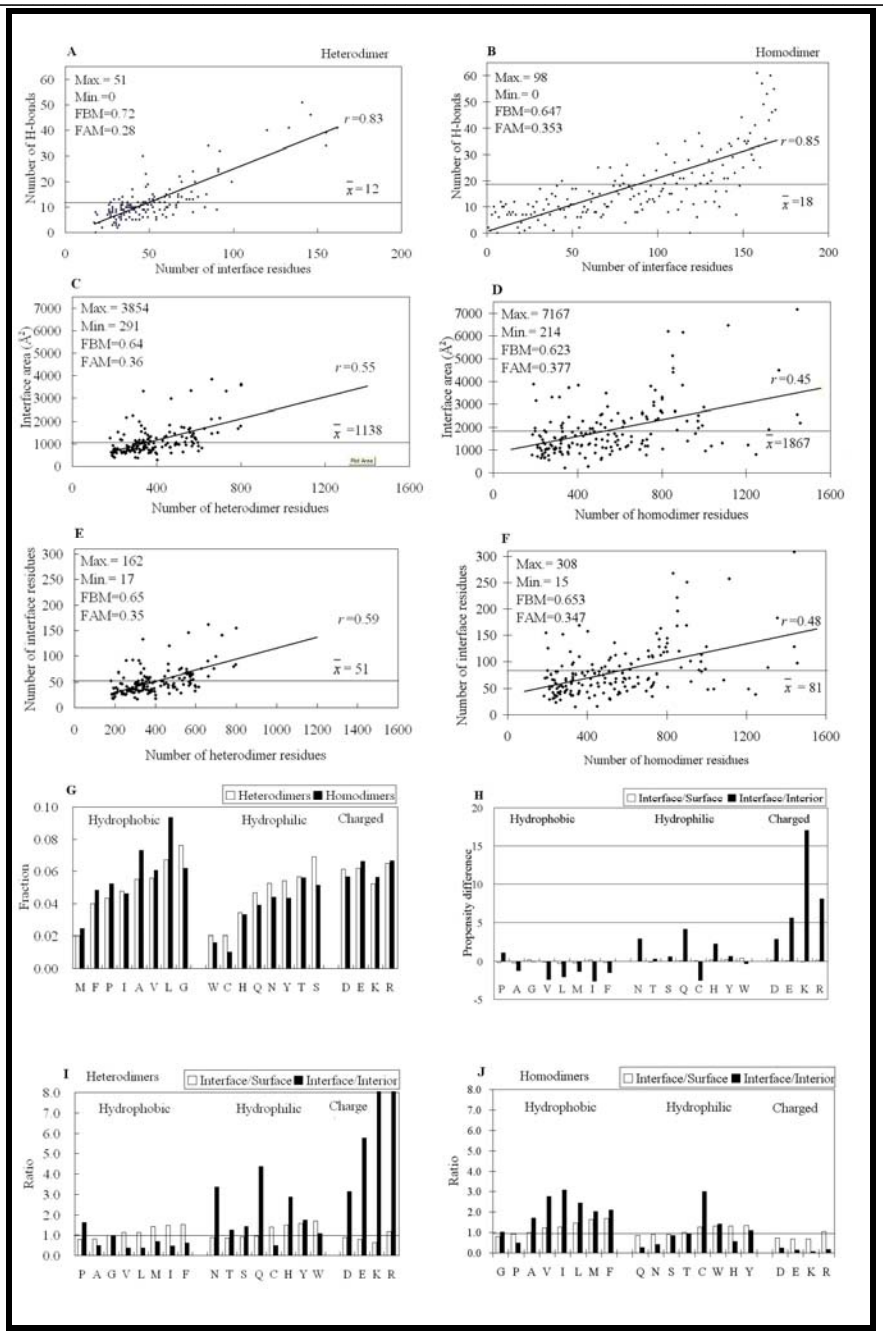

Figure 1: Difference between heterodimer and homodimer interface properties is shown
(A) Hydrogen bonds in heterodimer interface; (B) Hydrogen bonds in homodimer interface; (C) Interface area in heterodimer interface; (D) Interface area in homodimer interface; (E) Interface residues in heterodimers; (F) Interface residues in homodimers; (G) Hydrophobic, hydrophilic and charged residue fraction in heterodimers and homodimers; (H) Propensity difference in heterodimers and homodimers (heterodimers – homodimers); (I) Ratio of interface to surface & interface to interior propensity in heterodimers; (J) Ratio of interface to surface & interface to interior propensity in homodimers. FBM = Fraction below mean value; FAM = Fraction above mean value.

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Conclusion:
We performed a comprehensive analysis on the differences between 156 heterodimers and 170 homodimers. The homodimer and heterodimer datasets are mutually exclusive and is one of the unique features of the analysis. The analysis documents the differences between homodimer and heterodimer interfaces for the first time in a comprehensive manner. Homodimer interfaces have greater number of interface residues and H-bonds on average. However, the density of H-bonds per residue is greater for heterodimer interfaces. The study also shows that charged residues (D, E, K, R) and hydrophilic residues (N, T, S, Q, H, W, Y) are dominant at the heterodimer interfaces. Nonetheless, hydrophobic residues (A, V, L, M, I, F) are predominant at the homodimer interfaces. These data find utility in the development of independent models for the prediction of homodimer and heterodimer interaction sites.

Acknowledgement:
Cui Zhanhua acknowledges Nanyang Technological University, Singapore for support.

References:
[1] S. Jones & J. M. Thornton, *Proc. Natl. Acad. Sci.,* 93:13 (1996) [PMID: 8552589]
[2] D. Xu, *et al., Protein Engineering,* 10:999 (1997) [PMID: 9464564]
[3] C. J. Tsai, *et al.*, *Protein Science,* 6:53 (1997) [PMID: 9007976]
[4] L. Lo Conte, *et al., J. Mol. Biol.,* 285:2177 (1999) [PMID: 9925793]
[5] P. Chakrabarti & J. Janin, *Proteins,* 47:334 (2002) [PMID: 11948787]
[6] K. V. Brinda, *et al., Protein Eng.,* 15:265 (2002) [PMID: 11983927]
[7] R. P. Bahadur, *et al.*, *Proteins,* 53:708 (2003) [PMID: 14579361]
[8] I. M. Nooren & J. .M. Thornton, *J. Mol. Biol.*, 325:991 (2003) [PMID: 12527304]
[9] D. R. Caffrey, *et al., Protein Sci.*, 13:190 (2004) [PMID: 14691234]
[10] C. Zhanhua, *et al., Frontiers in Bioscience*, 10:844 (2005) [PMID: 15569623]
[11] S. Jones & J. M. Thornton, *Prog. Biophys. Mol. Biol.,* 63:31 (1995) [PMID: 7746868]
[12] P. Lijnzaad & P. Argos, *Proteins*, 28:333 (1997) [PMID: 9223180]
[13] S. Dasgupta, *et al., Proteins*, 28:494 (1997) [PMID: 9261866]
[14] W. S. Valdar & J. M. Thornton, *J Mol Biol.*, 313:399 (2001) [PMID: 11800565]
[15] L. Li, *et al., Frontiers in Bioscience*, 10:1977 (2005) [PMID: 15769678]
[16] B. Lee & E. M. Richard, *J. Mol. Biol.,* 55:379 (1971) [PMID: 5551392]
[17] R. Rodriguez, *et al.*, *Bioinformatics,* 14**:**523 (1998) [PMID: 9694991]
[18] S. Jones & J. M. Thornton, *J. Mol. Biol.,* 272:121 (1997) [PMID: 9299342]

Edited by K. Gunasekaran
Citation: Zhanhua *et al*., Bioinformation 1(2): 28-39 (2005)

ISSN 0973-2063
Bioinformation 1(2): 28-39 (2005)