Table 3S. Features considered for inclusion in the prediction model and their discriminatory power (F-score, *F*). Evolutionary based features were derived from the PSI-BLAST position specific scoring matrix (PSSM) generated after 3 iterations. Features highlighted with bold face were selected for the final model.

Acronym	***F*** [a]	***F*** [b]	***F*** [c]	**Description** [d]
dSS	0.64	0.68	0.65	Difference between similarity scores of wild type amino acid and mutation for a given position
Abs_dSS	0.68	0.73	0.67	Absolute difference between similarity scores of wild type amino acid and mutation for a given position
Entropy	0.63	0.66	0.63	Shannon entropy for a given position
EntropyRel	0.58	0.57	0.57	Shannon entropy for a given position relative to other positions computed similarly to the ConSurf procedure
zsEntropy7	0.35	0.41	0.36	Z-score for Shannon entropy at a given position based on a window of 7 neighboring amino acids
zsEntropy11	0.40	0.44	0.41	Z-score for Shannon entropy at a given position based on a window of 11 neighboring amino acids
zsEntropy15	0.44	0.46	0.45	Z-score for Shannon entropy at a given position based on a window of 15 neighboring amino acids
zsEntropy21	0.48	0.49	0.49	Z-score for Shannon entropy at a given position based on a window of 21 neighboring amino acids
varEntropy7	0.36	0.22	0.35	Variance of Shannon entropy for the window of 7 neighboring amino acids
varEntropy11	0.39	0.22	0.35	Variance of Shannon entropy for the window of 11 neighboring amino acids
varEntropy15	0.31	0.14	0.27	Variance of Shannon entropy for the window of 15 neighboring amino acids
varEntropy21	0.29	0.16	0.24	Variance of Shannon entropy for the window of 21 neighboring amino acids
zsPredRSA7	0.15	0.15	0.15	Z-score for predicted relative solvent accessibility at a given position based on a window of 7 neighboring amino acids
zsPredRSA11	0.16	0.16	0.16	Z-score for predicted relative solvent accessibility at a given position based on a window of 11 neighboring amino acids
zsPredRSA15	0.17	0.17	0.17	Z-score for predicted relative solvent accessibility at a given position based on a window of 15 neighboring amino acids
zsPredRSA21	0.22	0.22	0.22	Z-score for predicted relative solvent accessibility at a given position based on a window of 21 neighboring amino acids
varPredRSA7	0.29	0.29	0.29	Variance of predicted relative solvent accessibility for the window of 7 neighboring amino acids
varPredRSA11	0.37	0.37	0.37	Variance of predicted relative solvent accessibility for the window of 11 neighboring amino acids
varPredRSA15	0.40	0.40	0.40	Variance of predicted relative solvent accessibility for the window of 15 neighboring amino acids

varPredRSA21	0.45	0.45	0.45	Variance of predicted relative solvent accessibility for the window of 21 neighboring amino acids
SSref	0.45	0.51	0.45	Similarity score of wild type amino acid for a given position
SSsnp	0.59	0.60	0.61	Similarity score of mutation for a given position
dpAA	0.55	0.58	0.54	Difference between probabilities of wild type amino acid and mutation for a given position
Abs_dpAA	0.58	0.62	0.57	Absolute difference between probabilities of wild type amino acid and mutation for a given position
pAAref	0.52	0.57	0.51	Probability of wild type amino acid for a given position
pAAsnp	0.32	0.23	0.32	Probability of mutation amino acid for a given position
ss_Abs_dHP	0.34	0.39	0.35	Absolute difference between hydropathy indexes of wild type amino acid and mutation for a given position weighted by the difference of the corresponding similarity scores
p_Abs_dHP	0.33	0.36	0.33	Absolute difference between hydropathy indexes of wild type amino acid and mutation for a given position weighted by the difference of the corresponding probabilities
ss_Abs_dSize	0.56	0.61	0.54	Absolute difference between sizes of wild type amino acid and mutation for a given position weighted by the difference of the corresponding similarity scores
p_Abs_dSize	0.54	0.58	0.53	Absolute difference between sizes of wild type amino acid and mutation for a given position weighted by the difference of the corresponding probabilities
PredRSA	0.47	0.47	0.47	Relative solvent accessibility predicted by SABLE
PredTM	0.02	0.02	0.02	Binary value indicating whether a given position is at the predicted transmembrane region
HPref	0.08	0.08	0.08	Kyte-Doolittle hydropathy index for the wild type amino acid at a given position
HPsnp	0.04	0.04	0.04	Kyte-Doolittle hydropathy index for the new amino acid at a given position
dHP	0.04	0.04	0.04	Difference between hydropathy indexes of wild type amino acid and mutation for a given position
Abs_dHP	0.17	0.17	0.17	Absolute difference between hydropathy indexes of wild type amino acid and mutation for a given position
Size_ref	0.08	0.08	0.08	Size of the wild type amino acid at a given position
Size_snp	0.08	0.08	0.08	Size of the new amino acid at a given position
dSize	0.00	0.00	0.00	Difference between sizes of wild type amino acid and mutation for a given position
Abs_dSize	0.24	0.24	0.24	Absolute difference between sizes of wild type amino acid and mutation for a given position
RSA	0.35	0.35	0.35	3D structure based relative solvent accessibility computed by DSSP for a given position

Func_Cavity	0.27	0.27	0.27	Probability of the deleterious mutation of the amino acid residue known to be within the active site cavity
Func_Heme	0.28	0.28	0.28	Probability of the deleterious mutation of the amino acid residue known to be in contact with heme
Func_None	0.27	0.27	0.27	Probability of the deleterious mutation of the amino acid residue known to be outside the active site cavity
pPredPPI	0.08	0.08	0.08	Probability of being at a protein-protein interaction interface predicted by SPPIDER
Abs_Pred_dRSA	0.07	0.07	0.07	Absolute difference between predicted relative solvent accessibility and computed from 3D structure at a given position

[a] PSSM is based on the NCBI nr database used in SABLE predictions.
[b] PSSM is based on the reduced NCBI nr database after removing sequences with over 90% identity.
[c] PSSM is based on the reduced NCBI nr database after removing sequences with over 70% identity.
[d] Similarity scores are position specific scores derived from multiple sequence alignment (MSA) using, in this case, PSI-BLAST. They reflect likelihood of occurrence of a given amino acid at a given position based on a given sequence database used to generate MSA. Shannon entropy reflects variability of amino acids at a given position. Relative solvent accessibility measures solvent exposure of a residue in a given protein conformation normalized to a maximal solvent accessibility for a given type of amino acid.