

Supplemental Table 2

Protein	PDB	Function	Name	No. residues with matching ESM1v scores	No. mutations with matching ESM1v scores	References
GCK	P35557	Enzyme	Glucokinase	463	8396	Gersing, S., Schulze, T. K., Cagiada, M., Stein, A., Roth, F. P., Lindorff-Larsen, K., & Hartmann-Petersen, R. (2024). Characterizing glucokinase variant mechanisms using a multiplexed abundance assay. <i>Genome biology</i> , 25(1), 98.
PTEN	P60484	Enzyme	Phosphatase and tensin homolog	383	5083	Matreyek, K. A., Starita, L. M., Stephany, J. J., Martin, B., Chiasson, M. A., Gray, V. E., ... & Fowler, D. M. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. <i>Nature genetics</i> , 50(6), 874-882.
VKOR	Q9BQB6	Enzyme	Vitamin K epoxide reductase	162	2695	Chiasson, M. A., Rollins, N. J., Stephany, J. J., Sitko, K. A., Matreyek, K. A., Verby, M., ... & Fowler, D. M. (2020). Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact. <i>elife</i> , 9, e58026.
NUDT15	Q9NV35	Enzyme	Nucleotide triphosphate diphosphatase	163	2922	Suiter, C. C., Moriyama, T., Matreyek, K. A., Yang, W., Scaletti, E. R., Nishii, R., ... & Yang, J. J. (2020). Massively parallel variant characterization identifies NUDT15 associated with thiopurine toxicity. <i>Proceedings of the National Academy of Sciences</i> , 117(10), 5394-5401.
ASPA	P45381	Enzyme	Aspartoacylase	310	5843	Grønbaek-Thygesen, M., Voutsinos, V., Johansson, K. E., Schulze, T. K., Cagiada, M., Pedersen, L., ... & Hartmann-Petersen, R. (2024). Deep mutational scanning reveals a correlation between degradation and toxicity of thousands of aspartoacylase variants. <i>Nature Communications</i> , 15(1), 4026.
TPMT	P51580	Enzyme	Thiopurine S-methyltransferase	241	3648	Matreyek, K. A., Starita, L. M., Stephany, J. J., Martin, B., Chiasson, M. A., Gray, V. E., ... & Fowler, D. M. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. <i>Nature genetics</i> , 50(6), 874-882.
CYP2C9	P11712	Enzyme	Cytochrome P450 enzyme	486	6370	Amorosi, C. J., Chiasson, M. A., McDonald, M. G., Wong, L. H., Sitko, K. A., Boyle, G., ... & Dunham, M. J. (2021). Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. <i>The American Journal of Human Genetics</i> , 108(9), 1735-1751.
OCT	O15245	Transporter	Organic cation transporter 1	547	9803	Yee, S. W., Macdonald, C. B., Mitrovic, D., Zhou, X., Koleske, M. L., Yang, J., ... & Coyote-Maestas, W. (2024). The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics. <i>Molecular cell</i> , 84(10), 1932-1947.
PRKN	O60260	Enzyme	E3 ubiquitin-protein ligase parkin	465	8756	Clausen, L., Okarmus, J., Voutsinos, V., Meyer, M., Lindorff-Larsen, K., & Hartmann-Petersen, R. (2024). PRKN-linked familial Parkinson's disease: cellular and molecular mechanisms of disease-linked variants. <i>Cellular and Molecular Life Sciences</i> , 81(1), 223.

Supplemental Table 3

Protein	UniProt ID	Active Site Structure	Active Site Substrate	Active Site Annotation	Allosteric Site Structure	Allosteric Modulator	Allosteric Site Annotation	Type	Size	Slope	t-value	P-value	R2	Half Distance
GCK	P35557	1V4S, 3FGU	ANP,BGC, GLC	from 1V4S	1V4S,3ID8	MRK (activator)	literature (Kamata et al., 2004) (Larion & Miller, 2009)	kinase	465	-0.043143	-9.147	< 2.2e-16	0.1654	15.73543
PTP1B	P18031	1PTY	PTR	from 1PTY	1T48,1T49, 1T4J	892,BB3,FRJ (inhibitor)	consensus	phosphatase	435	-0.02548	-4.191	3.71E-05	0.05474	19.73865
PDK1	O15530	4RQK, 5LVM	ADE,ATP	consensus	3ORZ,4RQK	2A2 (activator), R1S (inhibitor)	from 4RQK	kinase	556	-0.048941	-7.152	7.56E-12	0.1519	12.28401
CHK1	O14757	2E9N, 5OOP	76A,ANP	consensus	3F9N	38M (inhibitor)	from 3F9N	kinase	476	-0.039685	-6.378	8.00E-10	0.1302	15.89498
CASP1	P29466	2HBQ, 6BZ9	C,PHQ	consensus	2FQQ	F1G (inhibitor)	from 2FQQ	protease	404	-0.024747	-3.794	0.000188	0.05308	22.64102
IDH	O75874	5YFN, 6ADG, 6B0Z, 6BKZ, 6IO0	FLC,ICT,MG, NAP,NDP	consensus from 6B0Z, 6BKZ, 6IO0	5SUN,6B0Z, 6BKZ,6BL2, 6IO0	70Q, AOU, C81, CIT, DWM, DWS (inhibitor)	consensus from 6B0Z, 6BKZ, 6IO0	oxidoreductase	414	-0.036131	-6.531	2.26E-10	0.1042	15.36862

Supplemental Table 4. PSD-95-PDZ3: linear regression models

Model Comparison		Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
whole protein	ESM1v ~ $\Delta\Delta G_f$	1565	21,260				
whole protein	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b$	1564	18,199	1	3,061.20	263.07	< 2.2e-16 ***
non-orthosteric sites	ESM1v ~ $\Delta\Delta G_f$	1323	15,721				
non-orthosteric sites	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b$	1322	13,441	1	2279.3	224.18	< 2.2e-16 ***
orthosteric sites	ESM1v ~ $\Delta\Delta G_f$	240	1845				
orthosteric sites	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b$	239	1704.8	1	140.21	19.657	1.413e-05 ***

Supplemental Table 5. GRB2-SH3: linear regression model

Model Comparison		Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
whole protein	ESM1v ~ $\Delta\Delta G_f$	754	7,968				
whole protein	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b$	753	5,795	1	2,173.30	282.41	< 2.2e-16 ***
non-orthosteric sites	ESM1v ~ $\Delta\Delta G_f$	583	3,265				
non-orthosteric sites	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b$	582	3,020	1	245.48	47.311	1.568e-11 ***
orthosteric sites	ESM1v ~ $\Delta\Delta G_f$	169	1864.3				
orthosteric sites	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b$	168	1343	1	521.27	65.205	1.245e-13 ***

Supplemental Table 6. PSD-95-PDZ3: linear mixed models

Model Comparison		npar	AIC	BIC	logLik	-2logLik	Chisq	Df	Pr(>Chisq)
whole protein	ESM1v ~ $\Delta\Delta G_f + (1 \text{pos})$	4	7575.2	7596.6	-3783.6	7567.2			
whole protein	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b + (1 \text{pos})$	5	7409.7	7436.5	-3699.9	7399.7	167.48	1	< 2.2e-16 ***
non-orthosteric sites	ESM1v ~ $\Delta\Delta G_f + (1 \text{pos})$	4	6354.3	6375.1	-3173.2	6346.3			
non-orthosteric sites	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b + (1 \text{pos})$	5	6215.8	6241.7	-3102.9	6205.8	140.57	1	< 2.2e-16 ***
orthosteric sites	ESM1v ~ $\Delta\Delta G_f + (1 \text{pos})$	4	1175.3	1189.2	-583.64	1167.3			
orthosteric sites	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b + (1 \text{pos})$	5	1155.6	1173	-572.78	1145.6	21.705	1	3.179e-06 ***

Supplemental Table 7. GRB2-SH3: linear mixed models

Model Comparison		npar	AIC	BIC	logLik	-2logLik	Chisq	Df	Pr(>Chisq)
whole protein	ESM1v ~ $\Delta\Delta G_f + (1 \text{pos})$	4	3339.9	3358.4	1666	3331.9			
whole protein	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b + (1 \text{pos})$	5	3304.1	3327.2	-1647	3294.1	37.858	1	7.609e-10 ***
non-orthosteric sites	ESM1v ~ $\Delta\Delta G_f + (1 \text{pos})$	4	2461.3	2478.8	-1226.7	2453.3			
non-orthosteric sites	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b + (1 \text{pos})$	5	2434.5	2456.4	-1212.3	2424.5	28.795	1	8.047e-08 ***
orthosteric sites	ESM1v ~ $\Delta\Delta G_f + (1 \text{pos})$	4	810.36	822.92	-401.18	802.36			
orthosteric sites	ESM1v ~ $\Delta\Delta G_f + \Delta\Delta G_b + (1 \text{pos})$	5	807.81	823.52	-398.91	797.81	4.5424	1	0.03307 *