Supporting Online Material

For

Stability of domain structures in multi-domain proteins

Ramachandra M. Bhaskara[1] & N. Srinivasan[1*]

[1]*Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India*.

***To whom correspondence should be addressed. E-mail: ns@mbu.iisc.ernet.in (N.S.)**

This PDF file includes:

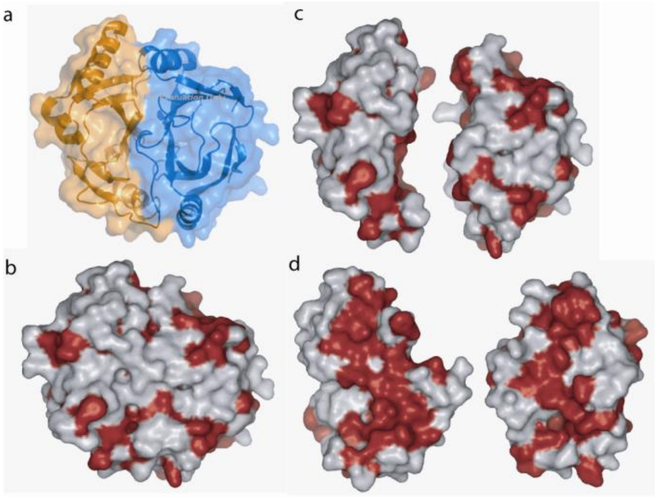
Supplementary Figure(S) S1-S4 with legends

Supplementary Table S1

Supplementary Dataset(s)

Supplementary References

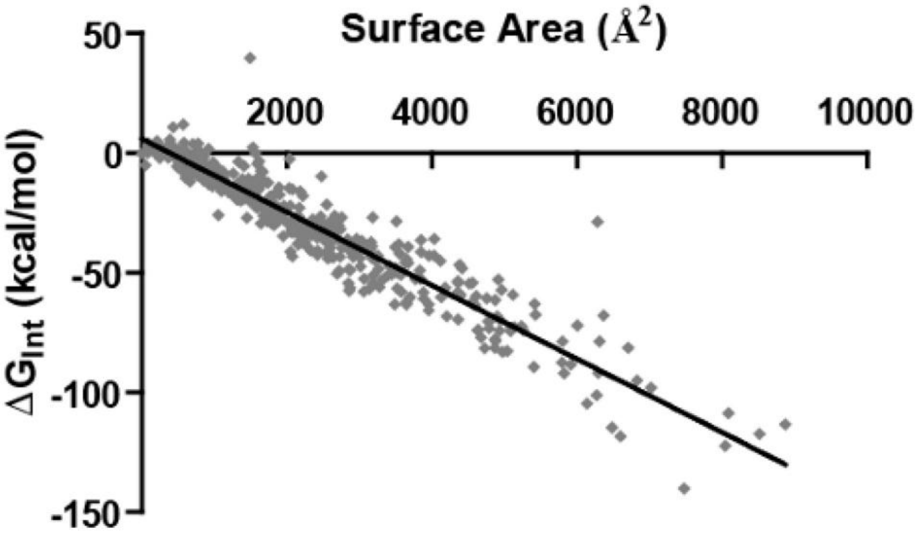
Supplementary Figure(s).



Supplementary Figure-S1 Srinivasan

Supplementary Figure S1

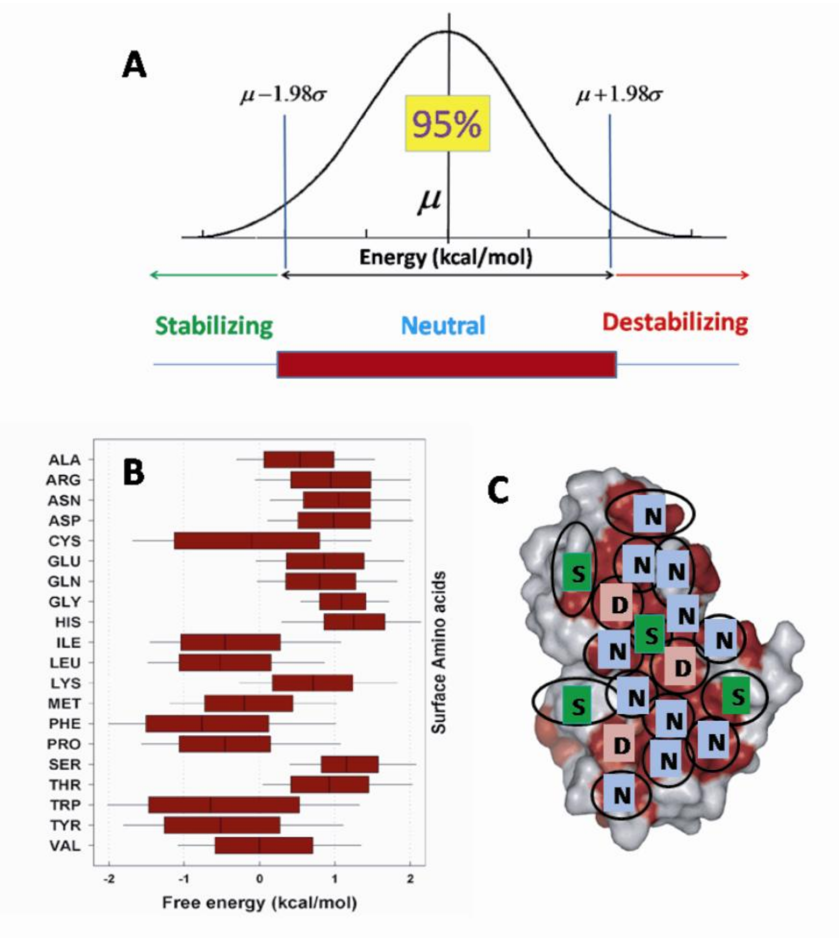
Stability and folding of a representative multi-domain enzyme (Dioxygenase, Bhpc). a. Ribbon diagram showing the topology of the two interacting domains. **b.** Surface representation of the same showing the apolar residues (Red). **c.** The artificial (in silico) parsing of the full length protein into two isolated domains. These isolated domains were used for free energy computations. Their stabilities are -16 kcal/mol (Domain 1) and 2 kcal/mol (Domain 2). **d.** 90^o rotation of the two domains showing the exposed hydrophobic patch. Hydrophobic bonding at the domain-domain interface holds the domains together. Apolar solvation energy of -54 kcal/mol contributes maximally towards the overall stability of -80 Kcal/mol.



Supplementary Figure-S2 Srinivasan

Supplementary Figure S2

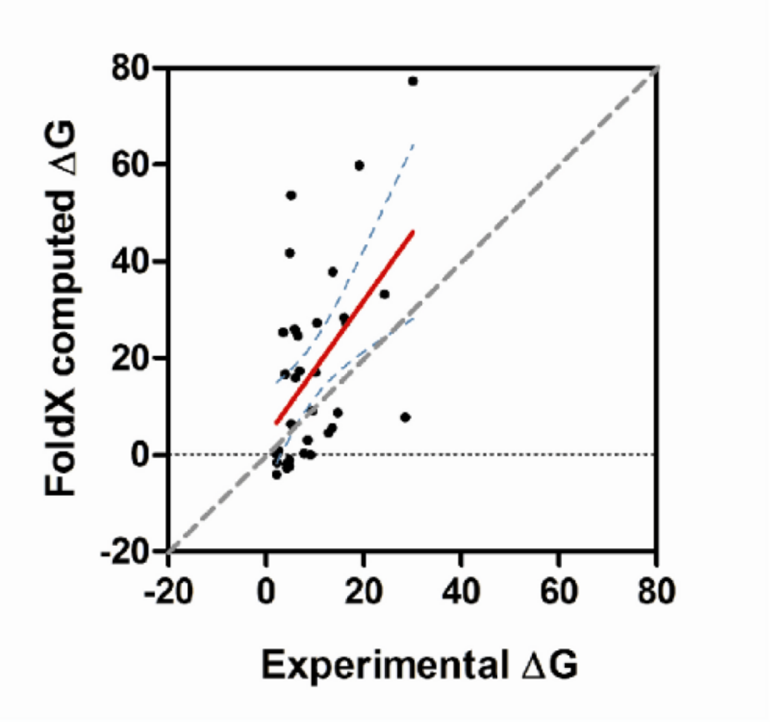
Interaction free energy as a function of interacting surface area. Scatter plot showing the linear relationship between the surface area of interaction and the FoldX computed interaction free energy (ΔG_{Int}) for a set of 268 obligate protein-protein complexes. Red line shows the best fit line in linear regression analysis: $r^2 = 0.6971$; $P < 0.0001$; $\Delta G_{\mathrm{Int}} = -0.01274 * SA + 1.426$.



Supplenentry Figure-S3 Srinivasan

Supplementary Figure S3

Schematic representation of classification of domain-domain interfacial residues. a. Computation of distribution of the free energy contribution of a typical surface residue of a single residue type. e.g. Alanine Showing the cutoff limits for the classification into the three categories. **b.** The computation of the distribution parameters (μ and s.d.) for all the 20 amino acids. The red bar shows the neutral zone with the mean at the center for the free energy contribution of surface residues. The lines extending the distribution are beyond the 95% confidence limits. **c.** Each interfacial residue from the MD_doms dataset was compared against the B and classified into one of the three categories. Shown is a representative example of interface of domain-1 of the BhpC enzyme. The Stabilizing residues are marked by the green squares with "S", the neutral residues are marked by blue squares with "N" and the destabilizing residues are marked by light red squares with "D".



Supplementary Figure-S4 Srinivasan

Supplementary Figure S4

Benchmarking FoldX computations. Correlation between experimentally determined and FoldX computed free energies of unfolding for a set of 35 proteins. Spearman rank correlation r = 0.5679; *P* = 0.0004. Red line represents best fit by linear regression. Blue dashed line indicates the 95% confidence limit. These proteins were selected as they were monomeric and showed two state unfolding at 298K and pH 7.0. The experimental unfolding free energies were obtained from Protherm database. The correlation shows that the FoldX computed values cannot be taken as absolute but are good for relative comparisons.

Supplementary Table S1 Srinivasan.

Supplementary Table S1. Natural variations (nsSNPs ; *n* = 33) in the interfaces of multi-domain proteins. The following table lists the nsSNPs found in the domain-domain interface of human proteins (*n* = 14). The improper functioning of these proteins have been implicated in various disease states. The effect of the nsSNP on the stability of the protein is computed by modelling the structure of the variant amino-acid and computing the ΔΔG using FoldX. The relative conservation of the site, the solvent accessibility and the functional relavence (Ligand binding) are also provided. The direct involvement of the nsSNP (*n* = 9) in prodiucing a non-functional protein and leading to disease state is shown by refering to pubmed id (PMID). The bold SNPs refer to change of a non-polar interfacial residue to a polar residue. The bold ΔΔG values (>1 kcal/mol) correspond to considerable destabilization of the structure to affect function. Note that the SNPs at residues participating in stabilization of the interface (See Supplementary Fig. S3online) lead to a larger destabilization effect.

Protein	PDB	SNPdb id	SNP	Solv Ac (%)	Conservation (%)	ΔΔG (kcal/ mol)	Ligand Binding	nsSNP Validation /PMID	Diseases associated with
hematopoietic prostaglandin D synthase	1iyh	rs56161897	F151L	0	0.16	**1.55**	---	----	Delayed type hypersensitivity
immunoglobulin lambda-like polypeptide 1	1l6x	rs11557948	Y436C	0.19	0.01	**1.25**	---	----	Agammaglobulinemia 2
immunoglobulin lambda-like polypeptide 1	1l6x	rs15893	E430K	0.15	0.18	-1.47	---	----	Agammaglobulinemia 2
sex comb on midleg-like 2 (Drosophila)	1oi1	rs73454384	I224V	0	0.66	0.83	---	----	Nil
tocopherol (alpha)	1oip	rs28936369	R192H	0.31	0.12	-0.5	---	8602747	Ataxia, Friedreich-like, with Isolated

transfer protein								(*1*)	vitamin E defieciency
FK506 binding protein 4	1q1c	rs1062478	R157H	0.42	0.51	-0.09	---	----	Associated with Refsum disease
acyl-CoA dehydrogenase family, member 8	1rx0	rs11550637	F87L	0	0.33	**1.35**	---	----	isobutyryl-CoA dehydrogenase (IBD) deficiency
acyl-CoA dehydrogenase family, member 8	1rx0	rs7483314	L375V	0.28	0.07	**1.02**	FAD	----	isobutyryl-CoA dehydrogenase (IBD) deficiency
breast cancer 1, early onset	1t15	rs1800757	**P1776S**	0.54	0.15	**1.1**	---	17063491 (*2*), 17719744 (*3*)	Breast-Ovarian cancer, familial, suseptibility
breast cancer 1, early onset	1t15	rs28897696	**A1708E**	0	0.19	**11.56**	---	17063491 (*2*), 17719744 (*3*)	Breast-Ovarian cancer, familial, suseptibility
breast cancer 1, early onset	1t15	rs41293459	**R1699Q**	0	0.11	**2.09**	---	17719744 (*3*)	Breast-Ovarian cancer, familial, suseptibility
breast cancer 1, early onset	1t15	rs41293463	**M1775R**	0.02	0.21	**4.41**	---	17063491 (*2*), 17719744 (*3*)	Breast-Ovarian cancer, familial, suseptibility
breast cancer 1, early onset	1t15	rs55770810	**R1669W**	0	0.11	**2.65**	---	17719744 (*3*)	Breast-Ovarian cancer, familial, suseptibility
breast cancer 1, early	1t15	rs55808233	**M1783T**	0	0.41	**4.01**	---	17719744	Breast-Ovarian cancer, familial,

onset								(*3*)	suseptibility
polo-like kinase 1 (Drosophila)	1umw	rs34359215	R581C	0.47	0.33	0.45	---	----	Cell cycle regulation and Cancer
polo-like kinase 1 (Drosophila)	1umw	rs35347525	R441H	0.19	0.41	**1.61**	---	----	Cell cycle regulation and Cancer
glutathione S-transferase omega 1	1eem	rs11509436	S86C	0.04	0.05	-0.34	GSH	----	Cancer, Xenobiotics
glutathione S-transferase omega 1	1eem	rs11509439	A236V	0	0.01	**4.85**	---	----	Cancer, Xenobiotics
glutathione S-transferase omega 1	1eem	rs15032	T217N	0.24	0.03	0.12	---	----	Cancer, Xenobiotics
glutathione S-transferase omega 1	1eem	rs45529437	**C32Y**	0.11	0.18	-1.06	GSH	----	Cancer, Xenobiotics
glutathione S-transferase omega 1	1eem	rs72559704	R7K	0.58	0	-0.08	---	----	Cancer, Xenobiotics
hydroxyacyl-CoA dehydrogenase	1f0y	rs1051519	Q140H	0.53	0.09	0.48	---	----	Hyperinsulinemic Hypoglycemia, familial
Fc fragment-IgG,low affinity IIa,recep(CD32)	1fcg	rs4986941	M107V	0.54	0.18	0.54	---	----	Lupus Nephritis
Fc fragment-IgG,low affinity IIIb,recep(CD16b)	1fnl	rs2290834	I88V	0.72	0.63	0.2	---	----	Susseptibility to recurrent viral infections
Fc fragment-IgG,low affinity	1fnl	rs403016	S18R	0.02	0.04	-1.9	---	2478590 (*4*)	Lupus Nephritis, Inflammatory Bowel Disease

IIIb,recep(CD16b)									
glutathione S-transferase alpha 1	1gse	rs1051558	**F10S**	0.07	0.25	**1.89**	EAA	----	Warfarin sensitivity
glutathione S-transferase alpha 1	1gse	rs1051566	**A12T**	0.08	0.06	0.95	EAA	----	Warfarin sensitivity
glutathione S-transferase alpha 1	1gse	rs1051573	A12V	0.08	0.06	**1.26**	EAA	----	Warfarin sensitivity
glutathione S-transferase alpha 1	1gse	rs1051983	**A216S**	0.02	0.02	0.67	EAA	----	Warfarin sensitivity
glutathione S-transferase alpha 1	1gse	rs1131849	R89K	0.08	0.43	-0.27	---	----	Warfarin sensitivity
glutathione S-transferase pi 1	11gs	rs45549733	R186W	0.29	0.14	0.74	---	----	Chemotherapy-induced carcinogenesis, acute myeloid leukemia
glutathione S-transferase pi 1	11gs	rs1695	I104V	0.28	0.1	0.76	EAA	9092542 (*5*)	Cancer, Xenobiotics
glutathione S-transferase pi 1	11gs	rs71534294	D157H	0.02	0.35	-0.18	---	----	Cancer, Xenobiotics

Supplementary Datasets:

Datasets used in the current analysis of Single and Multi-domain proteins.

MD_doms (856) : Astral domains of Multi-domain proteins; *SCOP version 1.75*

**d1lgsa1, d1lgsa2, d1a4ib1, d1a4ib2, d1a8d_1, d1a8d_2, d1a8h_1, d1a8h_2,
d1a8l_1, d1a8l_2, d1a8p_1, d1a8p_2, d1abwa1, d1abwa2, d1af0a1, d1af0a2,
d1af7_1, d1af7_2, d1afal1, d1afal2, d1afwb1, d1afwb2, d1akl_1, d1akl_2,
d1amm_1, d1amm_2, d1aqh_1, d1aqh_2, d1aqkl1, d1aqkl2, d1avaa1, d1avaa2,
d1aw7a1, d1aw7a2, d1b0na1, d1b0na2, d1b1ba1, d1b1ba2, d1b2ra1, d1b2ra2,
d1b3bc1, d1b3bc2, d1b43a1, d1b43a2, d1b71a1, d1b71a2, d1b8aa1, d1b8aa2,
d1b8pa1, d1b8pa2, d1b9ka1, d1b9ka2, d1b9wa1, d1b9wa2, d1bg0_1, d1bg0_2,
d1bgva1, d1bgva2, d1bhta1, d1bhta2, d1bif_1, d1bif_2, d1bli_1, d1bli_2,
d1bqub1, d1bqub2, d1bw9a1, d1bw9a2, d1c1aa1, d1c1aa2, d1c1el1, d1c1el2,
d1c2aa1, d1c2aa2, d1c5ka1, d1c5ka2, d1c8ua1, d1c8ua2, d1c9ya1, d1c9ya2,
d1ca1_1, d1ca1_2, d1cb6a1, d1cb6a2, d1ccza1, d1ccza2, d1cdy_1, d1cdy_2,
d1cf9a1, d1cf9a2, d1cfb_1, d1cfb_2, d1chma1, d1chma2, d1cjdb1, d1cjdb2,
d1clc_1, d1clc_2, d1coja1, d1coja2, d1cpo_1, d1cpo_2, d1cuna1, d1cuna2,
d1cvra1, d1cvra2, d1czya1, d1czya2, d1d2oa1, d1d2oa2, d1dd3a1, d1dd3a2,
d1dfx_1, d1dfx_2, d1dhka1, d1dhka2, d1dkxa1, d1dkxa2, d1dl5a1, d1dl5a2,
d1dmr_1, d1dmr_2, d1duga1, d1duga2, d1duvg1, d1duvg2, d1duza1, d1duza2,
d1dvpa1, d1dvpa2, d1dxrh1, d1dxrh2, d1dyka1, d1dyka2, d1dyqa1, d1dyqa2,
d1dzfa1, d1dzfa2, d1e1ca1, d1e1ca2, d1e1cb1, d1e1cb2, d1e3oc1, d1e3oc2,
d1e42a1, d1e42a2, d1e5ta1, d1e5ta2, d1e7la1, d1e7la2, d1eara1, d1eara2,
d1eb0a1, d1eb0a2, d1ebha1, d1ebha2, d1ec9d1, d1ec9d2, d1ecfa1, d1ecfa2,
d1ecfb1, d1ecfb2, d1edha1, d1edha2, d1eeja1, d1eeja2, d1eema1, d1eema2,
d1eerb1, d1eerb2, d1eg9a1, d1eg9a2, d1ehda1, d1ehda2, d1eila1, d1eila2,
d1ek4a1, d1ek4a2, d1ekxa1, d1ekxa2, d1emd_1, d1emd_2, d1enfa1, d1enfa2,
d1epfa1, d1epfa2, d1epfb1, d1epfb2, d1esl_1, d1esl_2, d1eu3a1, d1eu3a2,
d1evka1, d1evka2, d1ewca1, d1ewca2, d1ewfa1, d1ewfa2, d1f0ya1, d1f0ya2,
d1f1bb1, d1f1bb2, d1f1ha1, d1f1ha2, d1f1ra1, d1f1ra2, d1f1ua1, d1f1ua2,
d1f20a1, d1f20a2, d1f8ab1, d1f8ab2, d1fcga1, d1fcga2, d1fnla1, d1fnla2,
d1fnua1, d1fnua2, d1fp2a1, d1fp2a2, d1fpoa1, d1fpoa2, d1ftra1, d1ftra2,
d1fv1b1, d1fv1b2, d1fw1a1, d1fw1a2, d1fxla1, d1fxla2, d1g1qa1, d1g1qa2,
d1g38a1, d1g38a2, d1g5aa1, d1g5aa2, d1g6la1, d1g6la2, d1g87b1, d1g87b2,
d1g8ka1, d1g8ka2, d1g97a1, d1g97a2, d1giqa1, d1giqa2, d1gl4a1, d1gl4a2,
d1glpa1, d1glpa2, d1gmua1, d1gmua2, d1gmuc1, d1gmuc2, d1gqia1, d1gqia2,
d1gqza1, d1gqza2, d1gsa_1, d1gsa_2, d1gsea1, d1gsea2, d1gsma1, d1gsma2,
d1gupa1, d1gupa2, d1gv2a1, d1gv2a2, d1gv3a1, d1gv3a2, d1h03p1, d1h03p2,
d1h1oa1, d1h1oa2, d1h1pb1, d1h1pb2, d1h6ta1, d1h6ta2, d1h72c1, d1h72c2,
d1h8pa1, d1h8pa2, d1h9ra1, d1h9ra2, d1hbmb1, d1hbmb2, d1hf8a1, d1hf8a2,
d1hfel1, d1hfel2, d1hm9a1, d1hm9a2, d1hna_1, d1hna_2, d1hpua1, d1hpua2,
d1ht6a1, d1ht6a2, d1hw1a1, d1hw1a2, d1hw5a1, d1hw5a2, d1hx6b1, d1hx6b2,
d1hyob1, d1hyob2, d1i1aa1, d1i1aa2, d1i1ga1, d1i1ga2, d1i7na1, d1i7na2,
d1i7pa1, d1i7pa2, d1i8da1, d1i8da2, d1ig0a1, d1ig0a2, d1ig0b1, d1ig0b2,
d1ig3a1, d1ig3a2, d1ihga1, d1ihga2, d1ii2a1, d1ii2a2, d1iow_1, d1iow_2,
d1iqca1, d1iqca2, d1isza1, d1isza2, d1itkb1, d1itkb2, d1iyha1, d1iyha2,
d1iz4a1, d1iz4a2, d1iz6a1, d1iz6a2, d1iz9a1, d1iz9a2, d1j09a1, d1j09a2,**

d1j1la1, d1j1la2, d1j3na1, d1j3na2, d1j8mf1, d1j8mf2, d1jaka1, d1jaka2,
d1jdc_1, d1jdc_2, d1jfla1, d1jfla2, d1jhda1, d1jhda2, d1jhfa1, d1jhfa2,
d1jlva1, d1jlva2, d1jpna1, d1jpna2, d1js1x1, d1js1x2, d1js4a1, d1js4a2,
d1ju3a1, d1ju3a2, d1k0ma1, d1k0ma2, d1k0mb1, d1k0mb2, d1k8kd1, d1k8kd2,
d1k8wa3, d1k8wa5, d1k92a1, d1k92a2, d1kbva1, d1kbva2, d1khia1, d1khia2,
d1kkcx1, d1kkcx2, d1klua1, d1klua2, d1klub1, d1klub2, d1kqfb1, d1kqfb2,
d1ks9a1, d1ks9a2, d1ktba1, d1ktba2, d1kv9a1, d1kv9a2, d1kwma1, d1kwma2,
d1ky6a1, d1ky6a2, d1kzqa1, d1kzqa2, d1l1fb1, d1l1fb2, d1l3ka1, d1l3ka2,
d1l3sa1, d1l3sa2, d1l6xa1, d1l6xa2, d1l7ra1, d1l7ra2, d1lam_1, d1lam_2,
d1lara1, d1lara2, d1lbu_1, d1lbu_2, d1li1a1, d1li1a2, d1llda1, d1llda2,
d1lr7a1, d1lr7a2, d1lsla1, d1lsla2, d1lu9a1, d1lu9a2, d1ly2a1, d1ly2a2,
d1m0ua1, d1m0ua2, d1m1je1, d1m1je2, d1m1kg1, d1m1kg2, d1m1ta1, d1m1ta2,
d1m2tb1, d1m2tb2, d1m3ya1, d1m3ya2, d1m5ha1, d1m5ha2, d1m6za1, d1m6za2,
d1mdl_1, d1mdl_2, d1mgta1, d1mgta2, d1mixa1, d1mixa2, d1mlda1, d1mlda2,
d1mp9a1, d1mp9a2, d1mpga1, d1mpga2, d1ms5a1, d1ms5a2, d1mxda1, d1mxda2,
d1mzya1, d1mzya2, d1n1ea1, d1n1ea2, d1n1ic1, d1n1ic2, d1n1sa1, d1n1sa2,
d1n2aa1, d1n2aa2, d1n5wa1, d1n5wa2, d1n5wb1, d1n5wb2, d1n5wc1, d1n5wc2,
d1n63b1, d1n63b2, d1n67a1, d1n67a2, d1ne9a1, d1ne9a2, d1nija1, d1nija2,
d1nkr_1, d1nkr_2, d1nm8a1, d1nm8a2, d1np3a1, d1np3a2, d1np7a1, d1np7a2,
d1npya1, d1npya2, d1npyb1, d1npyb2, d1nr0a1, d1nr0a2, d1ntya1, d1ntya2,
d1nu5a1, d1nu5a2, d1nvma1, d1nvma2, d1nyta1, d1nyta2, d1nzoa1, d1nzoa2,
d1o0sa1, d1o0sa2, d1o0wa1, d1o0wa2, d1o0wb1, d1o0wb2, d1o50a1, d1o50a2,
d1o6va1, d1o6va2, d1o7ga1, d1o7ga2, d1o97d1, d1o97d2, d1obba1, d1obba2,
d1obzb1, d1obzb2, d1oe7a1, d1oe7a2, d1oe8b1, d1oe8b2, d1ogad1, d1ogad2,
d1ogae1, d1ogae2, d1ogmx1, d1ogmx2, d1oi1a1, d1oi1a2, d1oipa1, d1oipa2,
d1oiza1, d1oiza2, d1okta1, d1okta2, d1olla1, d1olla2, d1on2a1, d1on2a2,
d1on3a1, d1on3a2, d1on3e1, d1on3e2, d1ooya1, d1ooya2, d1ooyb1, d1ooyb2,
d1or7a1, d1or7a2, d1or7b1, d1or7b2, d1orb_1, d1orb_2, d1ospl1, d1ospl2,
d1ov3a1, d1ov3a2, d1ovna1, d1ovna2, d1ovnb1, d1ovnb2, d1owla1, d1owla2,
d1oxja1, d1oxja2, d1oxsc1, d1oxsc2, d1oxva1, d1oxva2, d1p1da1, d1p1da2,
d1pfqa1, d1pfqa2, d1pgs_1, d1pgs_2, d1pii_1, d1pii_2, d1pina1, d1pina2,
d1pn9a1, d1pn9a2, d1pnf_1, d1pnf_2, d1pprm1, d1pprm2, d1puoa1, d1puoa2,
d1puob1, d1puob2, d1q16a1, d1q16a2, d1q1ba1, d1q1ba2, d1q1ca1, d1q1ca2,
d1q1pa1, d1q1pa2, d1q4ga1, d1q4ga2, d1q7za1, d1q7za2, d1qd1b1, d1qd1b2,
d1qh4a1, d1qh4a2, d1qksa1, d1qksa2, d1qmga1, d1qmga2, d1qneb1, d1qneb2,
d1qnf_1, d1qnf_2, d1qnna1, d1qnna2, d1qr0a1, d1qr0a2, d1qsaa1, d1qsaa2,
d1qxma1, d1qxma2, d1qyab1, d1qyab2, d1r17a1, d1r17a2, d1r27a1, d1r27a2,
d1r4sa1, d1r4sa2, d1rc9a1, d1rc9a2, d1reqd1, d1reqd2, d1rj9a1, d1rj9a2,
d1rjya1, d1rjya2, d1rl6a1, d1rl6a2, d1rm6a1, d1rm6a2, d1rm6b1, d1rm6b2,
d1rqpa1, d1rqpa2, d1rrha1, d1rrha2, d1rwza1, d1rwza2, d1rx0a1, d1rx0a2,
d1rxza1, d1rxza2, d1rzhh1, d1rzhh2, d1rzpc1, d1rzpc2, d1sdwa1, d1sdwa2,
d1se4_1, d1se4_2, d1sgma1, d1sgma2, d1sh5a1, d1sh5a2, d1sjda1, d1sjda2,
d1sjdb1, d1sjdb2, d1skz_1, d1skz_2, d1sli_1, d1sli_2, d1sqga1, d1sqga2,
d1st0a1, d1st0a2, d1svb_1, d1svb_2, d1szna1, d1szna2, d1t0fa1, d1t0fa2,
d1t15a1, d1t15a2, d1t1ea1, d1t1ea2, d1t3qa1, d1t3qa2, d1t3qb1, d1t3qb2,
d1t3qc1, d1t3qc2, d1t56a1, d1t56a2, d1t6ca1, d1t6ca2, d1t6ka1, d1t6ka2,
d1t6sa1, d1t6sa2, d1t7va1, d1t7va2, d1ti6a1, d1ti6a2, d1ti6b1, d1ti6b2,
d1tjea1, d1tjea2, d1tjgl1, d1tjgl2, d1tjla1, d1tjla2, d1tola1, d1tola2,
d1tqya1, d1tqya2, d1tqyb1, d1tqyb2, d1tuaa1, d1tuaa2, d1tvfa1, d1tvfa2,

d1txga1, d1txga2, d1ty0a1, d1ty0a2, d1u00a1, d1u00a2, d1u0va1, d1u0va2,
d1u2ca1, d1u2ca2, d1u3ca1, d1u3ca2, d1u5pa1, d1u5pa2, d1u8va1, d1u8va2,
d1uasa1, d1uasa2, d1ud9a1, d1ud9a2, d1ueka1, d1ueka2, d1umdb1, d1umdb2,
d1umwa1, d1umwa2, d1uok_1, d1uok_2, d1upsa1, d1upsa2, d1uvqa1, d1uvqa2,
d1uxga1, d1uxga2, d1uxy_1, d1uxy_2, d1uypa1, d1uypa2, d1v0ea1, d1v0ea2,
d1v1oa1, d1v1oa2, d1v2aa1, d1v2aa2, d1v54b1, d1v54b2, d1v57a1, d1v57a2,
d1v5va1, d1v5va2, d1v6wa1, d1v6wa2, d1v8ca1, d1v8ca2, d1vcaa1, d1vcaa2,
d1vin_1, d1vin_2, d1vmaa1, d1vmaa2, d1vzia1, d1vzia2, d1vzya1, d1vzya2,
d1w5fa1, d1w5fa2, d1wd3a1, d1wd3a2, d1wejh1, d1wejh2, d1wf3a1, d1wf3a2,
d1wmda1, d1wmda2, d1wopa1, d1wopa2, d1woqa1, d1woqa2, d1wpca1, d1wpca2,
d1xjva1, d1xjva2, d1xo0a1, d1xo0a2, d1xqha1, d1xqha2, d1xsza1, d1xsza2,
d1xtia1, d1xtia2, d1xuua1, d1xuua2, d1y4wa1, d1y4wa2, d1y88a1, d1y88a2,
d2eifa1, d2eifa2, d2fbjh1, d2fbjh2, d2g3pa1, d2g3pa2, d2napa1, d2napa2,
d2pgd_1, d2pgd_2, d2pspa1, d2pspa2, d3hada1, d3hada2, d3hsc_1, d3hsc_2,
d5r1ra1, d5r1ra2, d6ldh_1, d6ldh_2, d7acn_1, d7acn_2, d7taa_1, d7taa_2.

MD_chains (428) : Multi-domain proteins each containing 2 domains, along with the chain identifier at the end of the *PDB* code.

11gsa, 1a4ib, 1a8d_, 1a8h_, 1a8l_, 1a8p_, 1abwa, 1af0a, 1af7_, 1afa1, 1afwb,
1akl_, 1amm_, 1aqh_, 1aqkl, 1avaa, 1aw7a, 1b0na, 1b1ba, 1b2ra, 1b3bc, 1b43a,
1b71a, 1b8aa, 1b8pa, 1b9ka, 1b9wa, 1bg0_, 1bgva, 1bhta, 1bif_, 1bli_, 1bqub,
1bw9a, 1c1aa, 1c1el, 1c2aa, 1c5ka, 1c8ua, 1c9ya, 1ca1_, 1cb6a, 1ccza, 1cdy_,
1cf9a, 1cfb_, 1chma, 1cjdb, 1clc_, 1coja, 1cpo_, 1cuna, 1cvra, 1czya, 1d2oa,
1dd3a, 1dfx_, 1dhka, 1dkxa, 1dl5a, 1dmr_, 1duga, 1duvg, 1duza, 1dvpa, 1dxrh,
1dyka, 1dyqa, 1dzfa, 1e1ca, 1e1cb, 1e3oc, 1e42a, 1e5ta, 1e7la, 1eara, 1eb0a,
1ebha, 1ec9d, 1ecfa, 1ecfb, 1edha, 1eeja, 1eema, 1eerb, 1eg9a, 1ehda, 1eila,
1ek4a, 1ekxa, 1emd_, 1enfa, 1epfa, 1epfb, 1esl_, 1eu3a, 1evka, 1ewca, 1ewfa,
1f0ya, 1f1bb, 1f1ha, 1f1ra, 1f1ua, 1f20a, 1f8ab, 1fcga, 1fnla, 1fnua, 1fp2a,
1fpoa, 1ftra, 1fv1b, 1fw1a, 1fxla, 1g1qa, 1g38a, 1g5aa, 1g6la, 1g87b, 1g8ka,
1g97a, 1giqa, 1gl4a, 1glpa, 1gmua, 1gmuc, 1gqia, 1gqza, 1gsa_, 1gsea, 1gsma,
1gupa, 1gv2a, 1gv3a, 1h03p, 1h1oa, 1h1pb, 1h6ta, 1h72c, 1h8pa, 1h9ra, 1hbmb,
1hf8a, 1hfel, 1hm9a, 1hna_, 1hpua, 1ht6a, 1hw1a, 1hw5a, 1hx6b, 1hyob, 1i1aa,
1i1ga, 1i7na, 1i7pa, 1i8da, 1ig0a, 1ig0b, 1ig3a, 1ihga, 1ii2a, 1iow_, 1iqca,
1isza, 1itkb, 1iyha, 1iz4a, 1iz6a, 1iz9a, 1j09a, 1j11a, 1j3na, 1j8mf, 1jaka,
1jdc_, 1jfla, 1jhda, 1jhfa, 1jlva, 1jpna, 1js1x, 1js4a, 1ju3a, 1k0ma, 1k0mb,
1k8kd, 1k8wa, 1k92a, 1kbva, 1khia, 1kkcx, 1klua, 1klub, 1kqfb, 1ks9a, 1ktba,
1kv9a, 1kwma, 1ky6a, 1kzqa, 1l1fb, 1l3ka, 1l3sa, 1l6xa, 1l7ra, 1lam_, 1lara,
1lbu_, 1li1a, 1llda, 1lr7a, 1lsla, 1lu9a, 1ly2a, 1m0ua, 1m1je, 1m1kg, 1m1ta,
1m2tb, 1m3ya, 1m5ha, 1m6za, 1mdl_, 1mgta, 1mixa, 1mlda, 1mp9a, 1mpga, 1ms5a,
1mxda, 1mzya, 1n1ea, 1n1ic, 1n1sa, 1n2aa, 1n5wa, 1n5wb, 1n5wc, 1n63b, 1n67a,
1ne9a, 1nija, 1nkr_, 1nm8a, 1np3a, 1np7a, 1npya, 1npyb, 1nr0a, 1ntya, 1nu5a,
1nvma, 1nyta, 1nzoa, 1o0sa, 1o0wa, 1o0wb, 1o50a, 1o6va, 1o7ga, 1o97d, 1obba,
1obzb, 1oe7a, 1oe8b, 1ogad, 1ogae, 1ogmx, 1oi1a, 1oipa, 1oiza, 1okta, 1olla,
1on2a, 1on3a, 1on3e, 1ooya, 1ooyb, 1or7a, 1or7b, 1orb_, 1ospl, 1ov3a, 1ovna,
1ovnb, 1owla, 1oxja, 1oxsc, 1oxva, 1p1da, 1pfqa, 1pgs_, 1pii_, 1pina, 1pn9a,
1pnf_, 1pprm, 1puoa, 1puob, 1q16a, 1q1ba, 1q1ca, 1q1pa, 1q4ga, 1q7za, 1qd1b,
1qh4a, 1qksa, 1qmga, 1qneb, 1qnf_, 1qnna, 1qr0a, 1qsaa, 1qxma, 1qyab, 1r17a,
1r27a, 1r4sa, 1rc9a, 1reqd, 1rj9a, 1rjya, 1rl6a, 1rm6a, 1rm6b, 1rqpa, 1rrha,
1rwza, 1rx0a, 1rxza, 1rzhh, 1rzpc, 1sdwa, 1se4_, 1sgma, 1sh5a, 1sjda, 1sjdb,

1skz_, 1sli_, 1sqga, 1st0a, 1svb_, 1szna, 1t0fa, 1t15a, 1t1ea, 1t3qa, 1t3qb,
1t3qc, 1t56a, 1t6ca, 1t6ka, 1t6sa, 1t7va, 1ti6a, 1ti6b, 1tjea, 1tjgl, 1tjla,
1tola, 1tqya, 1tqyb, 1tuaa, 1tvfa, 1txga, 1ty0a, 1u00a, 1u0va, 1u2ca, 1u3ca,
1u5pa, 1u8va, 1uasa, 1ud9a, 1ueka, 1umdb, 1umwa, 1uok_, 1upsa, 1uvqa, 1uxga,
1uxy_, 1uypa, 1v0ea, 1v1oa, 1v2aa, 1v54b, 1v57a, 1v5va, 1v6wa, 1v8ca, 1vcaa,
1vin_, 1vmaa, 1vzia, 1vzya, 1w5fa, 1wd3a, 1wejh, 1wf3a, 1wmda, 1wopa, 1woqa,
1wpca, 1xjva, 1xo0a, 1xqha, 1xsza, 1xtia, 1xuua, 1y4wa, 1y88a, 2eifa, 2fbjh,
2g3pa, 2napa, 2pgd_, 2pspa, 3hada, 3hsc_, 5r1ra, 6ldh_, 7acn_, 7taa_.

1D_homs (261) : Single domain Homologues of MD_doms; belong to the same *SCOP* class, but exist independently.

d101m__, d1a4ra_, d1a53__, d1aa2__, d1abq__, d1ae9a_, d1aera_,
d1ah7__, d1alq__, d1ame__, d1aohb_, d1arl__, d1axb__, d1b07a_,
d1b0b__, d1b0ua_, d1b13a_, d1b1ya_, d1b6e__, d1b88a_, d1bb9__,
d1bd7a_, d1bfea_, d1bg4__, d1bjre_, d1bmg__, d1bqca_, d1btn__,
d1c4rg_, d1c7za_, d1cef__, d1cem__, d1cen__, d1cfwa_, d1ci8a_,
d1cska_, d1cwya_, d1cyw__, d1d0ba_, d1d01a_, d1d8ua_, d1dil__,
d1dl3b_, d1dm1a_, d1dsl__, d1dtja_, d1dtp__, d1dv8a_, d1dy6a_,
d1dypa_, d1e5ja_, d1e8tb_, d1ea7a_, d1eaza_, d1eca__, d1ecea_,
d1edg__, d1edmb_, d1eeub_, d1egza_, d1ek0a_, d1elka_, d1erjc_,
d1eur__, d1f0ma_, d1f3oa_, d1f5oa_, d1f6bb_, d1f8ca_, d1faoa_,
d1fgya_, d1fhla_, d1fica_, d1fm5a_, d1fna__, d1fpra_, d1fuka_,
d1fxxa_, d1fyna_, d1g1ca_, d1g24a_, d1g43a_, d1g68a_, d1g6ha_,
d1g9oa_, d1ga0a_, d1gbg__, d1gcpb_, d1geqb_, d1ghsa_, d1gv5a_,
d1gwna_, d1gz2a_, d1h12a_, d1h1na_, d1h3lb_, d1h4pa_, d1h97a_,
d1hbg__, d1hh8a_, d1hiza_, d1hkfa_, d1hlb__, d1hpt__, d1hq8a_,
d1hura_, d1hxia_, d1i1ja_, d1i1na_, d1i3va_, d1i4na_, d1i71a_,
d1i94l_, d1i94q_, d1i9ea_, d1ia7a_, d1iona_, d1iq7a_, d1itha_,
d1j2aa_, d1j53a_, d1jc9a_, d1jfma_, d1jg1a_, d1jhea_, d1ji0a_,
d1jj7a_, d1jlna_, d1jo8a_, d1jqqb_, d1jvwa_, d1k38a_, d1k9ba_,
d1k9id_, d1kawa_, d1kdka_, d1knma_, d1ks8a_, d1ku1a_, d1ku7a_,
d1kw4a_, d1l1ya_, d1l5ta_, d1l8ja_, d1lfb__, d1luza_, d1lyva_,
d1m6ka_, d1mai__, d1mhqb_, d1mija_, d1mjc__, d1msba_, d1mt0a_,
d1mvea_, d1myt__, d1n4ob_, d1ncna_, d1nega_, d1neu__, d1npsa_,
d1npua_, d1nrfa_, d1ntea_, d1nz6b_, d1o4ya_, d1o4zb_, d1o7va_,
d1od7a_, d1odza_, d1oj6b_, d1ojqa_, d1p2xa_, d1p7id_, d1p9ga_,
d1pa1a_, d1pbk__, d1pht__, d1pkoa_, d1puja_, d1pw3a_, d1pz7a_,
d1q21__, d1q3oa_, d1q9ba_, d1qaua_, d1qdda_, d1qg2a_, d1qhla_,
d1qnoa_, d1qnxa_, d1qqea_, d1qu0c_, d1qvaa_, d1qxwa_, d1qzna_,
d1r6va_, d1r8me_, d1r9ha_, d1rt8a_, d1rzxa_, d1s02__, d1sgwa_,
d1sig__, d1skf__, d1smba_, d1smoa_, d1smxa_, d1sq9a_, d1sulb_,
d1t1ga_, d1t4ob_, d1t5ia_, d1t6na_, d1te5a_, d1ten__, d1thm__,
d1tjca_, d1tpka_, d1tq6a_, d1tu9a_, d1u1ka_, d1u2la_, d1ujpa_,
d1umoa_, d1umza_, d1unpa_, d1ur0b_, d1ur2a_, d1uuqa_, d1v5ca_,
d1vcub_, d1vera_, d1vg8a_, d1vjza_, d1vkaa_, d1vp6a_, d1w0ha_,
d1w1da_, d1w30b_, d1w3ha_, d1wcha_, d1wd5a_, d1wlja_, d1xa1b_,
d1xaua_, d1xc4a_, d1xmja_, d1xt0b_, d1xyza_, d1yfoa_, d2bema_,
d2cuab_, d2his__, d2rmca_, d2semb_, d3ovo__, d3raba_, d4mata_,
d4vhba_, d6rxn__.

1D_uniq (151) : Control dataset of Single domain proteins (No multi-domain Homologs found till date in *SCOP* families v 1.75).

d1bdb__, d1bdo__, d1bea__, d1bff__, d1bgc__, d1bgf__, d1bgp__, d1bhe__,
d1bhp__, d1bj7__, d1bk0__, d1bk9__, d1bky__, d1bm8__, d1bqk__, d1bs9__,
d1bv1__, d1bx7__, d1by2__, d1bzm__, d1c3d__, d1c52__, d1c53__, d1cbs__,
d1ccr__, d1cor__, d1cot__, d1cpq__, d1cwn__, d1cxc__, d1dhn__, d1din__,
d1doi__, d1dun__, d1enj__, d1ezm__, d1f3z__, d1fas__, d1frd__, d1fus__,
d1fxd__, d1gai__, d1gpr__, d1gsl__, d1gvp__, d1hep__, d1hmr__, d1hoe__,
d1hpi__, d1huw__, d1hyp__, d1icm__, d1idk__, d1iob__, d1jdw__, d1jer__,
d1jug__, d1kdi__, d1koe__, d1lci__, d1lis__, d1lki__, d1lml__, d1lz6__,
d1mdc__, d1mho__, d1mn1__, d1msc__, d1ncx__, d1nfp__, d1nox__, d1npc__,
d1nsj__, d1ntn__, d1nxb__, d1oaa__, d1onc__, d1orc__, d1pch__, d1pdo__,
d1pgx__, d1pmi__, d1pmy__, d1pne__, d1poc__, d1prn__, d1puc__, d1pvl__,
d1r69__, d1rds__, d1rkd__, d1rmg__, d1rro__, d1rss__, d1rtu__, d1sur__,
d1tca__, d1tgj__, d1tgl__, d1thg__, d1tia__, d1tib__, d1tif__, d1tig__,
d1tml__, d1tsp__, d1vcc__, d1vhh__, d1vie__, d1vjw__, d1vns__, d1vpe__,
d1wad__, d1wba__, d1wer__, d1whi__, d1xer__, d1xib__, d1yea__, d1zrm__,
d2a0b__, d2abk__, d2acy__, d2bce__, d2c2c__, d2cbp__, d2cts__, d2cut__,
d2cy3__, d2erl__, d2fua__, d2lbd__, d2mcm__, d2pii__, d2plc__, d2por__,
d2prd__, d2ptd__, d2pth__, d2sn3__, d2sns__, d3cbh__, d3cla__, d3dni__,
d3eng__, d3il8__, d3kvt__, d3lip__, d3nul__, d3vub__, d4bcl__, d4mt2__,
d5pal__, d7pcy__.

PPI_Obligate (268) : Obligate Protein-Protein Interaction dataset from the *PDB* for comparison against domain-domain interfaces.

12as, 1a3c, 1a4i, 1aa7, 1ad3, 1ade, 1af5, 1afw, 1ajs, 1aor, 1aoz, 1aq6, 1auo,
1b16, 1b3a, 1b5e, 1b67, 1b8a, 1b8j, 1b8z, 1b9h, 1bam, 1bd0, 1bdm, 1bif, 1biq,
1bis, 1bk5, 1bkp, 1brw, 1bsl, 1bsr, 1buo, 1bxg, 1byf, 1byi, 1byk, 1c4k, 1cd8,
1chm, 1cmb, 1cnz, 1coz, 1csh, 1ctt, 1czj, 1d0c, 1d2f, 1d9c, 1daa, 1db3, 1dbq,
1dcf, 1dor, 1dos, 1dpg, 1dqn, 1dqs, 1dxg, 1e1z, 1e4m, 1e58, 1e6b, 1e6u, 1e7n,
1ebh, 1ecy, 1elu, 1ew3, 1ex0, 1eyv, 1f3h, 1f6b, 1f89, 1fc4, 1fg7, 1fi4, 1fip,
1fjj, 1fp1, 1fp2, 1fp5, 1fro, 1g5c, 1g60, 1g8l, 1gd9, 1glj, 1gnw, 1gqi, 1gt3,
1gta, 1gte, 1gu7, 1gvj, 1gvp, 1gy6, 1gyo, 1h16, 1h1y, 1h7e, 1h9r, 1hf2, 1hpx,
1hq8, 1hss, 1hw1, 1hxp, 1hzj, 1i07, 1i24, 1i2k, 1i4u, 1i52, 1i69, 1i86, 1imb,
1is3, 1iup, 1ivy, 1ix9, 1iyz, 1j3b, 1j5p, 1j98, 1jd0, 1jdw, 1jhd, 1jhg, 1jm6,
1jsc, 1jsg, 1jtv, 1jye, 1jys, 1k04, 1k3y, 1k51, 1k66, 1k94, 1kc7, 1kcm, 1kic,
1kjq, 1kkj, 1kll, 1kly, 1knq, 1kpf, 1kso, 1kta, 1ktb, 1kzh, 1lc5, 1llf, 1lom,
1lxe, 1m6p, 1m7y, 1mh9, 1mkb, 1mo9, 1mor, 1mqi, 1mr8, 1mvc, 1mze, 1n1b, 1n26,
1n2s, 1n55, 1n57, 1n7f, 1nf9, 1nkd, 1nn5, 1noj, 1nox, 1ns5, 1nsj, 1nsy, 1o8b,
1oaa, 1oac, 1oc2, 1oh0, 1oi6, 1oke, 1on2, 1oqz, 1ov3, 1oyb, 1oys, 1p5f, 1pfq,
1pre, 1psr, 1puc, 1pv9, 1pxf, 1q3o, 1q4g, 1q6o, 1qfh, 1qhi, 1qi9, 1qj4, 1qjb,
1qkk, 1qks, 1qr2, 1qvz, 1qyc, 1r2f, 1r31, 1r3s, 1r7a, 1rdo, 1reg, 1rrm, 1ryt,
1s0a, 1s3e, 1sb8, 1se8, 1ses, 1sjy, 1smn, 1smt, 1sox, 1t0i, 1t5h, 1tc1, 1tox,
1trb, 1trk, 1tw3, 1ual, 1uby, 1ui8, 1ukg, 1utg, 1v25, 1v2z, 1vfr, 1vhz, 1vls,
1vok, 1xso, 2arc, 2gsa, 2hdh, 2ilk, 2mjp, 2nac, 2ohx, 2spc, 2tct, 2tgi, 3dap,
3grs, 3lad, 3ssi, 3vub, 5csm, 5rub, 8prk, 9wga.

Supplementary References

1. **A. Hentati *et al.*, *Ann Neurol* 39, 295 (Mar, 1996).**
2. **C. A. Gough, T. Gojobori, T. Imanishi, *Proteins* 66, 69 (Jan 1, 2007).**
3. **R. Rajasekaran, C. Sudandiradoss, C. G. Doss, R. Sethumadhavan, *Genomics* 90, 447 (Oct, 2007).**
4. **P. A. Ory, M. R. Clark, E. E. Kwoh, S. B. Clarkson, I. M. Goldstein, *J Clin Invest* 84, 1688 (Nov, 1989).**
5. **F. Ali-Osman, O. Akande, G. Antoun, J. X. Mao, J. Buolamwini, *J Biol Chem* 272, 10004 (Apr 11, 1997).**