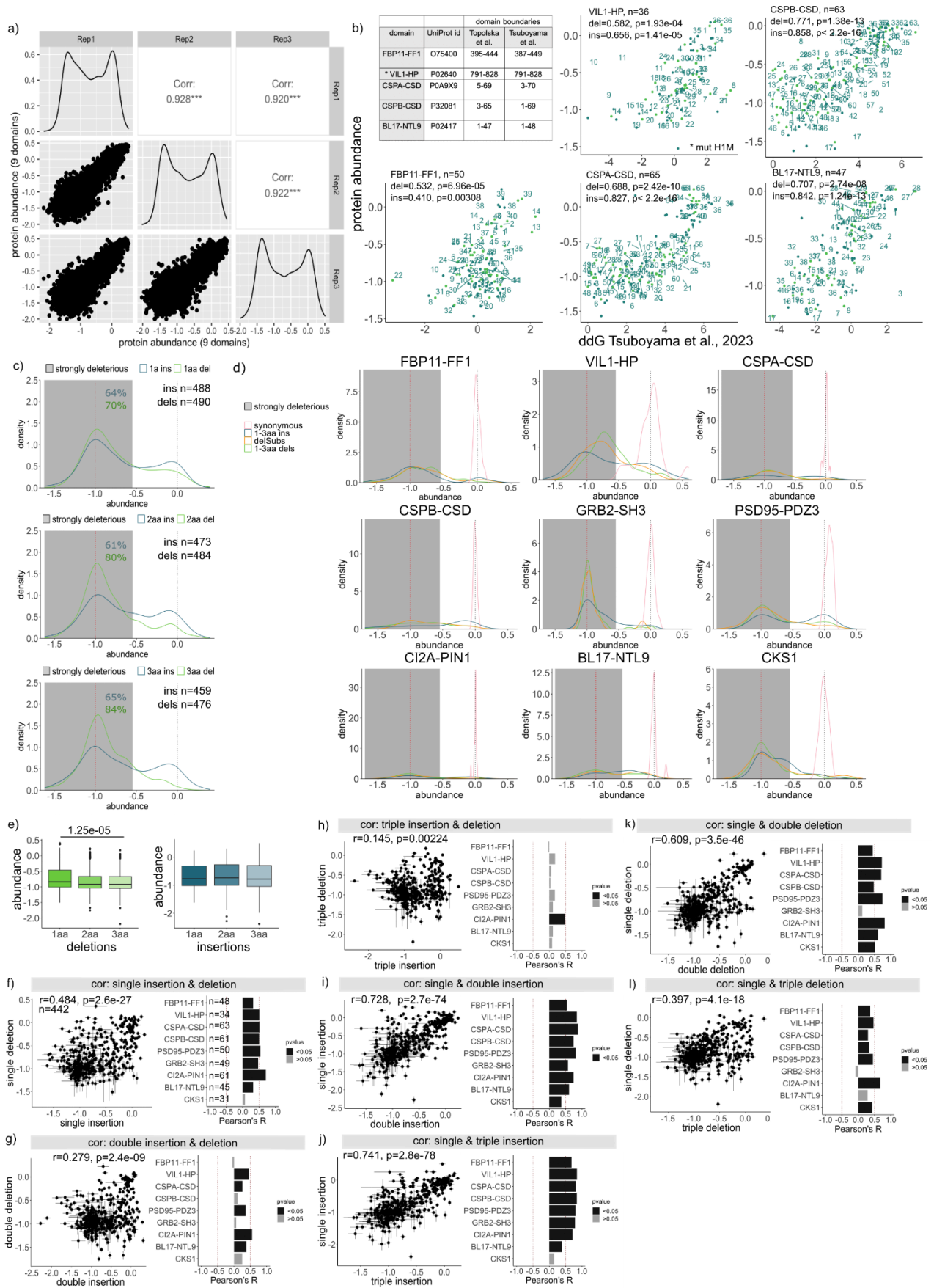


Supplementary Material

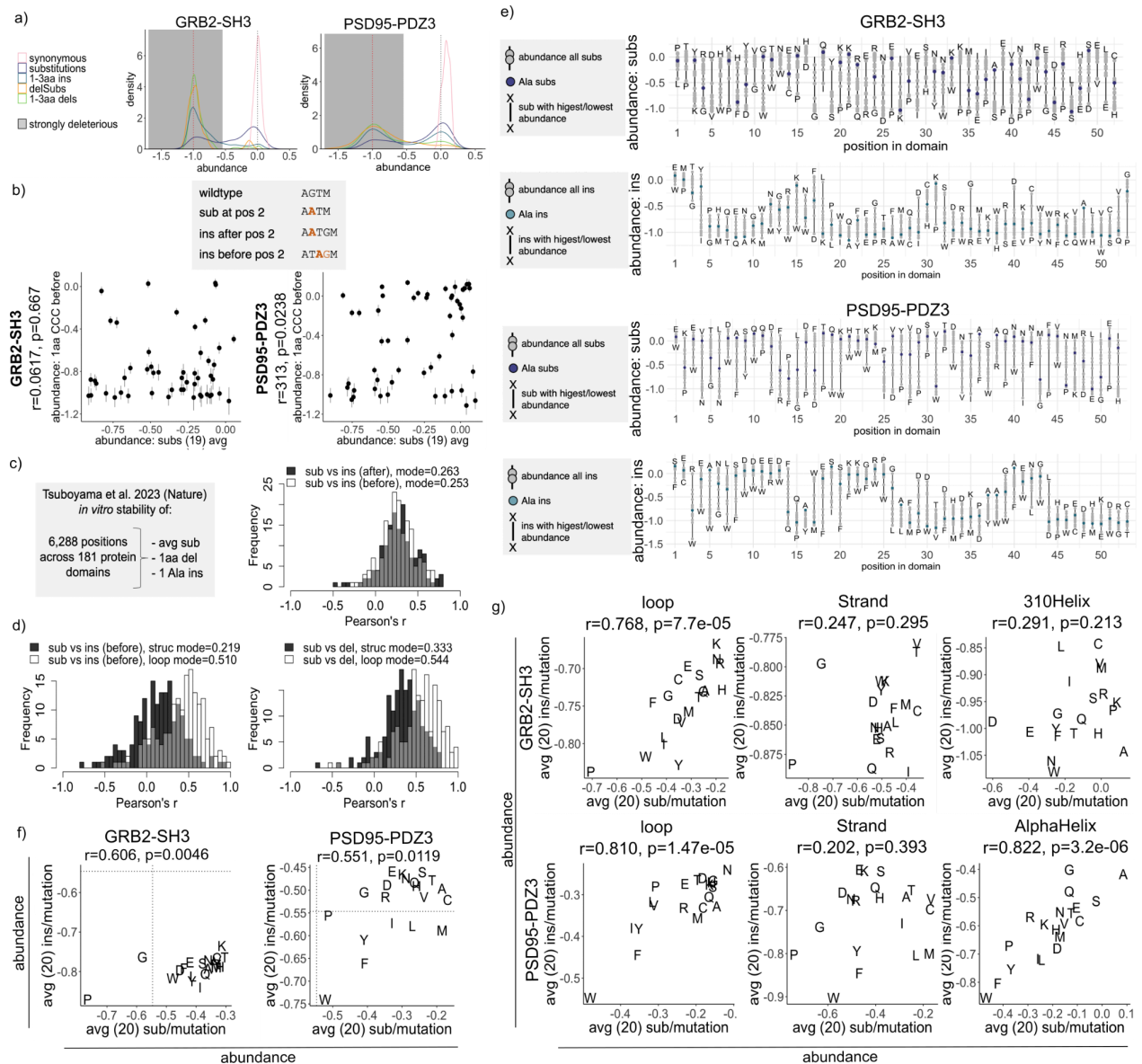
Deep indel mutagenesis reveals the impact of amino acid insertions and deletions on protein stability and function

Topolska, Beltran & Lehner

Supplementary Figures.

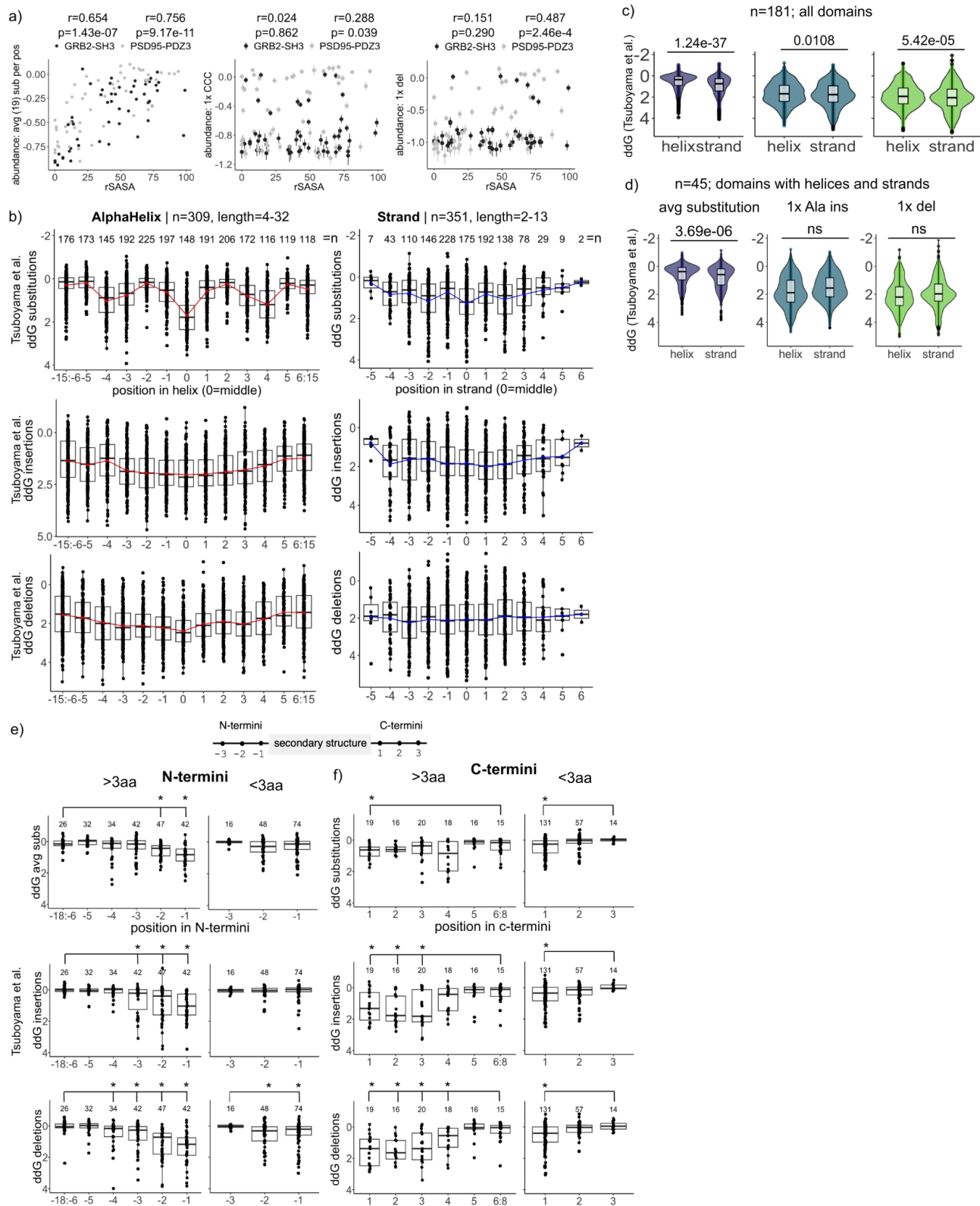


Supplementary Figure 1. Experimental reproducibility and comparisons of single and multi-aa indels. **a.** Scatter plots showing reproducibility of abundance scores for the 9 domains; n=7127. **b.** Scatter plots showing correlation of aPCA scores and inferred ddG stability scores from Tsuboyama et al., for indels across the overlapping 5 domains. Pearson's correlation and p-value reported for each domain. The table reports on differences in domain boundaries across the experiments; green: deletions; blue: insertions. **c.** Density distributions of 1-3aa deletions (green) and 1-3aa CCC insertions (blue) aPCA scores for all 9 domains. **d.** Density distributions of all mutation types for the 9 domains. **e.** Variation in indel deleteriousness across 1-3aa indel lengths. P-values were calculated using Wilcoxon one-sided t-test with multiple testing adjustment (Bonferroni). Boxplots represent median values and the lower and upper hinges correspond to the 25th and 75th percentiles, respectively. Sample sizes for 1-3aa deletions: n=490, 484 and 476. . Sample sizes for 1-3aa insertions: n=488, 473 and 459. **f-l.** Scatter plots of 1-3aa insertion versus deletion aPCA scores (f-h) and single versus multi-aa insertion or deletion aPCA scores (i-j, k-l respectively) across the 9 domains. Results are reported as scatter plots for all 9 domains and bar plots with Pearson's correlation coefficients and p-values for each individual domain. Sample sizes for f-l are displayed in f.

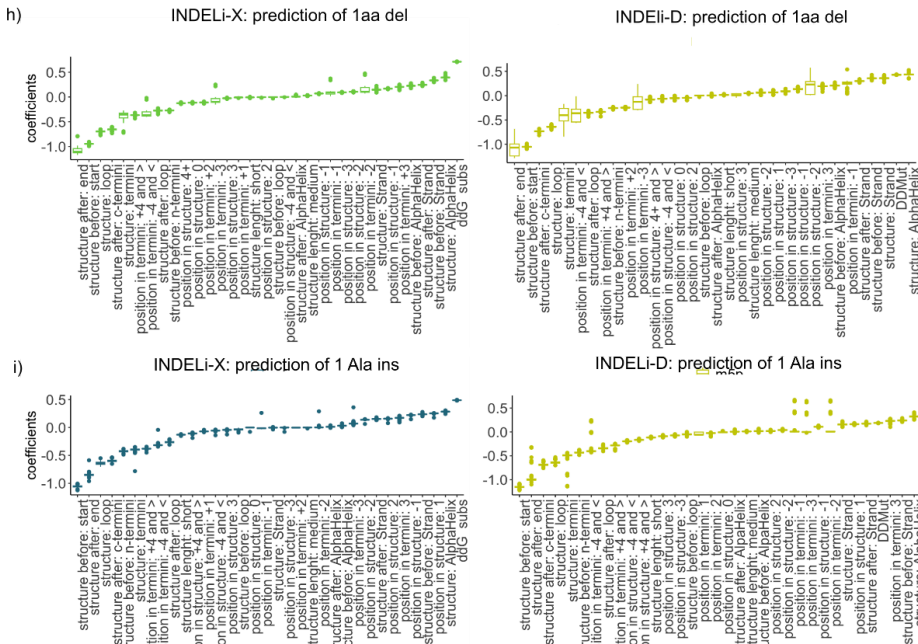
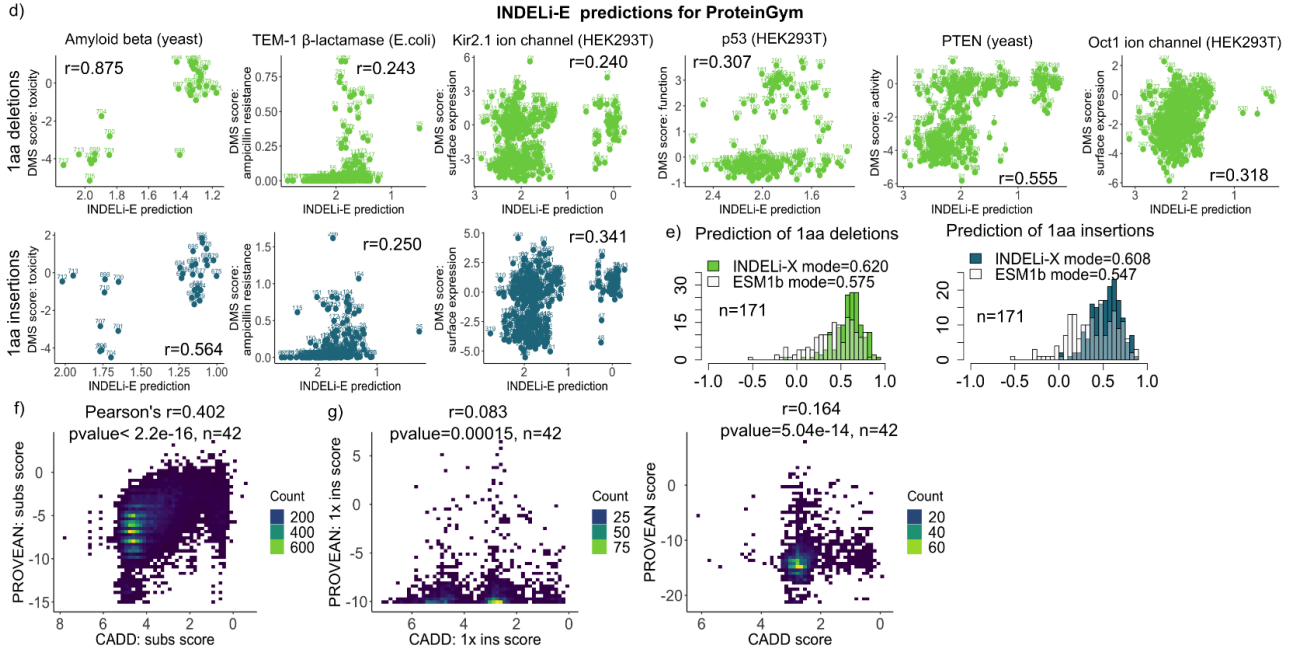
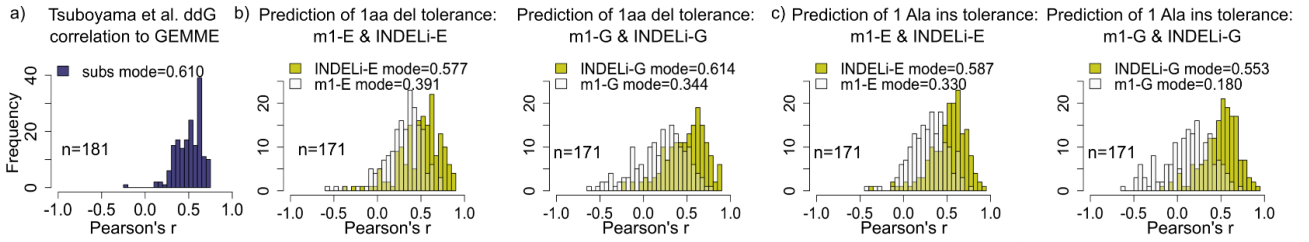


Supplementary Figure 2. Comparison of substitutions and insertion effects. **a.** Density distributions of all mutation types for GRB2-SH3 and PSD95-PDZ3, including all possible 20 insertions at each position. **b.** Scatter plots of 1aa CCC insertion versus effects of mean substitution/residue with Pearson's correlations and corresponding p-values. Insertion data is presented as mean values and error bars indicate 95% CI (n=3 biological replicates). Sample sizes were n=51 for GRB2-SH3 and n=52 for PSD95-PDZ3. **c.** Overview of the filtered Tsuboyama et al. dataset (See Methods) and histogram of Pearson's correlations between average substitution per position ddG and 1 Ala insertion before or after the substitution site. **d.** Histograms of Pearson's correlations between the all mutational types within loop (white) and structured (black) regions for the Tsuboyama et al. dataset. **e.** Domain abundance scores of 19 aa substitutions (upper panel) and 20 aa insertions (lower panel) at each position; purple dot: substitutions to Ala, blue dot: insertions of Ala. **f.** Scatter plots of mean insertion and mean substitution abundance scores/aa identity with Pearson's correlations. The dotted lines mark the threshold of severe deleteriousness for protein domain abundance (See Methods). The averages are based on 20 versus 20 aa comparisons, calculated from ins n=1055 and subs n=861 for GRB2-SH3 and ins n=1057

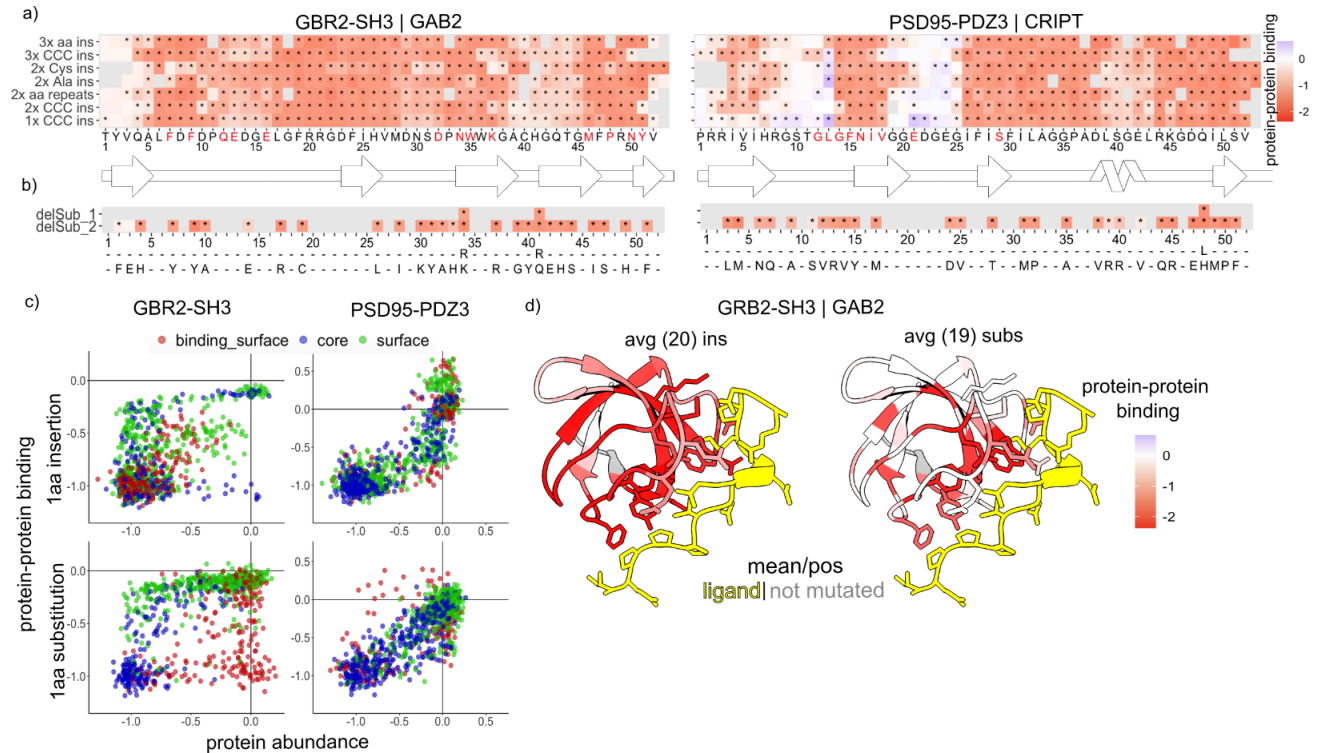
and subs n=988 for PSD95-PDZ3. **g.** Scatter plots of mean insertion and mean substitution abundance scores per aa identity within loops, strands, 310-helices and alpha-helices with Pearson's correlations and corresponding p-values. The averages are based on 20 versus 20 aa comparisons, calculated from ins n=1035 (loop: n=499, strand: n=476, helix: n=60) and subs n=861 (loop: n=407, strand: n=408, helix: n=46) for GRB2-SH3, and ins n=1037 (loop: n=479, strand: n=459, helix: n=99) and subs n=988 (loop: n=456, strand: n=437, helix: n=95) for PSD95-PDZ3.



Supplementary Figure 3. Variability of indel tolerance across secondary structure features. **a.** Scatter plots of protein abundance scores for average substitution/residue, 1aa CCC insertion and 1aa deletions versus the relative solvent accessible area (rSASA) with Pearson's correlation coefficients. Insertion and deletion data is presented as mean values and error bars indicate 95% CI (n=3 biological replicates). Sample sizes for substitution, insertion, and deletion comparisons were n=52, n=52, and n=51 for GRB2-SH3, and n=52 for all comparisons in PSD95-PDZ3. **b.** Patterns of ddG periodicity for substitutions, 1 Ala insertions and 1aa deletions across alpha-helices (red) and strands (blue). Red/blue line represents the mean/position. Position in helix/strand on the x-axis is counted from the middle of the structural element (See Methods). **c.** Violin plots of average substitution, 1 Ala insertion and 1aa deletion ddG between alpha-helices and strands across all 181 domains and **d.** across protein domains that contain both alpha-helices and strands. P-values from Wilcox one-sided t-tests are displayed above the plots. ns: p-value>0.05. Boxplots represent median values and the lower and upper hinges correspond to the 25th and 75th percentiles, respectively. Sample sizes for substitution, insertion, and deletion comparisons across the 181 domains were n=1157 for strands and n=2474 for helices, while across the 45 domains, they were n=479 for strands and n=737 for helices. **e-f.** Average substitution, 1 Ala insertion and 1aa deletion tolerance across short (1-3aa) and long (≥ 4 aa) N- and C-termini. Numbers of termini/position are indicated above the plots; x-axis: realigned termini positions explained by the legend above the plot. Significant (Wilcox one-sided t-test pvalue < 0.05) changes are denoted by “*”. Boxplots represent median values and the lower and upper hinges correspond to the 25th and 75th percentiles, respectively.



Supplementary Figure 4. Indel variant effect prediction. **a.** Histograms of Pearson's correlation coefficients for observed and predicted tolerance scores (>9) of average substitutions for the Tsuboyama et al. dataset using GEMME¹; $n=181$. **b-c.** Performance evaluation of models m1-E and -G and INDELi-E and -G for 1aa deletion (b) and 1aa insertion (c) prediction, as histograms of Pearson's correlation between the observed and predicted ddG, using ESM-1v² and GEMME. Mode of Pearson's r for each model is displayed in the legend. **d.** Scatter plots for observed and predicted 1aa deletion (green) and 1aa insertion (blue) scores across the indel ProteinGym benchmark dataset (<https://proteingym.org/>), not already included in the leave-one-domain-out cross-validation, using INDELi-E. **e.** Direct comparison of ESM-1b³ performance compared to INDELi-X shown as histograms of Pearson's correlation between the observed and predicted ddG. **f.** Plots of correlations between the predicted PROVEAN and CADD substitution scores. Results are shown as a scatter plot and histogram with the per-domain Pearson's correlation. **g.** Scatter plots of the correlation between the predicted PROVEAN and CADD insertion and deletion scores and the corresponding Pearson's correlation. **h-i.** Top significant coefficients from the deletion (f) and insertion (g) predictor models INDELi-X (left) and INDELi-D (right).



Supplementary Figure 5. Impact of multi-aa insertions and delSubs on protein interactions. **a-b.** Heatmaps of various types of 1-3aa insertions (a) and delSubs (b) bPCA scores. Changes in abundance significantly different from the weighted mean of synonymous variants, determined using a Bonferroni multiple testing correction of z-scores (two-tailed test, FDR ≤ 0.05), are marked with “*”. The two-tailed test compares the z-statistic for each mutation to the null hypothesis (mean = 0). The y-axis shows the identity, type, or length of mutation (multi-insertion repeats refer to duplicating the preceding amino acid 1-3 times; CCC indicates repeating the 1-3 amino acid sequence preceding the insertion site), and the x-axis shows the mutated position and amino acid sequence of the domain; x-axis: mutated position and identity of the substitution for delSub. We show the 2-dimensional structural representation of the domains (SSDraw⁴) under each heatmap. For PSD95-PDZ3 and GRB2-SH3, we mutated the first 52 residues, leaving, respectively, 32 and 4aa in the C-terminal as wild type. **c.** Scatter plots of 19 substitution and 20 insertion effects on protein domain abundance and binding; red: residues in binding surface; blue: residues with rSASA < 30; green: residues with rSASA > 30. **d.** Structure of GRB2-SH3 interacting with its ligand GAB2 (pdb 2VWF) coloured by the mean bPCA score of 20 aa insertions/residue (right) and mean of 19 aa substitutions/residue (left). Visualised using ChimeraX⁵.

Supplementary References.

1. Laine, E., Karami, Y. & Carbone, A. GEMME: A Simple and Fast Global Epistatic Model Predicting Mutational Effects. *Mol Biol Evol* **36**, 2604–2619 (2019).
2. Meier, J. *et al.* Language models enable zero-shot prediction of the effects of mutations on protein function. *bioRxiv* 2021.07.09.450648 (2021) doi:10.1101/2021.07.09.450648.
3. Brandes, N., Goldman, G., Wang, C. H., Ye, C. J. & Ntranos, V. Genome-wide prediction of disease variant effects with a deep protein language model. *Nat. Genet.* **55**, 1512–1522 (2023).
4. Chen, E. A. & Porter, L. L. SSDraw: Software for generating comparative protein secondary structure diagrams. *Protein Sci.* **32**, e4836 (2023).
5. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
6. Tsuboyama, K. *et al.* Mega-scale experimental analysis of protein folding stability in biology and design. *Nature* **620**, 434–444 (2023).