

Supplemental Table 1. List of PSFs that fail to overlap any MaCHoS by 50% (at $I=1.6$).

Interpro ID (Pfam)	Pfam family description	Coverage in residues	# proteins containing domain
IPR000031	1-(5-Phosphoribosyl)-5-amino-4-imidazole-carboxylate (AIR) carboxylase	1057	5
IPR000128	Progesterone receptor, N-terminal	2722	6
IPR000206	Ribosomal protein L7/L12	150	1
IPR000212	DNA helicase, UvrD/REP type	2070	3
IPR000310	Orn/Lys/Arg decarboxylase, major region	398	1
IPR000327	POU-specific	6214	52
IPR000353	MHC class II, beta chain, N-terminal	2929	30
IPR000523	Magnesium chelatase, ChII subunit	4940	16
IPR000534	Semialdehyde dehydrogenase, NAD - binding	367	1
IPR000536	Nuclear hormone receptor, ligand-binding, core	24482	198
IPR000612	Protein of unknown function UPF0057	357	4
IPR000622	K-Cl Co-transporter, type 1	1700	8
IPR000757	Glycoside hydrolase, family 16	2596	6
IPR000802	Arsenical pump membrane protein	1256	2
IPR000848	cAMP-type GPCR	1083	3
IPR000851	Ribosomal protein S5	901	2
IPR000876	Ribosomal protein S4e	261	1
IPR000877	Proteinase inhibitor I12, Bowman-Birk	470	1
IPR000896	Hemocyanin, copper-containing	3916	7
IPR000903	Myristoyl-CoA:protein N-myristoyltransferase	4439	11
IPR000961	Protein kinase, C-terminal	13506	184
IPR000976	Wilm's tumour protein	1040	3
IPR000990	Innexin	3439	9
IPR001083	Copper fist DNA-binding	919	2
IPR001139	Glycoside hydrolase, family 30	2541	5
IPR001155	NADH:flavin oxidoreductase/NADH oxidase, N-terminal	748	2
IPR001217	STAT transcription factor, core	643	1
IPR001220	Legume lectin, beta domain	256	1
IPR001289	CCAAT-binding transcription factor, subunit B	546	4
IPR001292	Oestrogen receptor, AF1 modulatory	700	3
IPR001305	DnaJ central region	3185	24
IPR001352	Ribonuclease HII/HIII	1584	6
IPR001383	Ribosomal protein L28	258	1
IPR001466	Beta-lactamase	1803	4
IPR001529	DNA-directed RNA polymerase, M/15 kDa subunit	874	9
IPR001531	Phospholipase C, zinc-binding, prokaryotic	100	1
IPR001573	Protein kinase A anchoring region	4347	8
IPR001602	Protein of unknown function UPF0047	266	1
IPR001628	Zinc finger, nuclear hormone receptor-type	22664	209
IPR001638	Bacterial extracellular solute-binding protein, family 3	756	3

IPR001708	60 kDa inner membrane insertion protein	2709	10
IPR001721	Threonine dehydratase, C-terminal	197	1
IPR001765	Carbonic anhydrase	255	1
IPR001774	Delta/Serrate/lag-2 (DSL) protein	1619	16
IPR001888	Transposase, type 1	101	1
IPR001893	Cysteine rich repeat	4264	4
IPR001904	Paxillin	1583	6
IPR001975	Ribosomal protein L40e	410	8
IPR002079	Gag polyprotein, inner coat protein p12	385	2
IPR002091	Aromatic amino acid permease	416	1
IPR002489	Glutamate synthase, alpha subunit, C-terminal	3089	2
IPR002494	Keratin, high sulphur B2 protein	14103	92
IPR002509	Polysaccharide deacetylase	2810	10
IPR002544	FMRFamide-related peptide	215	1
IPR002602	Protein of unknown function DB	570	1
IPR002651	Protein of unknown function DUF32	5784	19
IPR002656	Acyltransferase 3	7840	18
IPR002668	Na ⁺ dependent nucleoside transporter	2154	11
IPR002677	Ribosomal protein L32p	638	5
IPR002695	AICARFT/IMPCHase bienzyme	506	1
IPR002717	MOZ/SAS-like protein	7233	24
IPR002719	Retinoblastoma-associated protein, B-box	7253	12
IPR002720	Retinoblastoma-associated protein, A-box	6547	12
IPR002828	Survival protein SurE	336	1
IPR002922	Thiamine biosynthesis Thi4 protein	1523	7
IPR002932	Glutamate synthase, central-C	3385	3
IPR002989	Mycobacterial pentapeptide repeat	947	2
IPR003002	7TM chemoreceptor, subfamily 1	3716	12
IPR003020	Bicarbonate transporter, eukaryotic	729	2
IPR003090	Alpha-crystallin, N-terminal	350	5
IPR003101	Coactivator CBP, KIX	1276	8
IPR003115	ParB-like nuclease	127	1
IPR003142	Biotin protein ligase, C-terminal	2749	5
IPR003149	Iron hydrogenase, small subunit-like	958	2
IPR003156	Phosphoesterase, DHHA1	803	5
IPR003163	DNA-binding, yeast	1692	4
IPR003306	WIF domain	1599	10
IPR003327	Leucine zipper, Myc	375	3
IPR003333	Cyclopropane-fatty-acyl-phospholipid synthase	587	2
IPR003351	Dishevelled protein	1538	11
IPR003368	Polymorphic membrane protein, Chlamydia	286	3
IPR003379	Conserved carboxylase region	1937	6
IPR003460	Insect antifreeze protein	298	3
IPR003543	Macrophage scavenger receptor	267	3
IPR003605	TGF beta receptor, GS motif	2272	31
IPR003611	Intron-encoded nuclease 2	341	1
IPR003702	Acetyl-CoA hydrolase/transferase	1003	2
IPR003703	Acyl-CoA thioesterase	1495	5
IPR003734	Protein of unknown function DUF155	2583	8
IPR003769	Adaptor protein ClpS, core	2989	10
IPR003774	Protein of unknown function DUF179	218	2
IPR003778	Allophanate hydrolase subunit 2	424	1

IPR003833	Allophanate hydrolase subunit 1	402	1
IPR003937	Potassium channel, voltage dependent, KCNQ	302	1
IPR004009	Myosin, N-terminal, SH3-like	5416	74
IPR004060	Orexin receptor, type 2	222	3
IPR004101	Mur ligase, C-terminal	319	2
IPR004102	Poly(ADP-ribose) polymerase, regulatory region	4249	10
IPR004108	Iron hydrogenase, large subunit, C-terminal	3454	8
IPR004115	GAD	1331	3
IPR004118	TT viral ORF2 protein	117	1
IPR004126	Phospholipase A2 inhibitor	314	2
IPR004136	2-nitropropane dioxygenase, NPD	2413	5
IPR004167	E3 binding	2703	14
IPR004181	Zinc finger, MIZ-type	8910	27
IPR004191	Integrase, Tn916-type, N-terminal DNA binding	2057	8
IPR004193	Glycoside hydrolase, family 13, N-terminal	2180	8
IPR004206	mRNA capping enzyme, beta chain	869	2
IPR004244	Transposase, L1	128701	398
IPR004276	Glycosyl transferase, family 28	242	1
IPR004740	Nucleoside:H ⁺ symporter	486	1
IPR004792	HI0933-like protein	1289	5
IPR004813	Oligopeptide transporter OPT superfamily	2401	3
IPR004834	Chitin synthase	938	2
IPR004859	Putative 5-3 exonuclease	6859	10
IPR004860	LAGLIDADG DNA endonuclease	654	3
IPR004874	Borrelia repeat protein	308	1
IPR004886	Glycolipid anchored surface protein GAS1	1871	5
IPR004911	Gamma interferon inducible lysosomal thiol reductase GILT	1334	7
IPR004918	Cdc37	304	1
IPR004923	Iron permease FTR1	869	2
IPR004962	Mab-21 protein	1810	5
IPR004979	Transcription factor AP-2	266	1
IPR005016	TMS membrane protein/tumour differentially expressed protein	9914	22
IPR005100	Supt5 repeat	4523	5
IPR005106	Aspartate/homoserine dehydrogenase, NAD-binding	765	3
IPR005109	Anp1	1430	3
IPR005117	Nitrite/sulphite reductase, hemoprotein beta-component, ferredoxin-like	487	1
IPR005139	PCRF	1817	8
IPR005145	SUA5	362	1
IPR005155	Ribosome biogenesis factor NIP7-like	181	1
IPR005180	Protein of unknown function DUF302	203	1
IPR005185	Protein of unknown function DUF307	908	1
IPR005194	Glycoside hydrolase, family 65, C-terminal	532	1
IPR005196	Glycoside hydrolase family 65, N-terminal	420	1
IPR005200	Glycoside hydrolase, family 81	1496	2
IPR005240	Protein of unknown function DUF389	1314	2

IPR005312	Protein of unknown function DUF1759	87	1
IPR005480	Carbamoyl-phosphate synthetase large chain, oligomerisation	4351	10
IPR005514	Protein of unknown function DUF316	561	2
IPR005515	Vitelline membrane outer layer protein I (VOMI)	604	3
IPR005529	Protein of unknown function DUF321	852	1
IPR005560	Protein of unknown function DUF326	59	1
IPR005567	Fushi tarazu, N-terminal	254	1
IPR005575	Statherin	62	1
IPR005600	Gal4-like dimerisation domain	616	1
IPR005634	Male specific sperm protein	369	3
IPR005798	Cytochrome b/b6, C-terminal	768	4
IPR005839	Protein of unknown function UPF0004	794	2
IPR006031	XYPPX repeat	437	2
IPR006067	Nitrite and sulphite reductase 4Fe-4S region	568	1
IPR006124	Metalloenzyme	239	1
IPR006149	Nematode-specific EB region	541	3
IPR006190	Antifreeze-like/N-acetylneuraminic acid synthase C-terminal	96	1
IPR006218	DAHP synthetase I/KDSA	740	2
IPR006231	Malate:quinone-oxidoreductase	6227	3
IPR006297	GTP-binding protein LepA	117	1
IPR006530	YD repeat	10871	22
IPR006588	Protein of unknown function PAW	547	3
IPR006611	Protein of unknown function DUF1431, cysteine-rich, Drosophila	2023	11
IPR006711	Hox9 activation region	2503	11
IPR006718	Dec-1 repeat	1707	2
IPR006719	DEC-1 protein, N-terminal	1970	2
IPR006720	Dec-1 protein, C-terminal	1707	2
IPR006742	Yeast mating factor alpha hormone	285	2
IPR006763	Erythrocyte membrane-associated giant protein antigen 332	56	1
IPR006774	BAF1 / ABF1 chromatin reorganising factor	731	1
IPR006794	Zfx / Zfy transcription activation region	3759	12
IPR006809	TAFII28-like protein	1609	9
IPR006820	Caudal-like activation region	1734	9
IPR006821	Intermediate filament, DNA-binding region	2637	24
IPR006828	5-AMP-activated protein kinase, beta subunit, complex-interacting region	2672	10
IPR006847	Translation initiation factor IF-2, N-terminal	92	1
IPR006898	Hepatocyte nuclear factor 1, alpha isoform C-terminal	531	5
IPR006899	Hepatocyte nuclear factor 1, N-terminal	3593	10
IPR006903	Protein of unknown function DUF618	5245	18
IPR006947	EGF-like, alliinase	47	1
IPR006953	Uso1/p115 like vesicle tethering protein, head region	3643	5
IPR006955	Uso1/p115 like vesicle tethering protein, C-terminal	3830	5
IPR006982	Glutamate synthase, central-N	3089	2
IPR007073	RNA polymerase Rpb1, domain 7	2109	5
IPR007075	RNA polymerase Rpb1, domain 6	2407	5

IPR007131	SLA1 homology domain 1, SHD1	708	1
IPR007137	Protein of unknown function DUF348	850	1
IPR007147	Protein of unknown function DUF352	1145	2
IPR007177	Protein of unknown function DUF367	1119	5
IPR007187	Nucleoporin, Nup133-like, C-terminal	5000	5
IPR007250	Heat shock protein 9/12	109	1
IPR007337	RelB antitoxin	306	2
IPR007342	Indigoidine synthase A like protein	325	1
IPR007517	Rad50 zinc hook	1487	5
IPR007635	Tis11B like protein, N-terminal	803	6
IPR007638	Glutaminyl-tRNA synthetase, class Ic, non-specific RNA-binding region part 2	988	4
IPR007646	RNA polymerase Rpb2, domain 4	6409	10
IPR007698	Alanine dehydrogenase/PNT, C-terminal	5057	11
IPR007727	Spo12	310	2
IPR007775	LST-1	370	4
IPR007794	Ribosome receptor lysine or proline rich	381	3
IPR007830		253	1
IPR007850	RCSD	415	1
IPR007860	MutS II	8037	21
IPR007861	MutS IV	6368	20
IPR007862	Adenylate kinase, lid region	1724	16
IPR007886	Alanine dehydrogenase/PNT, N-terminal	3300	8
IPR007901	MoeZ/MoeB	1405	5
IPR007934	Alpha-L-arabinofuranosidase B	551	3
IPR008013	GATA-type transcription activator, N-terminal	2328	10
IPR008042	Retrotransposon, Pao	418	1
IPR008278	4'-phosphopantetheinyl transferase	1559	6
IPR008403	Apolipoprotein CIII	216	2
IPR008408	Brain acid soluble protein 1	674	3
IPR008442	Propeptide, carboxypeptidase Y	109	1
IPR008454	Collagen-binding surface protein Cna-like, B region	6345	6
IPR008539	Protein of unknown function DUF821, CAP10-like	1122	3
IPR008626	GAL11 transcriptional regulator	1081	1
IPR008630	Galactosyl transferase	815	2
IPR008640	Hep/Hag	289	1
IPR008671	Lycopene beta and epsilon cyclase	477	1
IPR008675	Mating factor alpha precursor, N-terminal	285	2
IPR008715	NodS	310	1
IPR008803	Root hair defective 3 GTP-binding	764	3
IPR008816	Rickettsia 17 kDa surface antigen	80	1
IPR008837	Serendipity locus alpha	1116	2
IPR008850	TEP1, N-terminal	3239	4
IPR008893	WGR	3905	11
IPR009223	APC cysteine-rich	8513	8
IPR009224	SAMP	5035	6
IPR009232	EB-1 binding	562	3
IPR009234	APC basic	4441	6
IPR009240	APC 15 residue	1957	5
IPR009255	Transcriptional coactivation	5219	8
IPR009398	Adenylate cyclase-like	3527	22
IPR009401	TRAP240	16799	9
IPR009435	Acid shock	138	1
IPR009449	GDPGTP exchange factor Sec2p	1894	8

IPR009462	Protein of unknown function DUF1086	3275	14
IPR009463	Protein of unknown function DUF1087	2304	14
IPR009535	Protein of unknown function DUF1154	803	7
IPR009581	Protein of unknown function DUF1193	3097	10
IPR009644	Fukutin-related	310	4
IPR009651	Aluminium resistance	336	1
IPR009674	RNA polymerase I, Rpa2 specific	2839	5
IPR009724	Protein of unknown function DUF1301	800	4
IPR010011	Protein of unknown function DUF1518	4385	13
IPR010407	Signaling lymphocytic activation molecule, N-terminal	709	3
IPR010456	Ribosomal L11 methyltransferase	708	4
IPR010465	DRF autoregulatory	730	14
IPR010481	CDC24 calponin	95	1
IPR010482	Integral peroxisomal membrane peroxin	2531	5
IPR010483	Alpha-2-macroglobulin RAP, C-terminal	1306	4
IPR010491	PRP1 splicing factor, N-terminal	1770	6
IPR010508	Protein of unknown function DUF1088	1783	7
IPR010526	Sodium ion transport-associated	10214	40
IPR010579	MHC class I, alpha chain, C-terminal	2307	38
IPR010619	Protein of unknown function DUF1212	69	1
IPR010643	Protein of unknown function DUF1227	2337	5
IPR010656	TRAP C4-dicarboxylate transport system permease DctM subunit	605	1
IPR010660	Notch, NOD region	838	10
IPR010684	RNA polymerase II transcription factor SIII, subunit A	3216	8
IPR010721	Protein of unknown function DUF1295	683	2
IPR010729	Ribosomal protein L47, mitochondrial	1095	5
IPR010822	Sporulation stage II, protein E C-terminal	255	1
IPR010828	Alcohol acetyltransferase	1060	2
IPR010893	Hydrogenase-1 expression HyaE	144	1
IPR010896	Nuclease-associated modular DNA-binding 1	148	1
IPR011063	PP-loop	3501	11
IPR011123	Two component regulator three Y	671	2
IPR011422	BRCA1-associated 2	1021	4
IPR011424	C1-like	723	10
IPR011430	Down-regulated in metastasis	1878	5
IPR011546	Peptidase M41, FtsH extracellular	3050	16
IPR011645	Heme NO binding associated	6266	19
IPR011666	Protein of unknown function DUF1604	2893	4
IPR011996	Potassium channel, calcium-activated, SK, conserved region	5183	15
IPR012280	Semialdehyde dehydrogenase, dimerisation region	367	1
IPR012435	Protein of unknown function DUF1632	1041	3
IPR012496	TMC	7763	25
IPR012508	ATPase, A1 complex, epsilon subunit	335	1
IPR012529	Attractin	125	1
IPR012532	BDHCT	446	3
IPR012541	DBP10CT	672	6
IPR012560	FerA	2604	10
IPR012561	FerB	3533	22
IPR012567	IlvGEDA operon leader peptide	38	1
IPR012572	Mad3-like	469	2
IPR012575	MNLL subunit	105	1

IPR012581	NUC156	2168	4
IPR012587	P68HR	563	4
IPR012592	PROCN	9535	5
IPR012599	Peptidase C1A, propeptide	413	4
IPR012603	RBB1NT	1547	8
IPR012642	Mitotic Wos2	1819	8
IPR012848	Propeptide, peptidase A1	1936	27
IPR012897	Potassium channel, voltage dependent, Kv1.4, tandem inactivation	216	2
IPR012920	Spb1, C-terminal	2483	5
IPR012923	Replication fork protection component Swi3	888	4
IPR012934	Zinc finger, AD-type	11288	92
IPR012946	X8	347	2
IPR012956	CBF, N-terminal	487	4
IPR012962	Peptidase M54, archaemetzincin	862	3
IPR012966	Protein of unknown function DUF1709, fungi	306	1
IPR012968	Fer1	5429	19
IPR012979	Protein of unknown function NUC189, C-terminal	1585	5
IPR012982	PADR1	1497	5
IPR012996	Zinc finger, CHHC-type	2277	12
IPR013086	Serotonin (5-HT) neurotransmitter transporter, N-terminal	208	3
IPR013113	FAD-binding 9, siderophore-interacting	628	5
IPR013124	Gap junction alpha-1 protein (Cx43), C-terminal	197	4
IPR013125	Gap junction alpha-8 protein (Cx50), C-terminal	222	2
IPR013180	Protein of unknown function DUF1716, eukaryotic	2143	4
IPR013182	Protein of unknown function DUF1720	764	2
IPR013221	Mur ligase, central	366	1
IPR013223	Ribonuclease B, OB region N-terminal	835	7
IPR013226	Pal1 cell morphology	865	2
IPR013235	Serine/threonine phosphatase, PPP5	1779	16
IPR013253	Kinetochore Spc7	917	1
IPR013257	SRI, Set2 Rpb1 interacting	129	1
IPR013498	DNA topoisomerase, type IA, zn finger	532	4
IPR013511		622	2
IPR013513	Integrin alpha chain, C-terminal cytoplasmic region	6139	53
IPR013530	Protein-arginine deiminase, C-terminal	8264	15
IPR013538	Activator of Hsp90 ATPase homologue 1-like	1169	6
IPR013544	Protein of unknown function DUF1725	15351	305
IPR013566	EF hand associated, type-1	3080	10
IPR013592	Maf transcription factor, N-terminal	1893	12
IPR013600	Ly49-like N-terminal	7387	45
IPR013612	Amino acid permease N-terminal	1754	16
IPR013621	Ion transport N-terminal	4014	13
IPR013629	Zinc finger-associated region	905	9
IPR013646	GTPase of unknown function C-terminal	206	1
IPR013656	PAS fold-4	427	3
IPR013662	RyR and IP3R Homology associated	36468	26
IPR013665	Sfi1 spindle body	656	1
IPR013673	Potassium channel, inwardly rectifying,	1851	6

	Kir, N-terminal		
IPR013680	Voltage-dependent calcium channel, alpha-2/delta subunit, conserved region	5782	8
IPR013697	Region of unknown function DUF1744	9292	6
IPR013705	Sterol methyltransferase C-terminal	383	1
IPR013706	3'5'-cyclic nucleotide phosphodiesterase N-terminal	1445	15
IPR013716	Adenylate cyclase G-alpha binding	67	1
IPR013732	Protein-arginine deiminase (PAD) N-terminal	2565	16
IPR013743	Nap binding NBP1, fungi	540	2
IPR013747	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III C-terminal	900	8
IPR013798	Indole-3-glycerol phosphate synthase	427	1
IPR013803	Amyloidogenic glycoprotein, amyloid-beta peptide	1058	3
IPR013851	Transcription factor Otx, C-terminal	664	5