

Supplementary data for

# Origin of Protein Aggregation: Identification of Some Characteristic Traits in Structured and Intrinsically Disordered Proteins

Uttam Pal, Anupam Roy, Supriya Das, Swagata Das, Mangaldeep Kundu, Khyati Bagga, Nakul Chandra Maiti\*

Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research (CSIR)-Indian Institute of Chemical Biology (IICB), 4, Raja S.C. Mullick Road, Kolkata 700032, India

\* Address correspondence to Nakul C. Maiti, Email: [ncmaiti@iicb.res.in](mailto:ncmaiti@iicb.res.in), Phone: +91-33-2499-5940, Phone: +91-33-2499-5886, Fax: +91-33-2473-5197

## Table of contents

Table S1 Disordered proteins in the dataset.	...page 2
Table S2 Structured proteins in the dataset.	...page 17
Figure S1	...page 35
Table S3 P-values	...page 35

**Table S1 Disordered proteins in the dataset.** Accession codes, name, sequence length and the computed physicochemical properties are listed.

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P49913</b>	DP00004_C002	Antibacterial protein LL-37	37	NON-AM	10.61	23.34	89.46	-0.724
<b>P27695</b>	DP00007	DNA-(apurinic or apyrimidinic site) lyase	318	AM	8.33	43.41	77.36	-0.583
<b>P50224</b>	DP00011	Sulfotransferase 1A3/1A4	295	AM	5.68	36.36	73.66	-0.489
<b>P13569</b>	DP00012	Cystic fibrosis transmembrane conductance regulator	1480	AM	8.91	43.8	102.82	0.024
<b>P01233</b>	DP00013	Choriogonadotropin subunit beta	145	NON-AM	8.65	64.41	75.24	-0.221
<b>P38936</b>	DP00016	Cyclin-dependent kinase inhibitor 1	164	NON-AM	8.69	80.9	56.52	-0.845
<b>P49918</b>	DP00017	Cyclin-dependent kinase inhibitor 1C	316	NON-AM	5.39	89.56	76.52	-0.21
<b>P46527</b>	DP00018	Cyclin-dependent kinase inhibitor 1B	198	NON-AM	6.54	66.05	44.39	-1.261
<b>P14061</b>	DP00023	Estradiol 17-beta-dehydrogenase 1	327	AM	5.47	42.16	94.86	0.051
<b>Q13541</b>	DP00028	Eukaryotic translation initiation factor 4E-binding protein 1	118	NON-AM	5.32	78.89	52.03	-0.701
<b>P04150</b>	DP00030	Glucocorticoid receptor	777	AM	6	51.67	78.16	-0.393
<b>P10912</b>	DP00033_C001	Growth hormone-binding protein	246	AM	5.25	42.24	66.5	-0.578
<b>P05204</b>	DP00039	Non-histone chromosomal protein HMG-17	89	NON-AM	10	30.39	28.88	-1.761
<b>P17096</b>	DP00040	High mobility group protein HMG-I/HMG-Y	107	NON-AM	10.31	82.35	26.45	-1.807
<b>P78356</b>	DP00054	Phosphatidylinositol-4-phosphate 5-kinase type II beta	416	AM	6.9	40.34	73.05	-0.503

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P27694</b>	DP00061	Replication protein A 70 kDa DNA-binding subunit	616	AM	6.92	35.83	78.96	-0.362
<b>P19793</b>	DP00062	Retinoic acid receptor RXR-alpha	462	AM	7.92	45.77	78.74	-0.309
<b>P63027</b>	DP00069	Vesicle-associated membrane protein 2	116	AM	7.84	41.34	101.03	-0.017
<b>P37840-1</b>	DP00070	Alpha-synuclein	140	AM	4.67	25.47	69.64	-0.403
<b>P04818</b>	DP00073	Thymidylate synthase	313	AM	6.51	47.24	83.19	-0.412
<b>P03372</b>	DP00074	Estrogen receptor [Isoform long]	595	AM	8.3	45.88	80.39	-0.354
<b>P11387</b>	DP00075	DNA topoisomerase 1	765	AM	9.33	45.74	58.92	-1.29
<b>P01100</b>	DP00078	Proto-oncogene protein c-fos	380	AM	4.77	78.82	65.32	-0.369
<b>Q92731-1</b>	DP00079	Estrogen receptor beta [Isoform 1 (Beta-1)]	530	AM	8.81	55.32	83.34	-0.288
<b>P61244</b>	DP00084	Protein max [Isoform 1]	160	AM	5.88	82.09	55.62	-1.324
<b>P61244-2</b>	DP00084_A002	Protein max [Isoform 2]	151	AM	6.07	79.27	58.28	-1.295
<b>Q04206</b>	DP00085	Transcription factor p65	551	AM	5.46	54.79	73.16	-0.465
<b>P04637</b>	DP00086	Cellular tumor antigen p53	393	AM	6.33	73.59	59.08	-0.756
<b>Q14722_2</b>	DP00090	Voltage-gated potassium channel subunit beta-1 [Isoform KvB1.1]	401	AM	9.39	50.28	88.5	-0.283
<b>Q08209</b>	DP00092	Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	521	AM	5.58	48.39	82	-0.316

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P50440</b>	DP00099	Glycine amidinotransferase, mitochondrial	423	AM	8.26	53.39	75.41	-0.424
<b>P04350</b>	DP00114	Tubulin beta-4 chain	444	AM	4.81	37.11	71.35	-0.322
<b>P10163</b>	DP00119	Basic salivary proline-rich protein 4	310	AM	10.52	99.82	17.97	-1.694
<b>P01730</b>	DP00123	T-cell surface glycoprotein CD4	458	AM	9.6	41.78	94.67	-0.257
<b>P10636-8</b>	DP00126	Microtubule-associated protein tau [Isoform Tau-F]	441	AM	8.24	47.59	57.3	-0.868
<b>P22531</b>	DP00130	Small proline-rich protein 2E	72	NON-AM	8.78	119.31	8.06	-1.258
<b>Q06787</b>	DP00134	Fragile X mental retardation 1 protein	632	AM	7	47.28	68.91	-0.847
<b>P12272</b>	DP00138	Parathyroid hormone-related protein	177	NON-AM	10.19	62.5	66.61	-1.053
<b>O95405-1</b>	DP00141	Zinc finger FYVE domain-containing protein 9 [Isoform 1]	1425	AM	4.89	46.15	76.13	-0.405
<b>Q13426</b>	DP00152	DNA repair protein XRCC4	336	AM	4.91	48.32	74.91	-0.749
<b>P01112</b>	DP00153	GTPase HRas	189	AM	5.16	43.27	81.96	-0.417
<b>P29353</b>	DP00154	SHC-transforming protein 1	583	AM	5.92	56.63	71.96	-0.507
<b>P54725</b>	DP00156	UV excision repair protein RAD23 homolog A	363	AM	4.56	55.63	74.21	-0.535
<b>O75324</b>	DP00162	Stannin	88	AM	5.19	42.28	108.52	0.263
<b>P19429</b>	DP00166	Troponin I, cardiac muscle	210	NON-AM	9.87	53.02	77.29	-0.93

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P10114</b>	DP00167	Ras-related protein Rap-2a	183	AM	4.73	44.34	83.01	-0.28
<b>Q9NVS9</b>	DP00168	Pyridoxine-5'-phosphate oxidase	261	AM	6.61	41.02	62.03	-0.654
<b>P04049</b>	DP00171	RAF proto-oncogene serine/threonine-protein kinase	648	AM	9.33	51.86	76.25	-0.392
<b>Q16633</b>	DP00172	POU domain class 2-associating factor 1	256	AM	4.87	71.1	79.3	-0.199
<b>O00273-1</b>	DP00173	DNA fragmentation factor subunit alpha [Isoform DFF45]	331	AM	4.68	62.19	93.17	-0.452
<b>P16949</b>	DP00174	Stathmin	149	AM	5.76	70.87	66.91	-1.311
<b>Q9NQB0</b>	DP00175	Transcription factor 7-like 2	619	AM	8.73	64.24	66.87	-0.745
<b>Q9Y5B0</b>	DP00177	RNA polymerase II subunit A C-terminal domain phosphatase	961	AM	5.16	61.05	66.36	-0.732
<b>P11473</b>	DP00184	Vitamin D3 receptor	427	AM	6.08	53.26	77.87	-0.41
<b>P00747</b>	DP00191	Plasminogen	810	AM	7.04	47.13	55.46	-0.688
<b>P20810</b>	DP00196	Calpastatin	708	NON-AM	4.99	49.83	53.56	-1.193
<b>P05814</b>	DP00199	Beta-casein	226	NON-AM	5.52	73.18	110.75	-0.094
<b>P20963-1</b>	DP00200	T-cell surface glycoprotein CD3 zeta chain [Isoform 1]	164	AM	9.28	58.34	80.43	-0.649
<b>P62993</b>	DP00210	Growth factor receptor-bound protein 2	217	AM	5.89	37.93	65.58	-0.668
<b>P24592</b>	DP00211	Insulin-like growth factor-binding protein 6	240	AM	8.15	68.08	57.79	-0.689

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P10451</b>	DP00214	Osteopontin	314	NON-AM	4.37	53.37	57.42	-1.1
<b>P42768</b>	DP00215	Wiskott-Aldrich syndrome protein	501	AM	6.81	69.36	64.51	-0.711
<b>O60927</b>	DP00219	Protein phosphatase 1 regulatory subunit 11	126	NON-AM	6.52	58.83	36.35	-1.205
<b>P07476</b>	DP00221	Non-histone chromosomal protein HMG-17	585	AM	4.62	81.44	73.09	-1.503
<b>P14635</b>	DP00223	G2/mitotic-specific cyclin-B1	433	AM	7.09	50.59	90.09	-0.239
<b>P15884</b>	DP00224	Transcription factor 4	667	AM	6.51	59.83	54.29	-0.848
<b>P05455</b>	DP00229	Lupus La protein	408	AM	6.68	38.02	70.32	-0.973
<b>Q14449</b>	DP00230	Growth factor receptor-bound protein 14	540	AM	8.56	52.45	74.81	-0.438
<b>P14859</b>	DP00231	POU domain, class 2, transcription factor 1	743	AM	6.34	54.75	83.96	-0.161
<b>P02686</b>	DP00236	Myelin basic protein	304	NON-AM	9.79	41.65	41.84	-1.174
<b>P38398</b>	DP00238	Breast cancer type 1 susceptibility protein	1863	AM	5.29	54.68	69.01	-0.785
<b>P35611-1</b>	DP00240	Alpha-adducin [Isoform 1]	737	AM	5.6	53.78	71.07	-0.632
<b>P35612-1</b>	DP00241	Beta-adducin [Isoform 1]	726	AM	5.67	58.35	68.5	-0.663
<b>P23025</b>	DP00243	DNA-repair protein complementing XP-A cells	273	NON-AM	6.29	58.66	70.48	-0.77
<b>O75365-1</b>	DP00254	Protein tyrosine phosphatase type IVA 3 [Isoform 1]	173	AM	9.35	39.47	81.68	-0.268

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q93096</b>	DP00255	Protein tyrosine phosphatase type IVA 1 [Isoform 1]	173	AM	9.17	39.11	86.76	-0.301
<b>P01106-1</b>	DP00260	Myc proto-oncogene protein [Isoform 1]	439	AM	5.33	92.23	66.42	-0.772
<b>Q9Y3R4</b>	DP00261	Sialidase-2	380	AM	6.35	46.07	77.08	-0.368
<b>Q16665-1</b>	DP00262	Hypoxia-inducible factor 1 alpha [Isoform 1]	826	AM	5.17	55.97	74.96	-0.573
<b>P35372-1</b>	DP00272	Mu-type opioid receptor [Isoform 1]	400	AM	8.62	37.39	97.72	0.311
<b>Q04656-1</b>	DP00282	Copper-transporting ATPase 1 [Isoform 4]	1500	AM	5.94	42.55	102.97	0.092
<b>P40337</b>	DP00287	Von Hippel-Lindau disease tumor suppressor [Isoform 1]	213	AM	4.7	68.65	75.45	-0.777
<b>P10415</b>	DP00297	Apoptosis regulator Bcl-2	239	AM	6.75	51.63	78.03	-0.136
<b>Q07817</b>	DP00298	Bcl-2-like protein 1	233	AM	4.86	39.64	73.26	-0.345
<b>P31751</b>	DP00304	RAC-beta serine/threonine-protein kinase	481	AM	5.98	35.09	77.01	-0.473
<b>P43155-1</b>	DP00305	Carnitine O-acetyltransferase [Isoform 1]	626	AM	8.63	45.31	86.61	-0.238
<b>Q07108</b>	DP00306	Early activation antigen CD69	199	AM	6.16	38.73	62.16	-0.413
<b>P51946</b>	DP00307	Cyclin-H	323	AM	6.73	50.65	89.07	-0.379
<b>P06132</b>	DP00308	Uroporphyrinogen decarboxylase	367	AM	5.77	49.12	88.56	-0.138
<b>P00533-1</b>	DP00309	Epidermal growth factor receptor [Isoform 1]	1210	AM	6.26	44.89	80.74	-0.316

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P62495</b>	DP00310	Eukaryotic peptide chain release factor subunit 1	437	AM	5.51	27.82	88.76	-0.311
<b>P24071-1</b>	DP00311	Immunoglobulin alpha Fc receptor [Isoform A.1]	287	AM	6.44	36.12	89.37	-0.136
<b>P36888</b>	DP00312	FL cytokine receptor	993	AM	5.52	49.93	79.1	-0.234
<b>P32455</b>	DP00313	Interferon-induced guanylate-binding protein 1	592	AM	5.97	45.42	83.34	-0.558
<b>Q9UJY5-1</b>	DP00314	ADP-ribosylation factor-binding protein GGA1 [Isoform 1]	639	AM	5.18	59.64	83.94	-0.443
<b>O43464-1</b>	DP00315	Serine protease HTRA2, mitochondrial [Isoform 1 (13B)]	458	AM	10.07	43.11	96.46	-0.056
<b>P08581-1</b>	DP00317	Hepatocyte growth factor receptor [Isoform 1]	1390	AM	7.02	42.61	87.95	-0.144
<b>Q8WUM0</b>	DP00318	Nuclear pore complex protein Nup133	1156	AM	4.98	46.95	92.98	-0.237
<b>O15350-1</b>	DP00319	Tumor protein p73 [Isoform Alpha]	636	AM	6.47	60.53	67.01	-0.52
<b>P05121</b>	DP00320	Plasminogen activator inhibitor 1	402	AM	6.68	35.33	87.76	-0.059
<b>Q9UM07</b>	DP00321	Protein-arginine deiminase type-4	663	AM	6.15	40.3	82.14	-0.24
<b>P11309-2</b>	DP00322	Proto-oncogene serine/threonine-protein kinase Pim-1 [Isoform 2]	313	AM	5.72	43.06	95.88	-0.247
<b>O75469-1</b>	DP00323	Nuclear receptor subfamily 1 group I member 2 [Isoform 1A (1)(PRR1-A)]	434	AM	8.7	51.73	75.3	-0.395
<b>P09651-1</b>	DP00324	Heterogeneous nuclear ribonucleoprotein A1 [Isoform A1-B]	371	AM	9.26	40.01	33.07	-0.912
<b>O14832</b>	DP00327	Phytanoyl-CoA dioxygenase, peroxisomal	338	AM	8.71	43.5	78.43	-0.409



UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P21815</b>	DP00332	Bone sialoprotein 2	317	AM	4.12	50.08	34.83	-1.35
<b>P27797</b>	DP00333	Calreticulin	417	AM	4.29	47.56	59.14	-1.104
<b>Q00987-1</b>	DP00334	E3 ubiquitin-protein ligase Mdm2 [Isoform Mdm2]	491	AM	4.6	57.63	65.85	-0.781
<b>Q8N726-1</b>	DP00336	Cyclin-dependent kinase inhibitor 2A [Isoform 4]	173	NON-AM	12.15	61.29	71.21	-0.433
<b>B0B1U2</b>	DP00338	Histo-blood group B transferase	354	AM	9.04	49.36	91.61	-0.068
<b>P16442</b>	DP00339	Histo-blood group A transferase	354	AM	9.14	48.64	92.43	-0.079
<b>Q9Y6Q9-3</b>	DP00343	Nuclear receptor coactivator 3 [Isoform 3, variant]	1412	AM	7	58.9	59.48	-0.748
<b>P02649</b>	DP00355	Apolipoprotein E	317	AM	5.65	44.88	87.16	-0.596
<b>Q99967-1</b>	DP00356	Cbp/p300-interacting transactivator 2 [Isoform 1]	270	AM	6.42	51.33	46.33	-0.6
<b>P62328</b>	DP00357	Thymosin beta-4	44	NON-AM	5.02	70.51	40	-1.62
<b>Q15185</b>	DP00358	Prostaglandin E synthase 3	160	AM	4.35	51.6	51.75	-1.049
<b>O14576-1</b>	DP00360	Cytoplasmic dynein 1 intermediate chain 1 [Isoform 1]	645	AM	5	51.45	70.82	-0.607
<b>Q16222-1</b>	DP00363	UDP-N-acetylhexosamine pyrophosphorylase [Isoform AGX2]	552	AM	5.92	34.7	86.09	-0.314
<b>Q16222-2</b>	DP00363_A002	UDP-N-acetylhexosamine pyrophosphorylase [Isoform AGX1]	505	AM	5.96	34.75	87.25	-0.292
<b>Q15382</b>	DP00364	GTP-binding protein Rheb	184	AM	5.65	33.42	94.78	-0.094

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P53041</b>	DP00365	Serine/threonine protein phosphatase 5	499	AM	5.88	37.25	75.27	-0.514
<b>Q9NR00</b>	DP00372	Uncharacterized protein C8orf4	106	AM	10.14	42.83	78.3	-0.717
<b>Q9H2S9-1</b>	DP00376	Zinc finger protein Eos [Isoform 1]	585	NON-AM	6.39	57.46	63.95	-0.649
<b>P08047</b>	DP00378	Transcription factor Sp1	785	AM	6.94	52.09	69.52	-0.436
<b>P35869</b>	DP00381	Aryl hydrocarbon receptor	848	AM	5.94	50.52	71.53	-0.56
<b>P05091</b>	DP00383	Aldehyde dehydrogenase, mitochondrial	500	AM	5.69	31.65	82.32	-0.142
<b>P49841-1</b>	DP00385	Glycogen synthase kinase-3 beta [Isoform 1]	420	AM	8.98	29.32	81.05	-0.319
<b>P02647</b>	DP00386_C001	Truncated apolipoprotein A-I	242	NON-AM	5.27	41.92	83.06	-0.829
<b>Q09161</b>	DP00392	Nuclear cap-binding protein subunit 1	790	AM	5.99	51.15	88.96	-0.328
<b>P52298-1</b>	DP00393	Nuclear cap-binding protein subunit 2 [Isoform 1]	156	AM	8.34	33.04	60.06	-0.908
<b>O00204</b>	DP00404	Sulfotransferase family cytosolic 2B member 1 [Isoform 1 (SULT2B1b)(B)]	365	AM	5.24	53.96	68.66	-0.545
<b>O00204-2</b>	DP00404_A002	Sulfotransferase family cytosolic 2B member 1 [Isoform 2 (SULT2B1a)(A)]	350	AM	6.1	55.85	67.14	-0.528
<b>Q04637</b>	DP00406	Eukaryotic translation initiation factor 4 gamma 1 [Isoform A]	1599	AM	5.25	61.36	72.83	-0.65
<b>P63000-1</b>	DP00408	Ras-related C3 botulinum toxin substrate 1 [Isoform A (Rac1A)]	192	AM	8.77	33.6	95.42	-0.101
<b>P78504</b>	DP00418	Protein jagged-1	1218	AM	5.7	43.87	52.13	-0.6

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q96T21</b>	DP00420	SECIS-binding protein 2	854	AM	8.31	57.33	68.75	-0.746
<b>P53350</b>	DP00428	Serine/threonine-protein kinase PLK1	603	AM	9.09	42.89	90.4	-0.323
<b>P43351</b>	DP00437	DNA repair protein RAD52 homolog	418	AM	8.49	47.64	67.68	-0.735
<b>Q14191</b>	DP00443	Werner syndrome ATP-dependent helicase	1432	AM	5.96	54.65	85.02	-0.441
<b>P02489</b>	DP00444	Alpha-crystallin A chain	173	AM	5.77	51.94	72.08	-0.49
<b>P02511</b>	DP00445	Alpha-crystallin B chain	175	NON-AM	6.76	<b>50.1</b>	<b>75.14</b>	<b>-0.531</b>
<b>Q9Y613</b>	DP00448	FH1/FH2 domain-containing protein 1	1164	AM	6.01	59.45	87.87	-0.3
<b>Q9Y3D6</b>	DP00457	Mitochondrial fission 1 protein	152	AM	8.84	27.51	100.07	-0.233
<b>Q07960</b>	DP00459	Rho GTPase-activating protein 1	439	AM	5.85	42.55	94.37	-0.367
<b>Q92838</b>	DP00460	Ectodysplasin-A [Isoform 1]	391	AM	8.54	47.37	67.62	-0.492
<b>Q92838-3</b>	DP00460_A003	Ectodysplasin-A [Isoform 3]	389	AM	8.69	47.57	67.22	-0.496
<b>Q13485</b>	DP00464	Mothers against decapentaplegic homolog 4	552	AM	6.5	51.12	75.49	-0.388
<b>P04156</b>	DP00466	Major prion protein	253	AM	9.13	43.11	52.37	-0.567
<b>Q01484</b>	DP00467	Ankyrin-2	1839	AM	5.83	45.69	86.34	-0.393
<b>P25963</b>	DP00468	NF-kappa-B inhibitor alpha	317	AM	4.57	43.29	88.61	-0.452

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P50542-3</b>	DP00472	Peroxisomal targeting signal 1 receptor [Isoform 3]	631	AM	4.45	48.27	79.16	-0.455
<b>Q49AN0</b>	DP00473	Cryptochrome-2	593	AM	8.66	50.43	79.19	-0.334
<b>Q9UBE0</b>	DP00485	SUMO-activating enzyme subunit 1	346	AM	5.17	36.74	85.9	-0.297
<b>Q9UBT2</b>	DP00486	SUMO-activating enzyme subunit 2	640	AM	4.57	43.29	88.61	-0.452
<b>P10275-1</b>	DP00492	Androgen receptor [Isoform 1 (AR-B)]	919	AM	6	56.54	68.12	-0.432
<b>Q6IBA2</b>	DP00501	PC4 protein	127	NON-AM	9.6	45.94	60.55	-1.094
<b>P11912-1</b>	DP00502	B-cell antigen receptor complex-associated protein alpha chain [Isoform 1 (Long)]	226	AM	4.87	43.03	87.96	-0.235
<b>P40259-1</b>	DP00503	B-cell antigen receptor complex-associated protein beta chain [Isoform Long]	229	AM	5.67	46.32	86.03	-0.337
<b>P52434</b>	DP00504	DNA-directed RNA polymerases I, II, and III subunit RPABC3	150	AM	4.5	31.36	87.07	-0.336
<b>P04234</b>	DP00505	T-cell surface glycoprotein CD3 delta chain	171	AM	5.3	18.8	100.88	-0.037
<b>P07766</b>	DP00506	T-cell surface glycoprotein CD3 epsilon chain	207	AM	6.32	42.85	77.63	-0.538
<b>P09693</b>	DP00508	T-cell surface glycoprotein CD3 gamma chain	182	AM	8.6	40.22	91.1	-0.348
<b>P30273</b>	DP00509	High affinity immunoglobulin epsilon receptor subunit gamma	86	AM	6.54	17.1	132.67	0.366
<b>O60356</b>	DP00510	Nuclear protein 1	82	NON-AM	9.98	72.76	41.83	-1.33
<b>Q76LA1</b>	DP00511	CSTB protein	98	AM	6.96	39.3	66.63	-0.546

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P02768-1</b>	DP00515	Serum albumin [Isoform 1]	609	AM	5.92	39.07	77.57	-0.354
<b>P52179-1</b>	DP00517	Myomesin-1	1685	AM	6.5	43.19	72.38	-0.551
<b>P25054-1</b>	DP00519	Adenomatous polyposis coli protein [Isoform Long]	2843	AM	7.92	66.31	61.65	-0.89
<b>Q969F2-1</b>	DP00520	Protein naked cuticle homolog 2 [Isoform 1]	451	AM	7.83	66.36	50.67	-1.156
<b>O95997</b>	DP00521	Securin	202	AM	6.18	64.9	79.65	-0.409
<b>O43236-1</b>	DP00537	Septin-4 [Isoform 1 (PNUTL2a)]	478	AM	5.77	48	69.92	-0.786
<b>P51608</b>	DP00539	Methyl-CpG-binding protein 2 [Isoform A (Beta)]	486	AM	9.95	61.06	52.59	-1.078
<b>Q14061</b>	DP00543	Cytochrome c oxidase copper chaperone	63	NON-AM	6.8	77.97	63.65	-0.589
<b>Q9NX55-1</b>	DP00546	Huntingtin-interacting protein K [Isoform 1]	175	AM	5.35	63.3	70.34	-0.811
<b>P16860</b>	DP00551	Natriuretic peptides B	134	AM	10.34	72.75	81.49	-0.486
<b>Q9NZ94-1</b>	DP00553	Neuroigin-3 [Isoform 1 (HNL3s)]	848	AM	5.94	42.41	79.91	-0.326
<b>Q9C000-1</b>	DP00554	NACHT, LRR and PYD domains-containing protein 1 [Isoform 1 (NAC beta)(DEFCAP-L)]	1473	AM	6.39	49.16	90.37	-0.263
<b>Q16143</b>	DP00555	Beta-synuclein	134	AM	4.41	48.74	67.76	-0.537
<b>P49354-1</b>	DP00558	Protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha [Isoform 1]	379	AM	4.96	51.03	81.82	-0.754
<b>Q9Y296</b>	DP00562	Trafficking protein particle complex subunit 4	219	AM	5.84	22.8	96.67	-0.013

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q66SS1</b>	DP00568	Paired box 6	422	AM	9.45	53.42	60.97	-0.703
<b>Q96IH6</b>	DP00569	EMILIN1 protein	347	AM	4.77	47.18	84.38	-0.336
<b>P09132</b>	DP00570	Signal recognition particle 19 kDa protein	144	AM	9.87	47.92	70.49	-0.697
<b>P39900</b>	DP00571	Macrophage metalloelastase	470	AM	8.75	25.91	73.23	-0.383
<b>P11831</b>	DP00574	Serum response factor	508	AM	7.83	44.73	68.01	-0.25
<b>P54252-1</b>	DP00576	Ataxin-3 [Isoform 1]	376	AM	4.86	64.61	77.07	-0.668
<b>P11233</b>	DP00581	Ras-related protein Ral-A	206	AM	6.66	39.88	75.73	-0.621
<b>P48539</b>	DP00592	Purkinje cell protein 4	62	AM	6.21	48.77	33.23	-1.332
<b>Q16595</b>	DP00607	Frataxin, mitochondrial	210	NON-AM	8.8	44.85	81.33	-0.465
<b>Q13573</b>	DP00608	SNW domain-containing protein 1	536	AM	9.52	51.45	59.22	-1.146
<b>P30291</b>	DP00611	Wee1-like protein kinase	646	AM	6.33	66.67	66.66	-0.683
<b>P02788</b>	DP00616	Lactotransferrin	710	AM	8.5	44.28	74.62	-0.337
<b>P60896</b>	DP00617	26S proteasome complex subunit DSS1	70	NON-AM	3.81	54.67	51.57	-1.36
<b>P00736</b>	DP00621	Complement C1r subcomponent	705	AM	5.82	42.9	68.3	-0.51
<b>P35638</b>	DP00624	DNA damage-inducible transcript 3	169	NON-AM	4.61	88.08	57.81	-1.147

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P15309-1</b>	DP00628	Prostatic acid phosphatase [Isoform 1]	386	AM	5.83	48.4	85.62	-0.301
<b>O76070</b>	DP00630	Gamma-synuclein	127	NON-AM	4.89	43.11	69.76	-0.53
<b>Q01844</b>	DP00632	RNA-binding protein EWS [Isoform EWS]	656	AM	9.37	49.64	22.53	-1.117
<b>Q09472</b>	DP00633	Histone acetyltransferase p300	2414	AM	8.81	67.04	59.74	-0.725
<b>P01270</b>	DP00637	Parathyroid hormone	115	NON-AM	9.72	39.02	90.7	-0.43
<b>P51671</b>	DP00641	Eotaxin	97	AM	9.98	38.2	87.63	-0.172
<b>O43521-2</b>	DP00643-A002	Bcl-2-like protein 11 [Isoform BimL]	138	AM	8.55	95.26	63.04	-0.77
<b>P31025</b>	DP00647	Lipocalin-1	176	AM	5.39	27.2	91.99	-0.241
<b>P00441</b>	DP00652	Superoxide dismutase [Cu-Zn]	154	AM	5.7	21.62	78.44	-0.344
<b>P32004-1</b>	DP00666	Neural cell adhesion molecule L1 [Isoform 1]	1257	AM	5.84	42.76	78.46	-0.449
<b>Q29983</b>	DP00670	MHC class I polypeptide-related sequence A	383	AM	6.49	49.59	74.07	-0.429
<b>Q92886</b>	DP00672	Neurogenin-1	237	AM	6.9	80.76	68.95	-0.589
<b>P11171</b>	DP00678	Protein 4.1	864	AM	5.42	45.33	73.83	-0.77
<b>Q99801</b>	DP00683	Homeobox protein Nkx-3.1	234	AM	9.23	61.09	61.41	-0.971
<b>P18754</b>	DP00691	Regulator of chromosome condensation	421	AM	7.18	34.88	77.93	-0.258

UniProt ID	DisProt ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q8N488</b>	DP00694	RING1 and YY1-binding protein	228	AM	9.66	49.98	41.93	-1.16
<b>Q9Y258</b>	DP00696	C-C motif chemokine 26	94	AM	10.22	19.56	82.98	-0.157
<b>P46937</b>	DP00702	Yorkie homolog	504	AM	5	65.51	59.29	-0.756
<b>P24522</b>	DP00704	Growth arrest and DNA damage-inducible protein GADD45 alpha	165	AM	4.36	36.25	100.48	-0.144
<b>Q15517</b>	DP00706	Corneodesmosin	529	AM	8.69	57.49	51.4	-0.343
<b>Q9UER7</b>	DP00707	Death domain-associated protein 6	740	AM	4.8	73.37	69.14	-0.848
<b>Q9Y3M2</b>	DP00709	Protein chibby homolog 1	126	AM	9.09	57.82	79.68	-0.823



**Table S2 Structured proteins in the dataset.** Accession codes, name, sequence length and the computed physicochemical properties are listed.

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q01082</b>	1BKR	Spectrin beta chain, non-erythrocytic 1	2364	AM	8.04	8.04	82.39	-0.505
<b>O14727</b>	1CY5	Apoptotic protease-activating factor 1	1248	AM	5.74	42.48	92.58	-0.397
<b>P50591</b>	1DG6	Tumor necrosis factor ligand superfamily member 10	281	AM	8.87	44.16	73.51	-0.629
<b>P78310</b>	1EAJ	Coxsackievirus and adenovirus receptor	365	AM	5.31	33.52	94.37	-0.338
<b>Q9HB21</b>	1EAZ	Pleckstrin homology domain-containing family A member 1	404	AM	8.91	53.58	67.04	-0.568
<b>P01270</b>	1ET1	Parathyroid hormone	115	NON-AM	9.3	62.67	94.41	-0.671
<b>P02766</b>	1F86	Transthyretin	147	AM	5.26	33.22	81.39	-0.094
<b>P13631</b>	1FCY	Retinoic acid receptor gamma	454	AM	5.74	49.98	97.2	-0.102
<b>P08670</b>	1GK7	Vimentin	466	AM	4.6	38.23	110	-0.667
<b>P11717</b>	1GP0	Cation-independent mannose-6-phosphate receptor	2491	NON-AM	8.12	39.89	67.48	-0.274
<b>P10153</b>	1GQV	Non-secretory ribonuclease	161	AM	9.2	57.66	58.44	-0.65
<b>P07320</b>	1H4A	Gamma-crystallin D	174	AM	6.71	50.72	64.22	-0.812
<b>P30043</b>	1HDO	Flavin reductase (NADPH)	206	AM	7.13	43.1	92.23	-0.06
<b>Q16674</b>	1I1J	Melanoma-derived growth regulatory protein	131	AM	8.83	25.34	71.3	-0.34
<b>P35269</b>	1I27	General transcription factor IIF subunit 1	517	NON-AM	10.07	43.85	81.37	-0.738

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>O95071</b>	1I2T	E3 ubiquitin-protein ligase UBR5	2799	AM	6.07	32.27	120.16	0.159
<b>P08519</b>	1I71	Apolipoprotein(a)	4548	NON-AM	6.29	41.17	25.78	-1.017
<b>P02545</b>	1IFR	Prelamin-A/C	664	AM	8.9	28.93	71.65	-0.584
<b>P41212</b>	1JI7	Transcription factor ETV6	452	AM	6.92	49.03	82.25	-0.985
<b>P61626</b>	1JSF	Lysozyme C	148	AM	9.28	32.13	69.85	-0.485
<b>O00299</b>	1K0M	Chloride intracellular channel protein 1	241	AM	5.09	35.99	88.63	-0.271
<b>P08263</b>	1K3Y	Glutathione S-transferase A1	222	AM	8.99	38.18	99.77	-0.292
<b>P03989</b>	1K5N	HLA class I histocompatibility antigen, B-27 alpha chain	362	AM	5.72	36.34	64.13	-0.817
<b>O14936</b>	1KGD	Peripheral plasma membrane protein CASK	926	AM	5.79	42.18	89.94	-0.493
<b>P52565</b>	1KMT	Rho GDP-dissociation inhibitor 1	204	NON-AM	8.77	14.81	79.43	-0.404
<b>P00374</b>	1KMV	Dihydrofolate reductase	187	AM	7.01	27.25	83.23	-0.477
<b>P14598</b>	1KQ6	Neutrophil cytosol factor 1	390	NON-AM	8.87	51.23	77.45	-0.446
<b>P12111</b>	1KTH	Collagen alpha-3(VI) chain	3177	AM	6.43	42.63	40.34	-0.8
<b>P09651</b>	1L3K	Heterogeneous nuclear ribonucleoprotein A1	372	AM	7.82	49.16	62.06	-0.773
<b>P22749</b>	1L9L	Granulysin	145	AM	10.83	47.67	69.73	-0.747

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P07360</b>	1LF7	Complement component C8 gamma chain	202	AM	9.01	42.57	83.63	-0.21
<b>P06239</b>	1LKK	Tyrosine-protein kinase Lck	509	AM	6.47	45.98	70.91	-0.646
<b>P37173</b>	1M9Z	TGF-beta receptor type-2	567	AM	4.61	55.42	55.32	-0.467
<b>P11215</b>	1MF7	Integrin alpha-M	1152	AM	8.87	22.2	83.87	-0.398
<b>P51687</b>	1MJ4	Sulfite oxidase, mitochondrial	545	NON-AM	4.95	45.15	87.93	-0.178
<b>P15692</b>	1MKK	Vascular endothelial growth factor A	232	AM	5.49	68.35	68.96	-0.425
<b>P54577</b>	1N3L	Tyrosine--tRNA ligase, cytoplasmic	528	AM	6.55	28.76	91.99	-0.394
<b>Q9Y286</b>	1NKO	Sialic acid-binding Ig-like lectin 7	467	NON-AM	8.49	34.67	54.62	-0.789
<b>Q9UBU9</b>	1OAI	Nuclear RNA export factor 1	619	AM	5.13	56.47	46.03	-0.735
<b>P53674</b>	1OKI	Beta-crystallin B1	252	AM	6.99	39.54	63.14	-0.59
<b>Q13526</b>	1PIN	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1	163	AM	8.95	64.63	53.93	-0.877
<b>P31151</b>	1PSR	Protein S100-A7	101	NON-AM	6.3	34.22	59.6	-0.777
<b>O95750</b>	1PWA	Fibroblast growth factor 19	216	NON-AM	6.4	56.89	81.85	-0.37
<b>P35247</b>	1PWB	Pulmonary surfactant-associated protein D	375	AM	5.12	35.07	67.29	-0.451
<b>Q03154</b>	1Q7L	Aminoacylase-1	408	AM	5.89	37.55	80.77	-0.216

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P05451</b>	1QDD	Lithostathine-1-alpha	166	AM	5.32	20.99	58.89	-0.497
<b>Q16775</b>	1QH5	Hydroxyacylglutathione hydrolase, mitochondrial	308	AM	6.93	26.37	77.96	-0.377
<b>P41182</b>	1R29	B-cell lymphoma 6 protein	706	AM	7.02	53.75	102.83	0.028
<b>O00560</b>	1R6J	Syntenin-1	298	AM	7.02	36.12	95.12	0.072
<b>P84080</b>	1R8S	ADP-ribosylation factor 1	181	AM	5.47	32.07	89.81	-0.421
<b>P05230</b>	1RG8	Fibroblast growth factor 1	155	AM	8.17	38.07	71.44	-0.783
<b>P02787</b>	1RYO	Serotransferrin	698	AM	7.16	34.05	73.12	-0.324
<b>O95881</b>	1SEN	Thioredoxin domain-containing protein 12	172	AM	5.66	42.67	66.04	-0.685
<b>P42126</b>	1SG4	Enoyl-CoA delta isomerase 1, mitochondrial	302	AM	6.01	38.73	101.69	0.01
<b>Q9NP99</b>	1SMO	Triggering receptor expressed on myeloid cells 1	234	AM	7.24	43.46	84.29	-0.543
<b>Q9Y3R4</b>	1SO7	Sialidase-2	380	AM	6.37	45.88	76.68	-0.369
<b>P06401</b>	1SR7	Progesterone receptor	933	AM	8.54	55.23	108.73	0.016
<b>Q14019</b>	1T3Y	Coactosin-like protein	142	AM	5.51	27.19	74.04	-0.43
<b>P05305</b>	1T7H	Endothelin-1	212	AM	8.2	92.65	37.78	-0.194
<b>O76074</b>	1TBF	cGMP-specific 3',5'-cyclic phosphodiesterase	875	AM	6.09	37.08	94.24	-0.259

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P05019</b>	1TGR	Insulin-like growth factor I	195	NON-AM	5.13	58.19	69.42	-0.05
<b>Q15772</b>	1U2H	Striated muscle preferentially expressed protein kinase	3267	NON-AM	8.99	50.95	70.91	-0.638
<b>P31749</b>	1UNQ	RAC-alpha serine/threonine-protein kinase	480	AM	5.78	46.38	63.92	-0.844
<b>P15121</b>	1US0	Aldose reductase	316	AM	6.52	37.19	93.13	-0.251
<b>P50552</b>	1USE	Vasodilator-stimulated phosphoprotein	380	AM	5.35	95.06	88.67	-1.009
<b>P07900</b>	1UYL	Heat shock protein HSP 90-alpha	732	AM	4.63	37.59	82.67	-0.47
<b>Q7Z589</b>	1UZ3	Protein EMSY	1322	NON-AM	9.17	60.01	99.41	-0.435
<b>P35555</b>	1UZK	Fibrillin-1	2871	NON-AM	4.24	30.45	61.91	-0.259
<b>Q14315</b>	1V05	Filamin-C	2725	AM	9.18	7.4	65.73	-0.246
<b>O15530</b>	1W1H	3-phosphoinositide-dependent protein kinase 1	556	NON-AM	8.92	54.08	76.16	-0.809
<b>P16152</b>	1WMA	Carbonyl reductase [NADPH] 1	277	NON-AM	8.55	43.77	91.09	-0.219
<b>P56945</b>	1WYX	Breast cancer anti-estrogen resistance protein 1	870	NON-AM	6.07	29.7	98.84	-0.3
<b>Q9UKM7</b>	1X9D	Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	699	AM	6.13	42.91	81.17	-0.445
<b>Q16625</b>	1XAW	Occludin	522	NON-AM	5.73	64.79	52.29	-1.501
<b>P15374</b>	1XD3	Ubiquitin carboxyl-terminal hydrolase isozyme L3	230	AM	5.01	39.83	87.57	-0.426

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P55265</b>	1XMK	Double-stranded RNA-specific adenosine deaminase	1226	AM	8.09	33.33	87.72	-0.505
<b>P39900</b>	1Y93	Macrophage metalloelastase	470	AM	6.1	13.23	73.02	-0.293
<b>P35285</b>	1Z0J	Ras-related protein Rab-22A	194	AM	5.08	46.64	92.88	-0.284
<b>Q8NBK3</b>	1Z70	Sulfatase-modifying factor 1	374	AM	6.24	43.59	52.73	-0.715
<b>Q9UIJ7</b>	1ZD8	GTP:AMP phosphotransferase AK3, mitochondrial	227	AM	9.15	33.05	92.78	-0.352
<b>Q14145</b>	1ZGK	Kelch-like ECH-associated protein 1	624	AM	6.61	41.52	74.64	-0.335
<b>P59666</b>	1ZMI	Neutrophil defensin 3	94	AM	8.87	56.41	67.59	0.324
<b>P61978</b>	1ZZK	Heterogeneous nuclear ribonucleoprotein K	463	AM	8.28	17.62	92.8	-0.368
<b>Q9HB71</b>	2A26	Calcyclin-binding protein	228	NON-AM	9.36	65.62	80	-1.082
<b>P05089</b>	2AEB	Arginase-1	322	AM	6.78	26.75	95.31	-0.164
<b>P09104</b>	2AKZ	Gamma-enolase	434	AM	5.17	34.02	91.57	-0.234
<b>Q00987</b>	2AXI	E3 ubiquitin-protein ligase Mdm2	491	AM	6.42	36.12	84.08	-0.522
<b>P53582</b>	2B3H	Methionine aminopeptidase 1	386	AM	6.71	45.77	75.53	-0.428
<b>Q8NBZ7</b>	2B69	UDP-glucuronic acid decarboxylase 1	420	AM	6.22	39.69	80.38	-0.453
<b>P12694</b>	2BFD	2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial	445	AM	5.78	41.39	78.52	-0.284

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q96T60</b>	2BRF	Bifunctional polynucleotide phosphatase/kinase	521	NON-AM	4.91	33.25	91.18	-0.309
<b>P00734</b>	2BVR	Prothrombin	622	AM	6.3	36.95	75.95	-0.544
<b>P00734</b>	2BVS	Prothrombin	622	AM	6.3	34.02	76.44	-0.526
<b>Q8N4Q0</b>	2C0C	Zinc-binding alcohol dehydrogenase domain-containing protein 2	377	AM	6.85	31.91	90.44	0.075
<b>Q99759</b>	2C60	Mitogen-activated protein kinase kinase kinase 3	626	AM	6.34	62.15	87.75	-0.664
<b>Q9BY32</b>	2CAR	Inosine triphosphate pyrophosphatase	194	AM	5.51	32.6	83.62	-0.174
<b>P07108</b>	2CB8	Acyl-CoA-binding protein	87	AM	6.15	37.01	59.54	-0.898
<b>P01112</b>	2CE2	GTPase HRas	189	AM	4.89	39.72	84.52	-0.413
<b>P02751</b>	2CG7	Fibronectin	2386	NON-AM	6.91	33.23	44.44	-0.737
<b>O43639</b>	2CIA	Cytoplasmic protein NCK2	380	AM	5.83	36.61	74.16	-0.576
<b>O60870</b>	2CKK	DNA/RNA-binding protein KIN17	393	AM	9.32	31.89	98.98	-0.513
<b>P42574</b>	2DKO	Caspase-3	277	AM	6.18	38.87	73.99	-0.394
<b>P30622</b>	2E3H	CAP-Gly domain-containing linker protein 1	1438	NON-AM	9.47	16.38	76.89	-0.186
<b>Q9Y5P4</b>	2E3N	Collagen type IV alpha-3-binding protein	624	AM	6.07	32.77	86.67	-0.299
<b>Q8WWI1</b>	2EAQ	LIM domain only protein 7	1683	AM	4.67	25.16	71.56	-0.421

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P07996</b>	2ERF	Thrombospondin-1	1170	AM	6.95	35.76	93.25	-0.229
<b>Q8WTS6</b>	2F69	Histone-lysine N-methyltransferase SETD7	366	AM	5.07	30.76	68.09	-0.487
<b>P30533</b>	2FCW	Alpha-2-macroglobulin receptor-associated protein	357	AM	5.39	50.69	61.48	-0.92
<b>P05067</b>	2FMA	Amyloid beta A4 protein	770	NON-AM	5.89	23	69.32	-0.258
<b>Q86YW5</b>	2FRG	Trem-like transcript 1 protein	311	NON-AM	6.1	76.94	99.25	-0.027
<b>O95989</b>	2FVV	Diphosphoinositol polyphosphate phosphohydrolase 1	172	AM	6.14	61.51	71.75	-0.576
<b>Q12888</b>	2G3R	Tumor suppressor p53-binding protein 1	1972	AM	6.16	31.32	82.36	-0.441
<b>P63279</b>	2GRR	SUMO-conjugating enzyme UBC9	158	AM	9.04	45.98	84.02	-0.205
<b>P06734</b>	2H2T	Low affinity immunoglobulin epsilon Fc receptor	321	AM	5.5	51.05	50.17	-0.671
<b>Q96RT1</b>	2H3L	Protein LAP2	1412	AM	5.89	31.77	89.81	-0.322
<b>P23467</b>	2HC1	Receptor-type tyrosine-protein phosphatase beta	1997	AM	7.63	39.82	83.99	-0.51
<b>O00213</b>	2HO2	Amyloid beta A4 precursor protein-binding family B member 1	710	AM	5.38	94.1	34.58	-0.923
<b>Q7KZF4</b>	2HQX	Staphylococcal nuclease domain-containing protein 1	910	AM	4.94	38.57	78.41	-0.418
<b>P07355</b>	2HYV	Annexin A2	339	AM	8.21	38.93	88.67	-0.561
<b>P26368</b>	2HZC	Splicing factor U2AF 65 kDa subunit	475	AM	6.36	42.9	80.8	-0.093



UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P08567</b>	2I5F	Pleckstrin	350	AM	8.62	52.89	81.47	-0.514
<b>Q8TCD5</b>	2I7D	5'(3')-deoxyribonucleotidase, cytosolic type	201	AM	6.19	55.08	93.37	-0.418
<b>Q9Y279</b>	2ICC	V-set and immunoglobulin domain-containing protein 4	399	NON-AM	6.06	41.27	84.12	-0.581
<b>P06239</b>	2IIM	Tyrosine-protein kinase Lck	509	NON-AM	4.63	52.62	78.71	-0.448
<b>Q16611</b>	2IMS	Bcl-2 homologous antagonist/killer	211	AM	5.13	46.54	84.44	-0.175
<b>Q9Y6M4</b>	2IZR	Casein kinase I isoform gamma-3	447	AM	9.06	33.86	79.18	-0.543
<b>P13861</b>	2IZX	cAMP-dependent protein kinase type II-alpha regulatory subunit	404	AM	4.78	50.02	111.1	-0.209
<b>P42574</b>	2J32	Caspase-3	277	AM	6.25	37.59	73.41	-0.404
<b>P49419</b>	2J6L	Alpha-aminoadipic semialdehyde dehydrogenase	539	AM	6.09	36.05	91.84	-0.055
<b>P13987</b>	2J8B	CD59 glycoprotein	128	AM	5.18	31.27	55.57	-0.544
<b>O00241</b>	2JJU	Signal-regulatory protein beta-1	398	AM	8.37	27.89	71.42	-0.46
<b>P00742</b>	2JKH	Coagulation factor X	488	AM	7.88	39	65.54	-0.485
<b>P60022</b>	2NLS	Beta-defensin 1	68	AM	8.31	34.76	46.11	-0.261
<b>Q9UJM8</b>	2NZL	Hydroxyacid oxidase 1	370	AM	7.15	35.58	96.51	-0.14
<b>Q99714</b>	2O23	3-hydroxyacyl-CoA dehydrogenase type-2	261	AM	8.1	19.26	99.43	0.211

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q9BX66</b>	2O9S	Sorbin and SH3 domain-containing protein 1	1292	AM	6.4	13.85	78.51	-0.445
<b>P04080</b>	2OCT'	Cystatin-B	98	AM	8.62	44.6	66.63	-0.549
<b>P52594</b>	2OLM	Arf-GAP domain and FG repeat-containing protein 1	562	NON-AM	9.28	61.05	60.5	-0.751
<b>O60885</b>	2OSS	Bromodomain-containing protein 4	1362	AM	7.67	41.59	76.77	-0.674
<b>Q9H0H5</b>	2OVJ	Rac GTPase-activating protein 1	632	NON-AM	5.96	35.66	101.34	-0.013
<b>Q9UHI6</b>	2OXC	Probable ATP-dependent RNA helicase DDX20	824	AM	6.99	46.39	109.43	0.1
<b>P31153</b>	2P02	S-adenosylmethionine synthase isoform type-2	395	AM	6.02	38.34	88.86	-0.282
<b>P04179</b>	2P4K	Superoxide dismutase [Mn], mitochondrial	222	AM	6.92	38.02	82.83	-0.487
<b>O43252</b>	2PEZ	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 1	624	AM	4.94	41.42	87.09	-0.258
<b>O76064</b>	2PIE	E3 ubiquitin-protein ligase RNF8	485	AM	5.42	45.26	82.62	-0.485
<b>P62942</b>	2PPN	Peptidyl-prolyl cis-trans isomerase FKBP1A	108	NON-AM	8.08	42.66	72.8	-0.433
<b>Q9H7Z6</b>	2PQ8	Histone acetyltransferase KAT8	458	AM	8.9	35.54	83.38	-0.436
<b>P61836</b>	2PUY	Histone H3	136	NON-AM	8.4	30.93	72	-0.382
<b>O60344</b>	2PXX	Endothelin-converting enzyme 2	883	AM	5.71	39.6	82.09	-0.099
<b>O15392</b>	2QFA	Baculoviral IAP repeat-containing protein 5	142	AM	5.38	55.71	76.65	-0.617

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q9Y6W3</b>	2QFE	Calpain-7	813	AM	8.85	42.89	73.72	-0.392
<b>Q15691</b>	2QJZ	Microtubule-associated protein RP/EB family member 1	268	AM	8.7	21.93	78.54	-0.363
<b>P29320</b>	2QOL	Ephrin type-A receptor 3	983	AM	6.5	26.89	86.76	-0.265
<b>Q9NWW6</b>	2QT1	Nicotinamide riboside kinase 1	199	AM	5.22	61.6	78.12	-0.468
<b>P49862</b>	2QXI	Kallikrein-7	253	NON-AM	9.03	27.68	73.04	-0.332
<b>Q15691</b>	2R8U	Microtubule-associated protein RP/EB family member 1	268	AM	5.02	35.99	81.16	-0.427
<b>P24821</b>	2RB8	Tenascin	2201	NON-AM	4.76	40.66	72.21	-0.661
<b>Q12830</b>	2RI7	Nucleosome-remodeling factor subunit BPTF	3046	AM	5.2	49.63	64.48	-0.595
<b>Q99497</b>	2RK3	Protein DJ-1	189	AM	6.5	26.75	97.06	-0.104
<b>P54619</b>	2UUF	5'-AMP-activated protein kinase subunit gamma-1	331	AM	6.26	34.73	76.69	-0.526
<b>P54619</b>	2UV4	5'-AMP-activated protein kinase subunit gamma-1	331	AM	6.09	31.56	112.7	0.066
<b>P49748</b>	2UXW	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	655	AM	6.62	35.09	87.91	-0.032
<b>Q96F86</b>	2VC8	Enhancer of mRNA-decapping protein 3	508	AM	5	26.22	92.74	-0.027
<b>P07311</b>	2VH7	Acylphosphatase-1	99	NON-AM	9.3	8.74	80.61	-0.557
<b>O14544</b>	2VIF	Suppressor of cytokine signaling 6	535	AM	6.18	38.87	68.55	-0.589

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q9NVD7</b>	2VZC	Alpha-parvin	372	AM	5.53	22.63	98.09	-0.173
<b>P69905</b>	2W72	Hemoglobin subunit alpha	142	AM	8.11	9.04	88.55	-0.032
<b>P62158</b>	2W73	Calmodulin	149	AM	4.32	31.59	70.54	-0.561
<b>Q15605</b>	2W7A	ORF1 codes for a 40 kDa product	338	AM	10.77	38.33	88.7	-0.696
<b>Q13427</b>	2WFI	Peptidyl-prolyl cis-trans isomerase G	754	AM	7.76	25.49	64.19	-0.365
<b>O15553</b>	2WL1	Pyrin	781	AM	8.69	43.65	85.71	-0.123
<b>O00602</b>	2WNP	Ficolin-1	326	AM	5.84	21.8	58.94	-0.502
<b>P02745</b>	2WNV	Complement C1q subcomponent subunit A	245	AM	8.9	37.57	75.49	-0.179
<b>O75747</b>	2WWE	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit gamma	1545	AM	6.33	56.31	71.34	-0.575
<b>P0CH28</b>	2WWZ	Polyubiquitin-C	690	AM	5.85	42.23	90.6	-0.501
<b>P00441</b>	2WYT	Superoxide dismutase [Cu-Zn]	154	AM	5.7	20.37	78.3	-0.356
<b>P00558</b>	2WZB	Phosphoglycerate kinase 1	417	AM	8.3	26.25	90.48	-0.083
<b>Q7Z2W4</b>	2X5Y	Zinc finger CCCH-type antiviral protein 1	902	AM	8.27	44.17	68.67	-0.583
<b>P17931</b>	2XG3	Galectin-3	250	AM	9.42	20.22	97.39	-0.242
<b>Q6P1N9</b>	2XIO	Putative deoxyribonuclease TATDN1	297	AM	6.16	27.7	84.88	-0.31

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P05543</b>	2XN6	Thyroxine-binding globulin	415	AM	6.13	36.62	92.47	-0.019
<b>Q96FJ2</b>	2XQQ	Dynein light chain 2, cytoplasmic	89	AM	7.17	32.73	64.06	-0.57
<b>P07711</b>	2XU3	Cathepsin L1	333	AM	4.68	34.69	56.73	-0.53
<b>P23468</b>	2YD6	Receptor-type tyrosine-protein phosphatase delta	1912	AM	7.24	32.07	68.02	-0.654
<b>Q96FJ0</b>	2ZNR	AMSH-like protease	436	AM	6.5	34.3	97.42	-0.037
<b>Q9HCB6</b>	2ZOU	Spondin-1	807	AM	5.96	53.32	66.78	-0.561
<b>Q96CS7</b>	3AJ4	Pleckstrin homology domain-containing family B member 2	222	AM	7.01	70.48	66.16	-0.706
<b>P09960</b>	3B7S	Leukotriene A-4 hydrolase	611	AM	6.21	39.46	86.95	-0.302
<b>P22392</b>	3BBB	Nucleoside diphosphate kinase B	152	AM	8.82	31.1	86.42	-0.291
<b>Q9UDY8</b>	3BFO	Mucosa-associated lymphoid tissue lymphoma translocation protein 1	824	AM	5.72	43.75	90.88	-0.395
<b>O43293</b>	3BHY	Death-associated protein kinase 3	454	AM	5.84	59.54	93.71	-0.399
<b>P07602</b>	3BQP	Proactivator polypeptide	524	NON-AM	4.52	49.08	93.75	0.01
<b>P08581</b>	3BUX	Hepatocyte growth factor receptor	1390	AM	9.11	33.01	79.01	-0.406
<b>P21964</b>	3BWY	Catechol O-methyltransferase	271	NON-AM	5.25	25.99	97.94	-0.132
<b>Q14896</b>	3CX2	Myosin-binding protein C, cardiac-type	1274	NON-AM	6.45	19.73	78.43	-0.244

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P04637</b>	3D06	Cellular tumor antigen p53	393	AM	8.7	73.79	64.25	-0.538
<b>O95166</b>	3D32	Gamma-aminobutyric acid receptor-associated protein	117	AM	7.32	39.61	76.64	-0.554
<b>P00390</b>	3DK9	Glutathione reductase, mitochondrial	522	AM	7.72	41.47	86.88	-0.088
<b>P13497</b>	3EDH	Bone morphogenetic protein 1	986	AM	8.75	36.61	67.81	-0.51
<b>P61964</b>	3EMH	WD repeat-containing protein 5	334	AM	8.67	28.42	82.18	-0.266
<b>Q9NZ45</b>	3EW0	CDGSH iron-sulfur domain-containing protein 1	108	NON-AM	9.18	38.1	58.5	-0.856
<b>P11172</b>	3EWY	Uridine 5'-monophosphate synthase	480	AM	6.01	23.58	103.58	0.105
<b>P01308</b>	3EXX	Insulin	110	AM	5.39	11.59	85.88	0.218
<b>P21579</b>	3F04	Synaptotagmin-1	422	AM	5.58	29.77	82.38	-0.273
<b>Q99814</b>	3F1P	Endothelial PAS domain-containing protein 1	870	AM	6.08	44.91	71.97	-0.491
<b>P28907</b>	3F6Y	ADP-ribosyl cyclase 1	300	NON-AM	5.85	50.52	69.16	-0.56
<b>Q9NQR1</b>	3F9X	N-lysine methyltransferase SETD8	393	AM	8.53	44.92	79.32	-0.674
<b>P13051</b>	3FCI	Uracil-DNA glycosylase	313	AM	9.04	30.06	73.81	-0.489
<b>O15151</b>	3FDO	Protein Mdm4	490	AM	7.09	46.45	101.27	-0.1
<b>O15151</b>	3FEA	Protein Mdm4	490	AM	6.31	46.51	87.75	-0.025

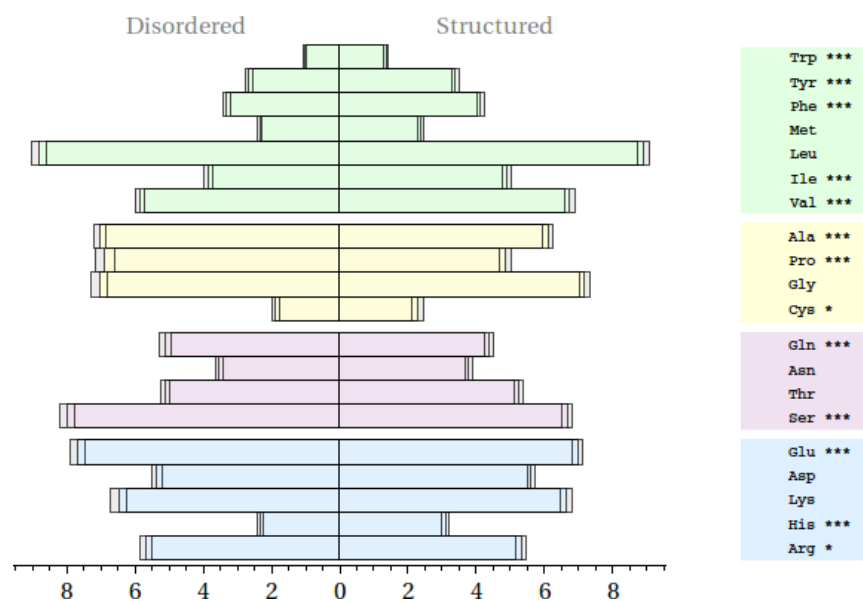
UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q9Y3Q0</b>	3FED	N-acetylated-alpha-linked acidic dipeptidase 2	740	AM	8.32	39.95	85.54	-0.342
<b>Q9Y259</b>	3FEG	Choline/ethanolamine kinase	395	AM	5.75	62.04	79.5	-0.466
<b>Q00059</b>	3FGH	Transcription factor A, mitochondrial	246	AM	8.9	42.41	43.73	-1.578
<b>Q15811</b>	3FIA	Intersectin-1	1721	AM	7.14	33.38	77.44	-0.296
<b>P16144</b>	3FSO	Integrin beta-4	1822	AM	6.49	32.77	85.45	-0.603
<b>Q9C005</b>	3G36	Protein dpy-30 homolog	99	NON-AM	9.16	61.73	108.18	-0.22
<b>Q15493</b>	3G4E	Regucalcin	299	AM	5.9	43.9	78.08	-0.315
<b>Q9BYX4</b>	3GA3	Interferon-induced helicase C domain-containing protein 1	1025	AM	8.82	37.52	82.03	-0.441
<b>Q9NX01</b>	3GIX	Thioredoxin-like protein 4B	149	AM	5.64	28.32	95.44	-0.146
<b>P62826</b>	3GJ0	GTP-binding nuclear protein Ran	216	AM	7.33	22.9	86.43	-0.279
<b>P53041</b>	3H63	Serine/threonine-protein phosphatase 5	499	AM	5.78	27.3	78.86	-0.386
<b>P00451</b>	3HNY	Coagulation factor VIII	2351	AM	9.23	35.34	75.97	-0.311
<b>Q9UNA0</b>	3HYG	A disintegrin and metalloproteinase with thrombospondin motifs 5	930	AM	5.83	46.6	90.54	-0.225
<b>O95396</b>	3I2V	Adenylyltransferase and sulfurtransferase MOCS3	460	NON-AM	6.4	41.1	111.34	-0.104
<b>P54652</b>	3I33	Heat shock-related 70 kDa protein 2	639	AM	6.33	38.08	83.07	-0.34

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q13526</b>	3I6C	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1	163	AM	7.25	55.42	61.14	-0.724
<b>P31947</b>	3IJ8	14-3-3 protein sigma	248	AM	6.29	23.94	68.17	-0.464
<b>P16442</b>	3IOH	Histo-blood group ABO system transferase	354	AM	8.58	45.88	85.94	-0.18
<b>P31947</b>	3IQU	14-3-3 protein sigma	248	AM	4.91	44.02	79.09	-0.536
<b>Q14623</b>	3IU1	Indian hedgehog protein	411	AM	8.57	46.79	88.22	-0.295
<b>P30046</b>	3IVV	D-dopachrome decarboxylase	118	AM	7.78	27.82	72.26	-0.394
<b>Q9BVM4</b>	3JUD	Gamma-glutamylaminecyclotransferase	153	AM	6.65	44.04	78.43	-0.482
<b>Q14676</b>	3K05	Mediator of DNA damage checkpoint protein 1	2089	AM	6.38	53.49	90.83	-0.078
<b>Q14623</b>	3K7I	Indian hedgehog protein	411	AM	9.38	30.75	61.02	-0.889
<b>P30046</b>	3KAN	D-dopachrome decarboxylase	118	AM	7.38	33.25	101.79	0.217
<b>P00918</b>	3KS3	Carbonic anhydrase 2	260	AM	6.87	21.68	76.46	-0.579
<b>P11940</b>	3KUS	Polyadenylate-binding protein 1	636	NON-AM	5.89	50.75	94.85	-0.177
<b>Q13009</b>	3KZD	T-lymphoma invasion and metastasis-inducing protein 1	1591	NON-AM	5.61	36.57	94.47	-0.26
<b>P55201</b>	3L42	Peregrin	1214	AM	9.07	51.4	87.69	-0.578
<b>Q9H8V3</b>	3L46	Protein ECT2	914	AM	5.12	50.44	60	-0.706



UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>Q7L1T6</b>	3LF5	Cytochrome b5 reductase 4	521	NON-AM	5.9	57.38	76.36	-0.505
<b>Q9HB90</b>	3LLU	Ras-related GTP-binding protein C	399	AM	6.08	37.65	80.56	-0.413
<b>Q12923</b>	3LNY	Tyrosine-protein phosphatase non-receptor type 13	2485	NON-AM	6.49	26.48	97.35	-0.211
<b>P10599</b>	3M9J	Thioredoxin	105	AM	4.81	23.3	75.14	-0.159
<b>Q9NZV6</b>	3MAO	Methionine-R-sulfoxide reductase B1	116	AM	8.73	50.39	49.24	-0.559
<b>Q96ES7</b>	3MEA	SAGA-associated factor 29 homolog	293	AM	6.75	48.62	70	-0.74
<b>P30044</b>	3MNG	Peroxiredoxin-5, mitochondrial	214	AM	7.8	17.62	91.27	-0.027
<b>Q16849</b>	3N01	Receptor-type tyrosine-protein phosphatase-like N	979	AM	5.46	38.82	112.81	0.038
<b>O43169</b>	3NER	Cytochrome b5 type B	146	AM	4.77	38.45	83.59	-0.666
<b>O75604</b>	3NHE	Ubiquitin carboxyl-terminal hydrolase 2	605	AM	8.62	43.1	76.32	-0.525
<b>Q01826</b>	3NZL	DNA-binding protein SATB1	763	NON-AM	9.27	49.95	54.1	-0.78
<b>P41222</b>	3O22	Prostaglandin-H2 D-isomerase	190	AM	8.76	42.47	59.07	-0.499
<b>Q92797</b>	3O2T	Symplekin	1274	AM	6.07	42.74	103.34	-0.235
<b>Q13451</b>	3O5Q	Peptidyl-prolyl cis-trans isomerase FKBP5	457	NON-AM	8.5	24.74	77.66	-0.407
<b>Q99816</b>	3OBQ	Tumor susceptibility gene 101 protein	390	AM	9.02	31.46	91.68	-0.139

UniProt ID	PDB ID	Protein name	Length	AM/NON-AM	pI	II	AI	GRAVY
<b>P15090</b>	3P6D	Fatty acid-binding protein, adipocyte	132	AM	7.92	22.38	76.26	-0.317
<b>Q9NXS2</b>	3PB6	GlutaminyI-peptide cyclotransferase-like protein	382	AM	7.87	65.02	101.06	-0.092
<b>Q92547</b>	3PD7	DNA topoisomerase 2-binding protein 1	1522	AM	7.35	44.92	67.38	-0.838
<b>Q99685</b>	3PE6	Monoglyceride lipase	303	AM	6.34	36.42	96.2	-0.001
<b>Q6ZUM4</b>	3PP2	Rho GTPase-activating protein 27	889	AM	10.15	38.26	73.23	-0.677
<b>P53350</b>	3Q1I	Serine/threonine-protein kinase PLK1	603	AM	6.66	47.12	85.47	-0.322
<b>P15090</b>	3Q6L	Fatty acid-binding protein, adipocyte	132	AM	7.84	23.36	74.21	-0.373
<b>Q92692</b>	3R0N	Poliovirus receptor-related protein 2	538	NON-AM	5.73	47.59	76.88	-0.285
<b>O15047</b>	3S8S	Histone-lysine N-methyltransferase SETD1A	1707	AM	9.56	18.88	91.18	-0.403



**Figure S1** Paired bar chart comparing the amino acid distribution in disordered and structured human proteins. Amino acids are categorized in four groups and shown bottom-up in the following order: charged (light blue), uncharged polar (light purple), special types (light yellow) and hydrophobic (light green). Significant variations are marked with asterisks.

**Table S3 P-values.**

Parameters	Disordered	Structured
pl acidic	0.52318	0.28916
pl basic	0.00015	0.00012
GRAVY	0.00004	0.18497
AI	0.00110	0.48286
II	0.00321	0.18456