

Supplementary Table S2. 20 most important features based on the sequence and structure-based models detected by the Information Gain algorithm. In the structure-based model, we indicate the pair of atomic types, and the maximum cutoff distance used as a parameter.

#	Sequence-based model		Structure-based model	
	Features	Info. gain	Features	Info. gain
1	P34S	0.245	donor x hydrophobic (12.0-0)	0.481
2	No mutations	0.065	donor x hydrophobic (11.2-0)	0.481
3	E155K	0.065	acceptor x acceptor (11.0-0)	0.447
4	S486T	0.037	donor x hydrophobic (14.8-0)	0.427
5	K404Q	0.036	donor x hydrophobic (10.8-0)	0.421
6	E383K	0.036	donor x hydrophobic (10.4-0)	0.421
7	V338M	0.036	aromatic x hydrophobic (15.8-0)	0.412
8	K281del	0.036	aromatic x hydrophobic (15.6-0)	0.412
9	P267H	0.036	acceptor x sulfide (15.0-0)	0.410
10	T261I	0.036	acceptor x hydrophobic (5.6-0)	0.380
11	G169R	0.036	aromatic x sulfide (13.0-0)	0.374
12	E156A	0.036	neutral x sulfide (9.4-0)	0.354
13	V136M	0.036	acceptor x acceptor (6.0-0)	0.352
14	T107I	0.036	acceptor x acceptor (6.4-0)	0.345
15	R26H	0.031	aromatic x hydrophobic (10.6-0)	0.345
16	V11M	0.031	aromatic x hydrophobic (9.8-0)	0.345
17	E410K	0.015	donor x sulfide (14.6-0)	0.344
18	A237S	0.015	aromatic x neutral (5.6-0)	0.340
19	A122S	0.015	donor x sulfide (15.2-0)	0.340
20	A90V	0.015	acceptor x donor (4.0-0)	0.338