Supplementary Material

Figures

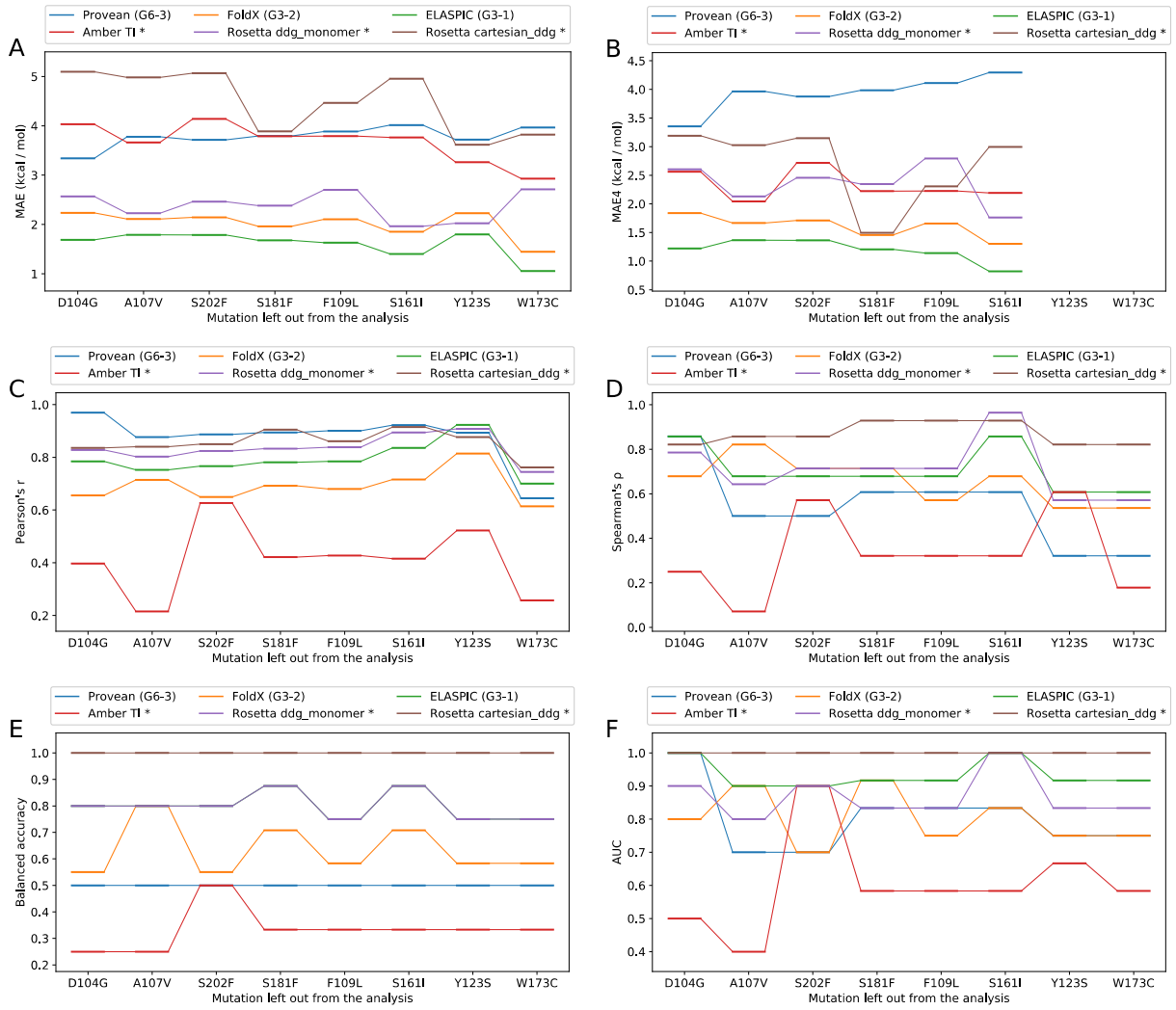

Figure S1. The robustness of predictions made using Provean, FoldX, ELASPIC, Amber TI, Rosetta ddb_monomer, and Rosetta cartesian_ddg is examined using six different metrics while removing one of the mutations in tern from the analysis. **(A)** Mean absolute error between predicted and actual values. **(B)** Mean absolute error considering only those mutations that have an experimental ΔΔG less than 4 kcal / mol. **(C)** Pearson's correlation coefficient between predicted and actual values. **(D)** Spearman's correlation coefficient between predicted and actual values. **(E)** Balanced accuracy, describing the average of the recall for neutral mutations and for destabilizing mutations. **(F)** Area under the receiver Receiver Operating Characteristic (ROC) Curve. * Predictions made by methods marked with an asterisk were not submitted to the CAGI5 frataxin challenge.

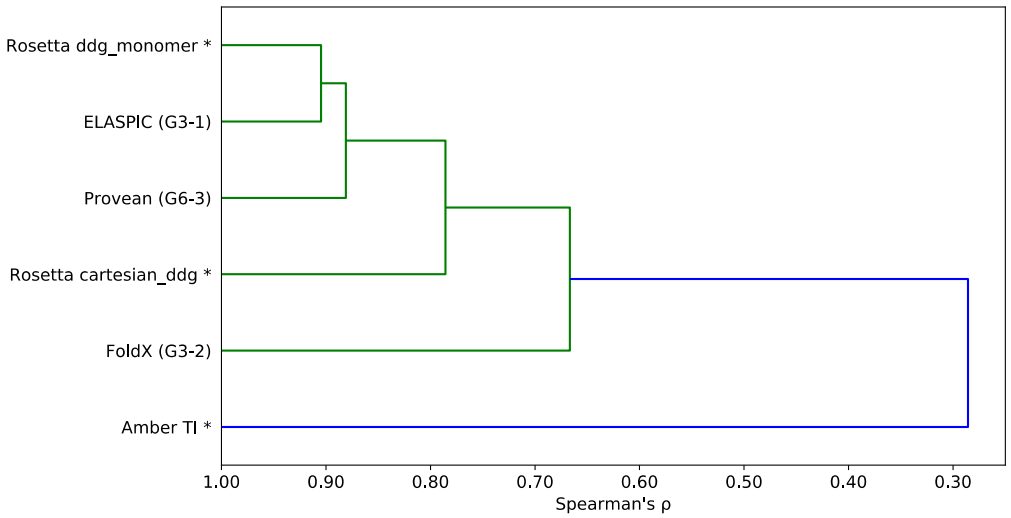

Figure S2. Hierarchical clustering dendrogram showing the correlations between predictions made by the six methods evaluated in this study. Distances between clusters were calculated by finding the shortest distance between two points in those clusters (i.e. using the "single" linkage criterion). The distance function used was 1 - Spearman's ρ. * Predictions made by methods marked with an asterisk were not submitted to the CAGI5 frataxin challenge.

Tables

Table S1. Experimentally-determined ΔΔG values (in kcal / mol), and predictions by the six different methods evaluated in this study, for eight mutations in human FXN (NP_000135.2). * Predictions made by methods marked with an asterisk were not submitted to the CAGI5 frataxin challenge.

Mutation	$\Delta\Delta G_{exp}$	Provean (G6-3)	FoldX (G6-2)	ELASPIC (G6-1)	Amber TI *	Rosetta ddg_monomer *	Rosetta cartesian_ddg *
p.D104G	-0.255	6.564	0.177	0.752	-1.402	-1.317	-0.046
p.A107V	-0.220	3.540	1.085	0.077	-3.963	-3.664	0.803
p.S202F	0.685	4.887	-0.392	0.373	1.063	-1.125	0.260
p.S181F	2.035	5.696	-0.308	0.952	-0.815	-0.331	10.722
p.F109L	2.645	5.676	1.286	1.227	-0.188	2.774	7.268
p.S161I	3.440	5.547	0.322	0.436	0.429	-1.842	2.247
p.Y123S	4.480	8.655	4.973	4.734	-2.050	9.364	15.064
p.W173C	9.500	11.952	3.544	4.059	0.629	9.438	18.642

Table S2. System commands used to run Rosetta's ddg_monomer protocol (left) and cartesian_ddg protocol (right).

Rosetta's ddg_monomer protocol

```
$ ddg_monomer.static.linuxgccrelease \
    -in:file:s '{structure_file}' \
    -in::file::fullatom \
    -database '{rosetta_db}' \
    -ignore_unrecognized_res true \
    -ignore_zero_occupancy false \
    -fa_max_dis 9.0\
    -ddg::weight_file soft_rep_design \
    -ddg::local_opt_only true \
    -ddg::mut_file '{mutation_file}' \
    -ddg::iterations 50 \
    -ddg::dump_pdbs true \
    -ddg::suppress_checkpointing true \
    -ddg::mean true \
    -ddg::min true \
    -ddg::output_silent true
```

Rosetta's cartesian_ddg protocol

```
$ cartesian_ddg.static.linuxgccrelease \
    -in:file:s '{structure_file}' \
    -in::file::fullatom \
    -database '{rosetta_db}' \
    -ignore_unrecognized_res true \
    -ignore_zero_occupancy false \
    -fa_max_dis 9.0 \
    -ddg::cartesian \
    -ddg::mut_file '{mutation_file}' \
     -ddg::iterations 3 \
    -ddg::dump_pdbs true \
    -ddg::suppress_checkpointing true \
    -ddg::mean true \
    -ddg::min true \
    -ddg::output_silent true \
    -bbnbr 1 \
    -beta_nov16_cart
```