

Supplementary data

Proteome-*pI*: Proteome Isoelectric Point Database

Lukasz P. Kozlowski^{1,*}

¹ Quantitative and Computational Biology Group, Max Planck Institute for Biophysical Chemistry, Göttingen, Lower Saxony, 37077, Germany

* Corresponding author: lukasz.kozlowski.lpk@gmail.com

Database web address: <http://isoelectricpointdb.org>

Supplementary table S1. Detailed statistics of the Proteome-*pl* database

	IDM number of proteins	Median number of protein	IDM size of proteins	Median size of protein	IDM mw of proteins	Median mw of protein
Viruses	22	7	225	182	26	21
Archaea	2,254	2,065	255	241	28	27
Bacteria	3,011	2,950	279	265	31	29
Eukaryote (all isoforms)	13,187	11,684	369	336	41	37
Eukaryote (main isoform)	12,618	11,585	367	334	41	37
Eukaryote (minor isoforms)	387	15	407	362	45	40

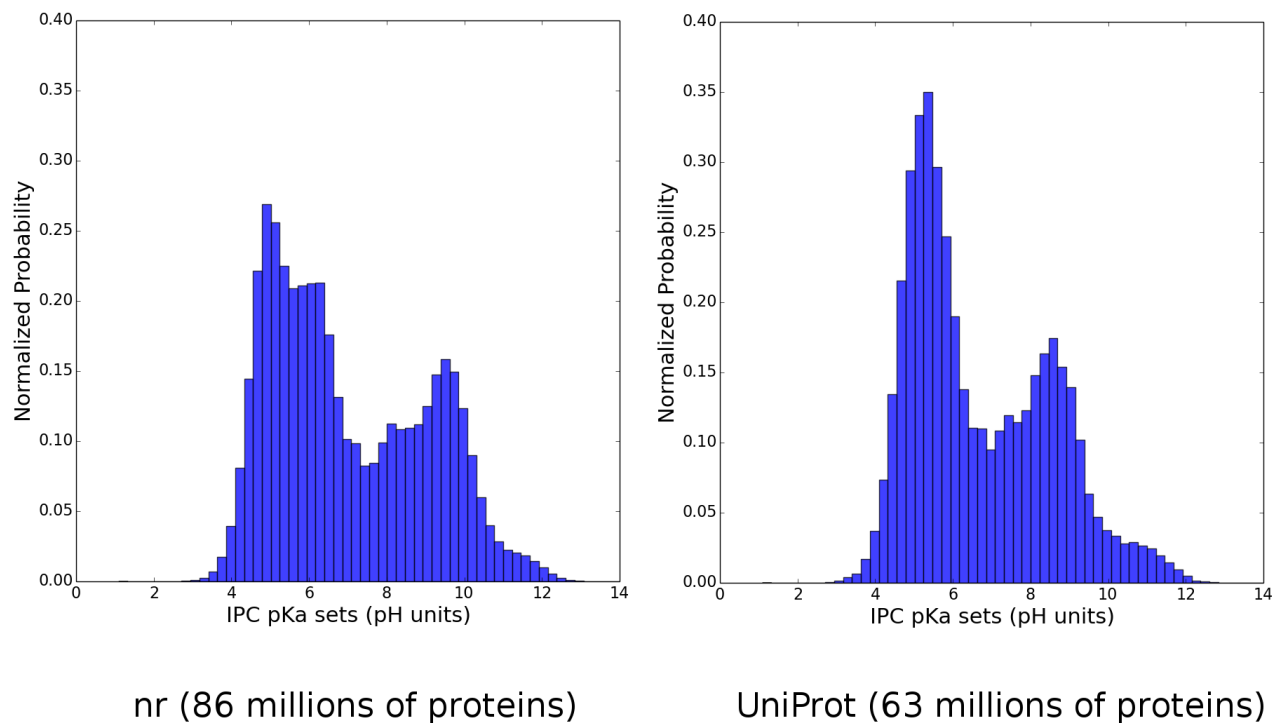
mw - molecular weight in kDa; IDM - interdecile mean

Supplementary table S2. Di-amino acid frequency in the Proteome-*pl* database (average from 5,029 proteomes)

	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
Ala	9.16	1.12	4.67	5.48	3.21	6.70	1.77	4.57	3.96	8.98	2.18	2.79	4.13	3.28	5.16	5.93	4.83	6.38	1.06	2.21
Cys	1.01	0.30	0.73	0.73	0.58	1.12	0.36	0.73	0.64	1.30	0.29	0.52	0.72	0.49	0.80	1.06	0.72	0.88	0.18	0.41
Asp	4.80	0.69	3.24	4.03	2.30	4.21	1.17	3.26	2.64	5.28	1.26	2.01	2.85	1.80	3.02	3.46	2.73	3.84	0.78	1.77
Glu	5.58	0.71	3.58	4.76	2.15	3.85	1.38	3.74	4.11	6.06	1.60	2.72	2.48	2.70	4.00	3.69	3.32	4.08	0.75	1.74
Phe	3.15	0.63	2.40	2.25	1.70	2.97	0.88	2.22	1.81	3.79	0.87	1.63	1.72	1.35	1.94	3.03	2.20	2.66	0.53	1.29
Gly	5.99	0.99	3.77	4.14	2.94	5.33	1.62	4.05	3.74	6.59	1.77	2.63	2.81	2.50	4.22	5.02	4.02	4.96	1.04	2.27
His	1.77	0.38	1.17	1.23	0.96	1.66	0.72	1.21	0.94	2.29	0.49	0.82	1.40	0.94	1.42	1.52	1.16	1.44	0.32	0.73
Ile	4.84	0.83	3.35	3.54	2.22	3.85	1.21	3.08	2.95	5.10	1.15	2.42	2.77	1.97	2.88	4.04	3.14	3.75	0.61	1.68
Lys	4.05	0.61	2.85	3.84	1.69	3.08	1.10	3.13	3.64	4.73	1.31	2.47	2.37	2.11	3.03	3.35	2.89	3.25	0.58	1.63
Leu	9.04	1.35	5.39	5.90	3.75	6.53	2.22	4.99	5.06	9.02	2.14	3.69	5.21	3.88	5.83	7.21	5.35	6.42	1.11	2.59
Met	2.22	0.27	1.27	1.46	0.80	1.54	0.48	1.26	1.38	2.21	0.61	1.00	1.18	0.91	1.33	1.70	1.41	1.52	0.23	0.57
Asn	2.95	0.55	2.07	2.32	1.58	2.86	0.87	2.54	2.16	3.72	0.92	1.94	2.15	1.54	1.98	2.75	2.13	2.51	0.50	1.33
Pro	4.64	0.56	2.89	3.47	1.87	3.62	1.12	2.31	2.21	4.49	1.02	1.79	2.97	1.95	2.63	3.97	2.89	3.58	0.61	1.37
Gln	3.37	0.47	1.81	2.42	1.34	2.35	0.97	2.13	2.11	3.87	0.96	1.59	1.96	2.27	2.46	2.51	2.05	2.46	0.49	1.09
Arg	4.87	0.77	3.10	3.75	2.27	3.69	1.46	3.17	3.04	5.86	1.43	2.13	2.82	2.37	4.12	3.82	2.98	3.72	0.82	1.73
Ser	5.73	1.00	3.70	3.89	2.93	5.34	1.57	3.81	3.43	6.86	1.61	2.82	3.84	2.64	3.91	5.89	4.16	4.56	0.89	2.03
Thr	5.04	0.74	2.88	3.11	2.16	4.26	1.18	3.14	2.51	5.48	1.17	2.07	3.33	1.90	2.81	4.10	3.29	4.09	0.68	1.58
Val	6.24	0.96	3.90	4.25	2.69	4.51	1.44	3.75	3.26	6.74	1.56	2.51	3.40	2.34	3.78	4.72	3.97	4.82	0.81	1.88
Trp	0.94	0.18	0.68	0.69	0.51	0.80	0.31	0.69	0.67	1.37	0.34	0.55	0.52	0.56	0.86	0.85	0.70	0.78	0.22	0.37
Tyr	2.20	0.47	1.72	1.71	1.35	2.13	0.71	1.61	1.44	2.93	0.64	1.29	1.34	1.19	1.74	1.99	1.58	1.84	0.40	1.04

* Similar statistics for all 5,029 proteomes, grouped proteomes from different kingdoms of life are available online in "[Statistics](#)" section and on individual subpages of the database.

All values are presented as per milles (‰), therefore need to be multiplied by 10⁻³. For better visibility di amino acids which are more than expected are marked by red, and those which are underrepresented are marked by blue in the table.



Supplementary Figure S1. Isoelectric point distributions for proteins from *nr* and UniProt databases

Isoelectric point for biological databases

Pre-computed data with isoelectric point for *nr*, UniProt and PDB databases are available as separate files from RepOD repository:

Isoelectric point for all *nr* (non redundant) NCBI proteins (86 million proteins) April 2016. RepOD. <http://dx.doi.org/10.18150/repod.2872103>

Isoelectric point for UniProtKB/TrEMBL proteins (63 million proteins) April 2016 RepOD. <http://dx.doi.org/10.18150/repod.9948646>

Isoelectric point for PDB proteins (~339,000proteins) Dec 2015. RepOD. <http://dx.doi.org/10.18150/repod.1549954>