

Table 1 Protein size summary.

Group	Species Code	Species Name	Length aa		Percentiles				
			Mean	SD	10%	25%	50%	75%	90%
ARCHAEA	ARC_PRO	<i>Archaeoglobus profundus</i> DSM5631	263	187	80	128	221	346	479
ARCHAEA	CAN_KOR	<i>Candidatus Korarchaeum cryptofilum</i> OPF8	296	191	104	160	262	379	501
ARCHAEA	CEN_SYM	<i>Cenarchaeum symbiosum</i> A	308	535	74	117	213	348	521
ARCHAEA	DES_KAM	<i>Desulfurococcus kamchatkensis</i> 1221n	272	188	75	129	238	369	499
ARCHAEA	MET_JAN	<i>Methanococcus jannaschii</i>	283	204	98	149	241	365	492
ARCHAEA	NAN_EQU	<i>Nanoarchaeum equitans</i> Kin4-M	276	203	91	142	225	352	512
ARCHAEA	SUL_ACI	<i>Sulfolobus acidocaldarius</i> DSM 639	284	183	96	146	249	375	511
ARCHAEA	THE_NEU	<i>Thermoproteus neutrophilus</i> V24Sta	268	182	91	142	230	346	463
ARCHAEA	THE_VOL	<i>Thermoplasma volcanium</i> GSS1	297	198	98	157	258	390	518
BACTERIA	ACI_FER	<i>Acidimicrobium ferrooxidans</i> DSM 10331	322	203	109	174	287	415	553
BACTERIA	BAC_FRA	<i>Bacteroides fragilis</i> NCTC 9343	361	249	107	182	310	455	691
BACTERIA	BAC_SUB	<i>Bacillus subtilis</i> 168	294	266	85	145	254	382	504
BACTERIA	BIF_ADO	<i>Bifidobacterium adolescentis</i> ATCC 15703	369	233	136	218	325	461	654
BACTERIA	BRA_JAP	<i>Bradyrhizobium japonicum</i> USDA 110	317	229	107	170	277	403	552
BACTERIA	BUR_CEP	<i>Burkholderia cepacia</i> AMMD	330	250	110	180	295	410	549
BACTERIA	CAM_JEJ	<i>Campylobacter jejuni</i> RM1221	294	202	83	150	254	392	538
BACTERIA	CHL_MUR	<i>Chlamydia muridarum</i> Nigg	355	296	105	172	290	446	650
BACTERIA	COR_AUR	<i>Corynebacterium aurimucosum</i> ATCC 700975	325	225	105	177	283	417	557
BACTERIA	DEL_DES	<i>Deinococcus deserti</i> VCD115	314	209	117	169	274	395	552
BACTERIA	ESC_COL	<i>Escherichia coli</i> O157:H7 str. EC4115	287	236	58	121	239	384	548
BACTERIA	GLO_VIO	<i>Gloeobacter violaceus</i> PCC 7421	313	233	95	151	256	398	593
BACTERIA	HYD_THE	<i>Hydrogenobacter thermophilus</i> TK-6	293	198	93	149	251	389	540
BACTERIA	KOC_RHI	<i>Kocuria rhizophila</i> DC2201	337	213	118	189	300	434	578
BACTERIA	LEP_BIF	<i>Leptospira biflexa</i> Patoc 1 (Ames)	338	216	123	184	292	430	611
BACTERIA	MYC_ABS	<i>Mycobacterium abscessus</i>	317	250	115	174	273	400	524
BACTERIA	PER_MAR	<i>Persephonella marina</i> EX-H1	304	240	95	152	256	392	569
BACTERIA	STA_AUR	<i>Staphylococcus aureus aureus</i> MW2	298	285	84	149	254	385	522
BACTERIA	STR_AVE	<i>Streptomyces avermitilis</i> MA-4680	341	308	115	182	289	422	578
BACTERIA	SUL_DEL	<i>Sulfurospirillum deleyianum</i> DSM 6946	312	223	101	166	266	403	577
BACTERIA	SYN_SP	<i>Synechocystis</i> sp. PCC 6803	319	256	96	153	264	404	584
BACTERIA	THE_ELO	<i>Thermosynechococcus elongatus</i> BP-1	314	214	98	157	273	403	577
BACTERIA	THE_THE	<i>Thermus thermophilus</i> HB27	303	199	109	167	264	390	529
BACTERIA	XAN_CAM	<i>Xanthomonas campestris</i> pv <i>armoraciae</i>	311	258	59	134	257	412	623
APICOMPLEXA	CRY_PAR	<i>Cryptosporidium parvum</i>	597	628	155	251	433	729	1192
APICOMPLEXA	PLA_FAL	<i>Plasmodium falciparum</i>	753	866	145	253	453	930	1707
APICOMPLEXA	TOX_GON	<i>Toxoplasma gondii</i>	682	766	139	224	441	843	1486
CILIOPHORA	PAR_TET	<i>Paramecium tetraurelia</i>	457	438	127	205	348	541	854
CILIOPHORA	TET_THE	<i>Tetrahymena thermophila</i>	649	660	110	229	456	839	1396
AMOEBOZOA	DIC_DIS	<i>Dictyostelium discoideum</i>	533	513	92	198	392	702	1123
DIPLOMONADIDA	GUL_LAM	<i>Giardia lamblia</i>	543	630	84	180	369	689	1110
PLACOZOA	TRI_ADH	<i>Trichoplax adhaerens</i>	453	426	141	217	345	539	854
FUNGI_ASC	PIC_STI	<i>Pichia stipitis</i>	492	346	161	263	416	613	893
FUNGI_ASC	SAC_CER	<i>Saccharomyces cerevisiae</i>	497	382	137	239	409	632	951
FUNGI_ASC	TRI_REE	<i>Trichoderma reesei</i>	491	452	154	262	408	600	891
FUNGI_BAS	LAC_BIC	<i>Laccaria bicolor</i>	370	312	88	153	289	488	749
FUNGI_BAS	PHA_CHR	<i>Phanerochaete chrysosporium</i> strain RP78	456	327	157	246	373	556	856
FUNGI_BAS	UST_MAY	<i>Ustilago maydis</i>	613	454	176	298	501	793	1198
STRAM_DIA	PHA_TRI	<i>Phaeodactylum tricomutum</i>	462	343	162	249	381	562	841
STRAM_DIA	THA_PSE	<i>Thalassiosira pseudonana</i>	499	424	159	249	391	613	947
STRAM_OOM	PHY_RAM	<i>Phytophthora ramorum</i>	479	407	152	237	373	584	903

Table 1 Protein size summary. (Continued)

STRAM_OOM	PHY_SOJ	<i>Phytophthora sojae</i>	502	447	146	234	382	616	986
CNIDARIA	NEM_VEC	<i>Nematostella vectensis</i>	335	336	95	145	250	405	646
INSECTA	ANO_GAM	<i>Anopheles gambiae</i>	529	547	132	223	389	632	1065
INSECTA	DRO_MEL	<i>Drosophila melanogaster</i>	584	642	141	242	427	700	1164
NEMATODA	CAE_ELE	<i>Caenorhabditis elegans</i>	444	484	124	211	342	522	820
NEMATODA	PRI_PAC	<i>Pristionchus pacificus</i>	288	285	76	116	206	359	583
VERT_AVE	GAL_GAL	<i>Gallus gallus</i>	490	508	108	184	346	608	1007
VERT_AVE	MEL_GAL	<i>Meleagris gallopavo</i>	479	463	116	197	351	595	968
VERT_MAM	BOS_TAU	<i>Bos taurus</i>	495	490	145	246	356	592	947
VERT_MAM	EQU_CAB	<i>Equus caballus</i>	564	606	147	247	393	688	1139
VERT_MAM	HOM_SAP	<i>Homo sapiens</i>	456	540	98	163	311	562	947
VERT_MAM	MON_DOM	<i>Monodelphis domestica</i>	574	489	174	295	457	719	1069
VERT_MAM	ORN_ANA	<i>Ornithorhynchus anatinus</i>	445	416	123	202	327	540	868
VERT_MAM	RAT_NOR	<i>Rattus norvegicus</i>	520	500	130	224	374	643	1039
VERT_SAU	ANO_CAR	<i>Anolis carolinensis</i>	462	436	128	207	346	559	903
VERT_TEL	DAN_RER	<i>Danio rerio</i>	473	456	151	234	363	565	879
VERT_TEL	TAK_RUB	<i>Takifugu rubripes</i>	634	536	215	324	494	780	1177
PLANT_BRY	PHY_PAT	<i>Physcomitrella patens</i>	363	308	115	165	278	461	711
PLANT_CHL	CHL_REI	<i>Chlamydomonas reinhardtii</i>	503	589	97	173	335	608	1074
PLANT_CHL	MIC_CCM	<i>Micromonas CCMPI545</i>	426	390	123	202	334	522	799
PLANT_CHL	MIC_RCC	<i>Micromonas RCC299</i>	485	475	146	236	371	571	920
PLANT_CHL	OST_LUC	<i>Ostreococcus lucimarinus</i>	397	343	121	199	319	486	726
PLANT_CHL	OST_TAU	<i>Ostreococcus tauri</i>	387	349	114	186	307	476	716
PLANT_DIC	ARA_THA	<i>Arabidopsis thaliana</i>	403	299	115	202	345	513	749
PLANT_DIC	CAR_PAP	<i>Carica papaya</i>	296	249	68	112	225	411	611
PLANT_DIC	GLY_MAX	<i>Glycine max</i>	422	354	139	220	353	529	768
PLANT_DIC	MED_TRU	<i>Medicago truncatula</i>	245	245	59	78	149	334	550
PLANT_DIC	POP_TRI	<i>Populus trichocarpa</i>	375	292	101	167	306	490	732
PLANT_LYC	SEL_MOE	<i>Selaginella moellendorffii</i>	382	300	124	191	316	481	699
PLANT_MON	BRA_DIS	<i>Brachypodium distachyon</i>	428	303	146	223	361	537	788
PLANT_MON	ORY_SAT	<i>Oryza sativa ssp. japonica</i>	448	389	108	174	332	574	960
PLANT_MON	SOR_BIC	<i>Sorghum bicolor</i>	361	282	103	167	288	476	706
PLANT_MON	ZEA_MAY	<i>Zea mays</i>	345	258	97	164	286	455	655
RHODOPHYTA	CYA_MER	<i>Cyanidioschyzon merolae</i>	504	404	158	259	412	628	918

Statistical summary of protein length values in the proteomes of dataset 1 including 84 different species (9 archeal, 24 bacterial and 51 eukaryotic organisms). The mean, standard deviation (SD) and the 10%, 25%, 50%, 75% and 90% percentiles were calculated for each organism individually (see methods)