

TABLE S1: List of human proteins with histidine enrichment in their propeptides

UniProt identifier	Name	Length propeptide	Histidine content	P(X>=k)	Cellular location	Propeptide type
P12821	Angiotensin-converting enzyme	74	6.76%	2.80E-02	Secreted, Cell membrane	C-terminus
O14672	ADAM10	194	7.73%	5.08E-05	Cell membrane, Endomembrane	N-terminus
O75078	ADAM11	202	4.46%	4.54E-02	Cell membrane	N-terminus
Q9Y3Q7	ADAM18	168	4.76%	4.15E-02	Cell membrane	N-terminus
Q9P0K1	ADAM22	197	6.09%	2.22E-03	Cell membrane	N-terminus
O75077	ADAM23	227	4.85%	1.69E-02	Cell membrane	N-terminus
Q9UKF2	ADAM30	171	4.68%	4.52E-02	Cell membrane	N-terminus
Q9BZ11	ADAM33	174	5.17%	2.01E-02	Cell membrane	N-terminus
O15204	ADAM-like protein decysin-1	175	6.29%	2.61E-03	Secreted	N-terminus
Q9H324	ADAM-TS10	208	5.29%	9.33E-03	Secreted	N-terminus
P58397	ADAM-TS12	215	6.05%	1.59E-03	Secreted	N-terminus
Q8TE57	ADAM-TS16	255	5.88%	9.60E-04	Secreted	N-terminus
Q8TE60	ADAM-TS18	237	5.91%	1.34E-03	Secreted	N-terminus
P59510	ADAM-TS20	232	4.74%	1.96E-02	Secreted	N-terminus
Q9UKP5	ADAM-TS6	223	8.52%	1.31E-06	Secreted	N-terminus
Q9P2N4	ADAM-TS9	269	4.09%	4.88E-02	Secreted	N-terminus
O95972	Bone morphogenetic protein 15	249	5.22%	5.55E-03	Secreted	N-terminus
P12643	Bone morphogenetic protein 2	259	5.79%	1.12E-03	Secreted	N-terminus
P12644	Bone morphogenetic protein 4	273	6.23%	2.38E-04	Secreted	N-terminus
P18075	Bone morphogenetic protein 7	263	5.70%	1.30E-03	Secreted	N-terminus
P55287	Cadherin-11	31	12.90%	5.36E-03	Cell membrane	N-terminus
Q13634	Cadherin-18	29	17.24%	4.82E-04	Cell membrane	N-terminus
P12830	Cadherin-1	132	6.06%	1.17E-02	Cell membrane, Endosome, Golgi apparatus	N-terminus
P14091	Cathepsin E	34	8.82%	4.29E-02	Endosome	N-terminus
P09668	Cathepsin H	85	8.24%	3.52E-03	Lysosome	N-terminus
P43235	Cathepsin K	99	6.06%	2.71E-02	Lysosome	N-terminus
P25774	Cathepsin S	98	8.16%	1.97E-03	Lysosome	N-terminus
Q6YHK3	CD109 antigen	25	12.00%	1.92E-02	Cell membrane	C-terminus
P0CG37	Cryptic protein	65	7.69%	1.70E-02	Cell membrane	C-terminus
Q14126	Desmoglein-2	26	11.54%	2.13E-02	Cell membrane	N-terminus
P12259	Coagulation factor V	836	3.59%	1.27E-02	Secreted	Internal
P02765	Alpha-2-HS-	40	12.50%	2.17E-03	Secreted	Internal

	glycoprotein					
P09958	Furin	83	6.02%	4.28E-02	Golgi apparatus, Cell membrane	N-terminus
O60383	Growth/differentiation factor 9	295	4.41%	2.04E-02	Secreted	N-terminus
P07686	Beta-hexosaminidase subunit beta	79	6.33%	3.58E-02	Lysosome	N-terminus
P55103	Inhibin beta C chain	218	4.59%	3.07E-02	Secreted	N-terminus
P58166	Inhibin beta E chain	217	5.07%	1.25E-02	Secreted	N-terminus
P51460	Insulin-like 3	47	14.89%	9.55E-05	Secreted	Internal
P19827	ITI heavy chain H1	246	4.47%	2.84E-02	Secreted	C-terminus
Q99538	Legumain	110	9.09%	2.40E-04	Lysosome	C-terminus
P09848	Lactase-phlorizin hydrolase	847	3.78%	5.15E-03	Apical cell membrane	N-terminus
P10253	Lysosomal alpha-glucosidase	42	9.52%	1.56E-02	Lysosome	N-terminus
P14151	L-selectin	10	20.00%	2.11E-02	Cell membrane	N-terminus
Q9NRE1	Matrix metalloproteinase-26	72	8.33%	6.34E-03	Secreted	N-terminus
P16519	PCSK2	84	8.33%	3.29E-03	Secretory vesicle	N-terminus
P29122	PCSK6	86	5.81%	4.86E-02	Secreted, Endoplasmic reticulum	N-terminus
P01127	PDGF subunit B	112	5.36%	4.53E-02	Secreted	N- and C-terminus
Q96B86	Repulsive guidance molecule A	147	5.44%	2.10E-02	Cell membrane	N- and C-terminus
P10600	Transforming growth factor beta-3	280	5.00%	5.96E-03	Secreted	N-terminus
Q9BZD6	Proline-rich Gla protein 4	32	9.38%	3.68E-02	Cell membrane	N-terminus
Q8N2E6	Prosalusin	163	5.52%	1.37E-02	Secreted	N-terminus
O43915	Vascular endothelial growth factor D	216	6.02%	1.66E-03	Secreted	N- and C-terminus

All human proteins with annotated propeptides in the UniProt database that have more histidine in their propeptides than expected, assuming a 2.3% probability of histidine at each sequence position, with a significance level of 5%.