Structure, Volume *32*

Supplemental Information

PMSPcnn: Predicting protein stability changes upon single point mutations with convolutional neural network

Xiaohan Sun, Shuang Yang, Zhixiang Wu, Jingjie Su, Fangrui Hu, Fubin Chang, and Chunhua Li

Methods S1. Parametrization of PMSPcnn, related to Figure 1.

Considering the limited available training data in our study, we tried some smaller architectures to decide the optimal one with cross-validation, rather than simply adopting the widely used, much larger network architectures in computer vision applications. The network structure of CNN contains six 1D convolutional layers, two Dropout layers, two MaxPooling1D layers, one Flatten layer and a fully connected layer. Each convolutional layer uses Rectified Linear Unit (ReLU) as its activation function. The max pooling of size 3 applies to all max pooling layers. The dropout rates of the first and second dropout layers are set to 0.2 and 0.5 respectively. The batch_size and epoch of CNN model are set to 20 and 1000, respectively. CNN models are trained using the Adam optimizer with learning rate = 0.0001. Mean squared error (MSE) is used as the loss function during the training. The model architecture uses Keras with TensorFlow as the backend. The number of filters in the convolutional layers is tuned using five-fold cross-validation, and the results indicate that the architecture of the 32×64×128 convolutional configuration achieves the best performance (as shown in Figure S4).

For membrane proteins, the model PMSPcnn_m is constructed to predict their stability changes upon mutations. For it, the same five types of characteristics including physicochemical properties, secondary structural features, depth and protrusion indices, SNB-PSSM and PSSM based evolutionary information, and topological features from PH analysis are calculated. And then these features are integrated into the same model framework as that in PMSPcnn. The parameters batch_size and epoch are set to batch_size = 8 and epoch = 500 due to the small samples in M223. Also the 5 fold RScv is used to train the model.

Methods S2. Mapping mutation sites on predicted structures, related to STAR Methods.

In this work, five datasets (Q6422, S^{sym}, p53, myoglobin and M223) are used, all of which provide mutation positions according to the structures of proteins in Protein Data Bank (PDB).[1] Alphafold2 (AlphaFold v2) is used to predict protein structures based on their sequences from Uniprot,[2] which is available on GitHub (https://github.com/deepmind/alphafold). Thus, the mutation sites need to be located on the predicted structures. First, the protein structure from the PDB is downloaded and aligned to the predicted one through PyMOL,[3] and then the mutation positions are mapped on the predicted structure by examining the residues near the mutated residue. If a mismatch is spotted which results from the inconsistency in the residue numbering between PDB structure and its sequence from Uniprot, it will be corrected by cross-checking its sequence and structure information. All the mutation data and the corresponding positions in the sequences from Uniprot are available from the website https://github.com/ChunhuaLab/PMSPcnn.

Methods S3. Performances of SPOT-1D-LM on protein sequences and DSSP on the predicted protein structures by AlphaFold2, related to STAR Methods.

We choose SPOT-1D-LM, which is a good ensemble machine learning method based on convolution and Long Short-Term Memory (LSTM), to extract the secondary structure information. Here, to detect the prediction performances of the two ways: SPOT-1D-LM on protein sequences and DSSP on the predicted protein structures by AlphaFold2, we predicted the secondary structure information (eight states: 3_{10}-helix, α-helix, π-helix, β-strand, bridge, turn, bend and others) for the 147 proteins from the training set with the two ways respectively, and then compared their results with the experimental data (obtained by DSSP on experimentally resolved structures). The average accuracies of the two ways are 0.844 for SPOT-1D-LM on sequences and 0.883 for DSSP on the predicted structures, which indicates that the two ways are comparable. The structures with accuracy below 0.6 are the same for the two ways, which are proteins with PDB ID: 1SSO and 1ZNJ. The better results of 0.597 and 0.586 are obtained for the former way than those of 0.548 and 0.310 for the latter way. The performance of AlphFold2 depends on the multiple sequence alignment (MSA) to some extent, while SPOT-1D-LM does not. The author of SPOT-1D-LM states that the highly accurate prediction of secondary structure information may be made by SPOT-1D-LM without homologous sequences, the remaining obstacle in the post AlphaFold2 era.[4] Considering the dependence of AlphaFold2 on MSA, we used SPOT-1D-LM to extract protein secondary structure information.

Methods S4. Optimizing the region size around mutation site in calculating topological properties, related to STAR Methods.

Topological properties are calculated from the atoms in the area within a distance cutoff of the mutated residue (Cα atom). Determining the appropriate distance cutoff is crucial to the topological property extraction for protein stability change prediction. Generally, the area within a distance of 14 Å from a mutated residue (Cα atom) can cover the range of structural changes caused by the mutation. Thus, to determine the optimal distance cutoff, we performed a comparative analysis of various cutoff values, including 6 Å, 8 Å, 10 Å, 12 Å, and 14 Å. A smaller cutoff may overlook significant structural changes, while a larger cutoff may bring some irrelevant information. The purpose for optimizing the distance cutoff is to hopefully capture the main structural changes caused by residue mutation and meanwhile to avoid excessive noise. Thus, we cross-validated the distances of 6 Å, 8 Å, 10 Å, 12 Å and 14 Å through Extreme Gradient Boosting (XGBoost). In this way, five XGBoost models were generated, and the parameters in each model were adjusted to obtain the optimal model. Finally, the distance cutoff corresponding to the best model among the five is selected as the optimal distance cutoff.

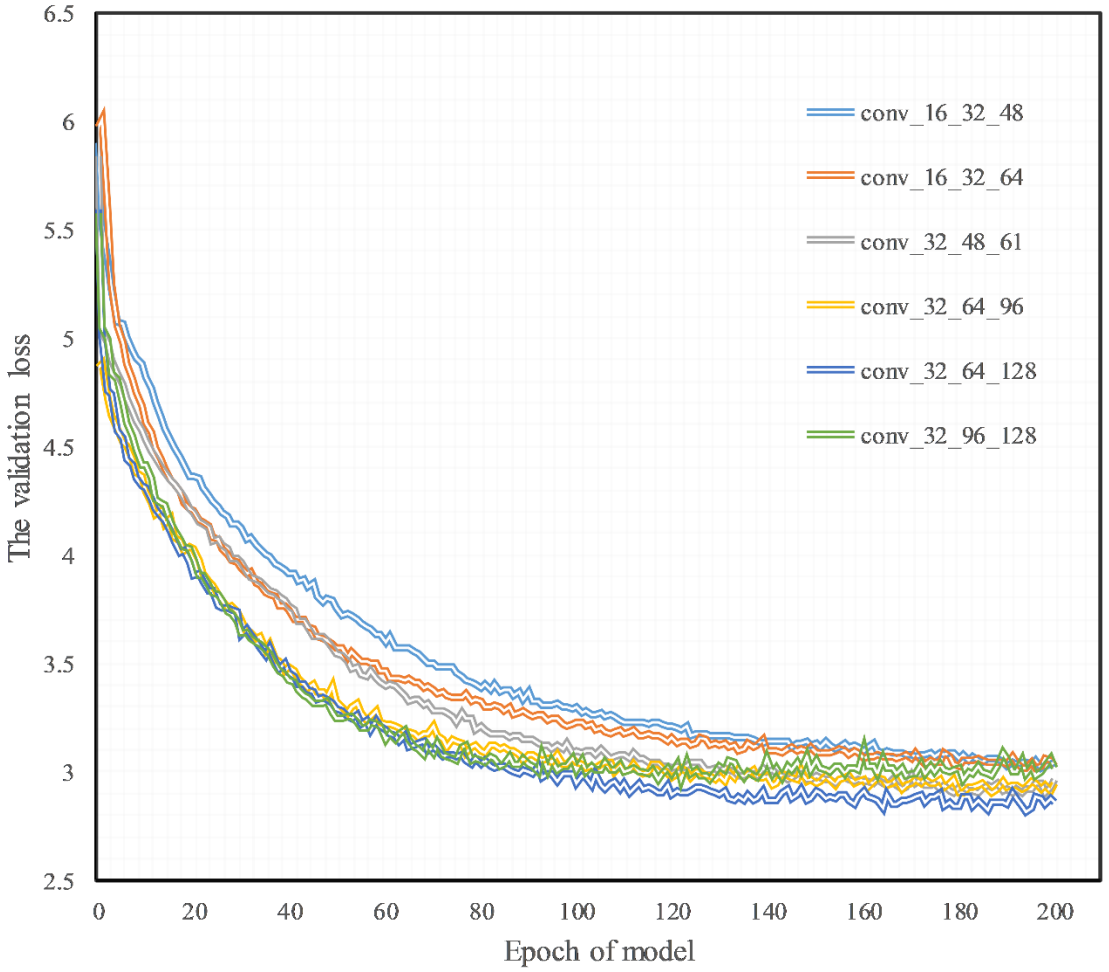

Figure S1. The results from the models with different filter combinations, related to Figure 1. The numbers of filters in convolutional layers are tuned via five-fold cross-validation. Evidently, the architecture of the 32×64×128 convolutional configuration achieves the best performance.

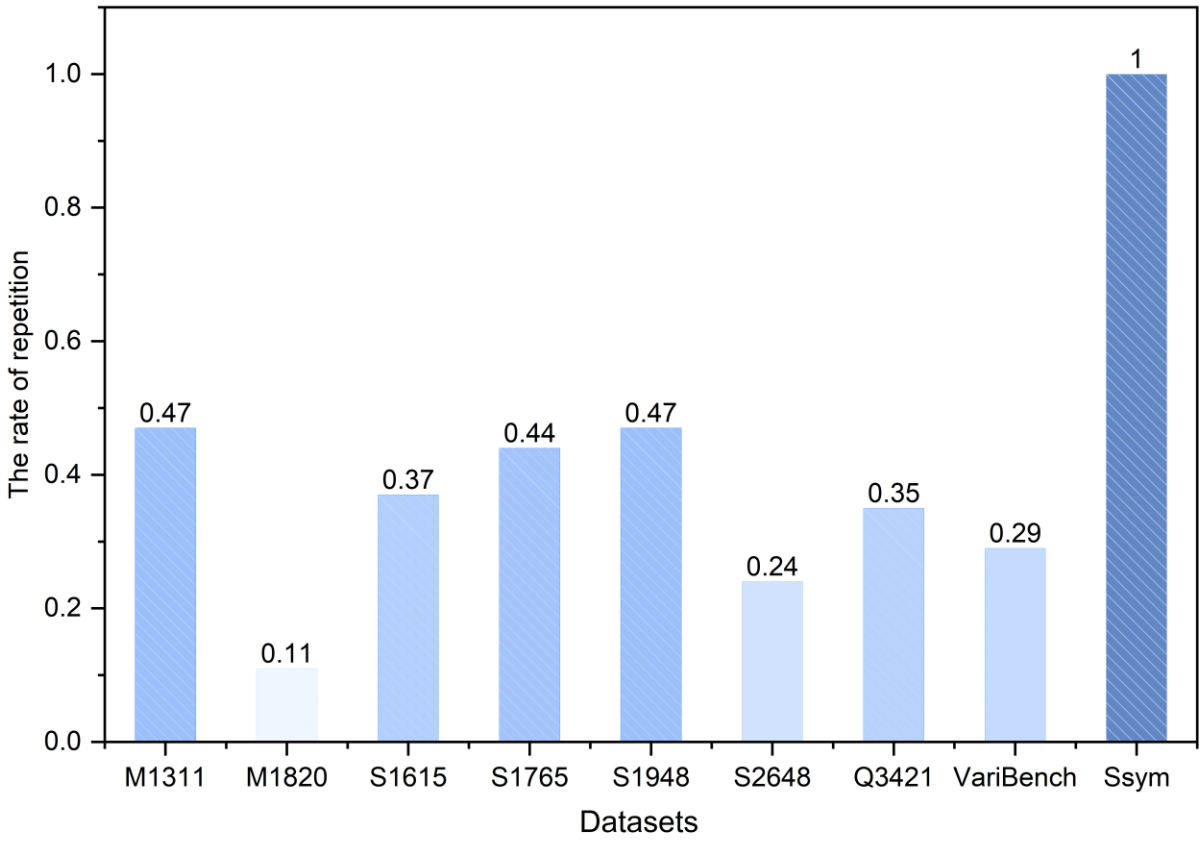

Figure S2. The repetition rate between each of the eight datasets and S^{sym} dataset, related to Table 1. The darker the bar, the higher the repetition rate.

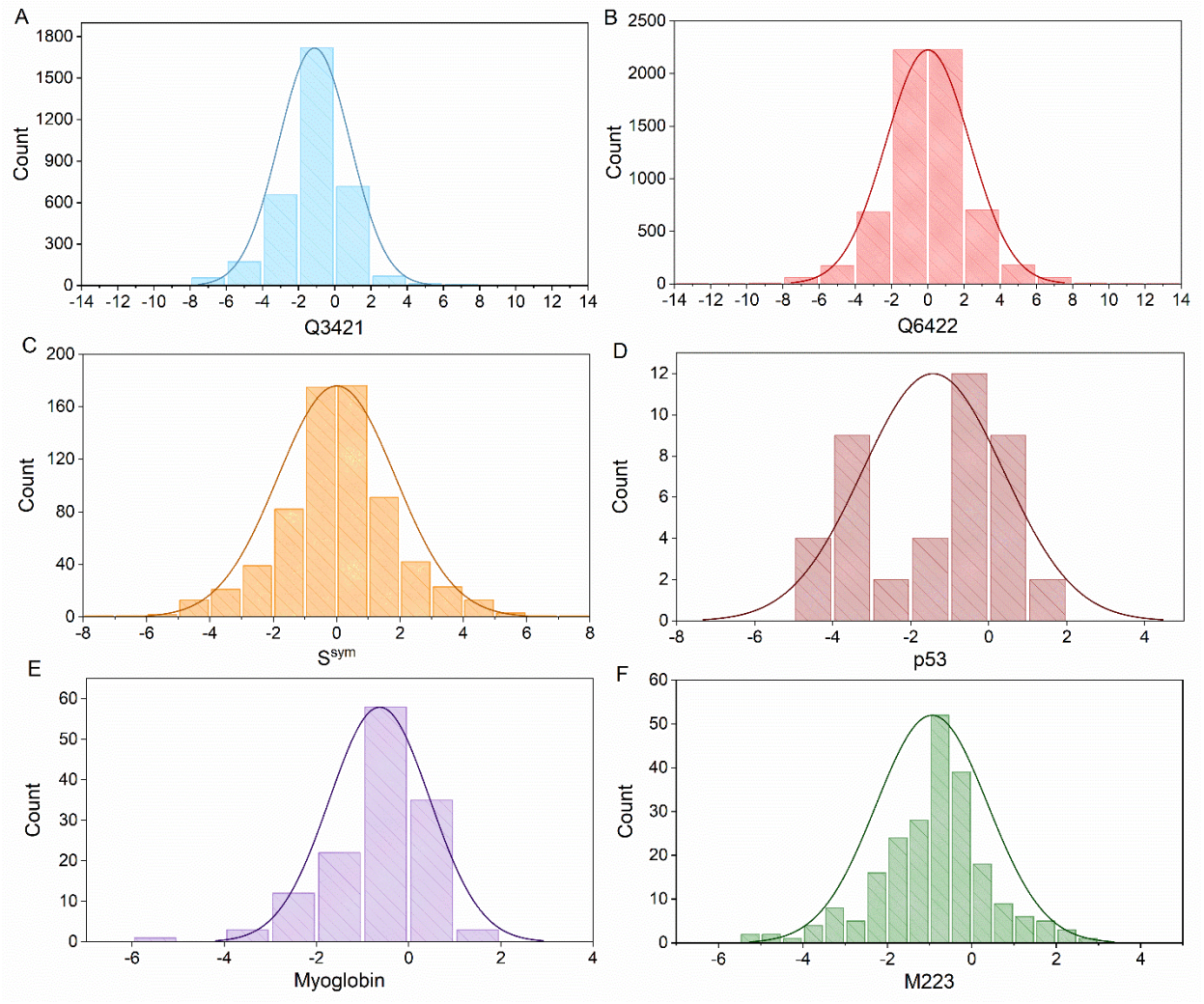

Figure S3. Histogram distributions of the experimental ΔΔG values in six datasets, related to STAR Methods. The six datasets include Q3421 (A), Q6422 (B), S^{sym} (C), p53 (D), Myoglobin (E) and M223 (F).

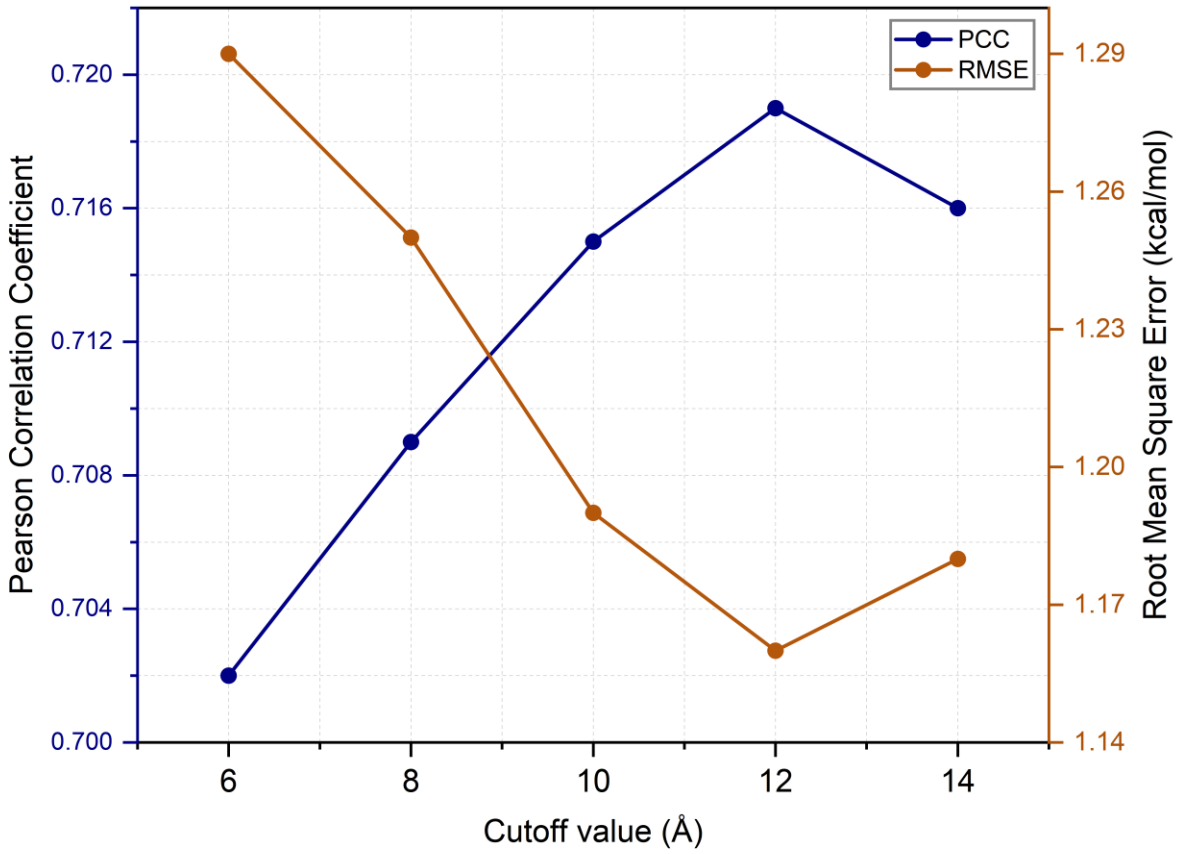

Figure S4. The average results are from the five-fold cross-validation with the distance cutoff being 6 Å, 8 Å, 10 Å, 12 Å and 14 Å respectively, related to STAR Methods. The blue line represents the PCC and the red line represents the RMSE. Evidently, the distance of 12 Å gives the best prediction performance.

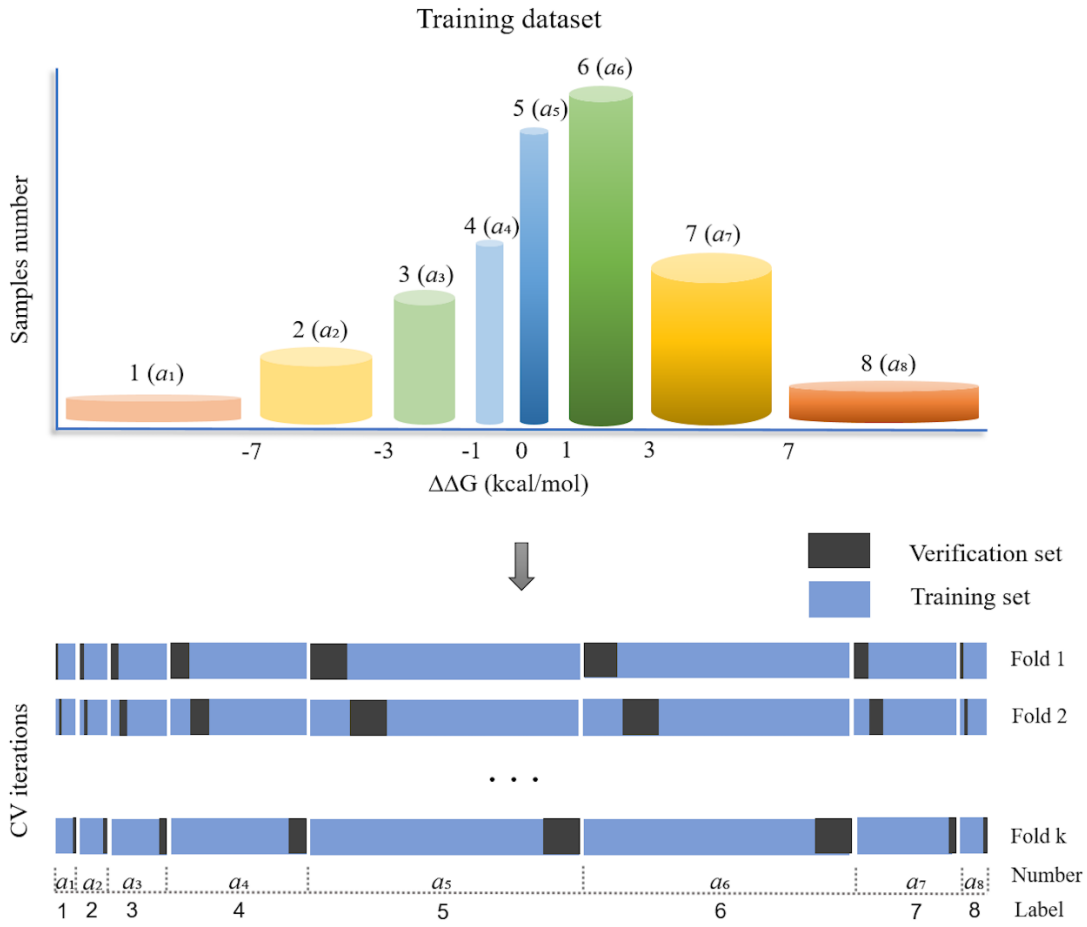

Figure S5. The process of RScv, related to STAR Methods. Step 1: the range of the ΔΔG values in the dataset is divided into eight bins (labeled as 1-8) with different intervals (including (-∞, -7], (-7, -3], (-3, -1], (-1, 0], (0,1], (1, 3], (3, 7], (7, +∞)). Step 2: k-fold cross validation is adopted when training the model. For each fold, the data of each bin is equally divided into k parts with one part of each bin combined as the validation set, while the remaining k-1 parts used for the training set. Step 3: repeat step 2 k times. Finally, n trained sub-models and the corresponding evaluation metrics can be obtained, and their average values are calculated as the final evaluation metric of the predictor. $a_1, a_2, \ldots, a_8$ represent the number of samples per bin. When RScv is applied to a general prediction problem, the number of bins and interval of each bin need to be set based on the distribution of the training data used for model construction. Generally, the number of bins should be set to a suitable value to make sure there are a certain number of samples in each bin. And the bin interval setting needs to ensure the balanced samples in different bins.

Table S1. Detailed information on currently existing methods for protein stability change prediction, related Table 1.

Methods	Basic algorithm	Training Dataset	Overlap of training set with S^{sym}	Anti-symmetry operation
ThermoNet[5]	CNN	Q3214, Q1744	no	yes
DDGun3D[6]	Linear parametric model	S2648, VariBench	yes	yes
DDGun[6]	Linear parametric model	S2648, VariBench	yes	yes
PoPMuSic 2.0[7]	ANN + Statistical	S2648	yes	no
FoldX[8]	Energy function	S339	no	no
Rosetta[9]	Energy function	771	yes	no
MAESTRO[10]	LR, ANN and SVM	S2648, S1948, S1765	yes	no
SDM[11]	Statistical	S2648	yes	yes
PoPMuSiC 2.1[12]	Linear parametric model	S2648	yes	no
mCSM[13]	RF	S2648, S1948	yes	no
DUET[14]	SVM	S2648	yes	no
MUPRO[15]	ANN and SVM	S1615	yes	no
CUPSAT[16]	Energy function	4024	-	no
NeEMO[17]	ANN	S2648	yes	no
AUTOMUTE[18]	SVM, RF	S1948, S1615	yes	no
I-Mutant[19]	SVM	S2648	yes	no
iSTABLE[20]	SVM	M1311, M1820	yes	no
STRUM[21]	GBRT	Q3421	yes	no
ACDC-NN-Seq[22]	CNN	S2648, Varibench	yes	yes
SAAFEC-SEQ[23]	GBDT	S2648	yes	no

Table S3. Predicted ΔΔG values of our model PMSPcnn on the test set p53, related to STAR Methods, related to Figure 6.

	Mutation sites			Experimental	Predicted ΔΔG (kcal/mol)	
PDB ID	Position	Wild	Mutant	ΔΔG (kcal/mol)	Direct	Reverse
2ocj	104	Q	H	-0.24	0.29	0.09
2ocj	104	Q	P	-0.11	-1.43	1.19
2ocj	123	T	A	0.13	1.16	-1.03
2ocj	129	A	D	0.7	-0.23	0.08
2ocj	129	A	E	0.38	-0.04	-0.26
2ocj	129	A	S	0.19	-0.01	-0.05
2ocj	133	M	L	-0.3	-0.35	0.09
2ocj	134	F	L	4.78	1.24	-0.58
2ocj	143	V	A	3.5	2.88	-2.83
2ocj	145	L	Q	2.98	2.77	-3.44
2ocj	148	D	E	0.43	-0.29	0.20
2ocj	148	D	S	-0.22	-0.12	0.32
2ocj	150	T	P	0.08	0.42	-0.30
2ocj	151	P	S	4.49	1.63	-1.15
2ocj	157	V	F	3.88	2.58	-3.00
2ocj	165	Q	K	1.27	-0.07	-0.37
2ocj	167	Q	E	0.43	0.19	0.10
2ocj	168	H	R	2.75	0.29	0.11
2ocj	174	R	K	0.22	1.40	-1.89
2ocj	175	R	A	0.73	2.03	-0.89
2ocj	175	R	H	3.52	2.20	-2.13
2ocj	182	C	S	-0.16	0.63	-0.35
2ocj	195	I	T	4.12	3.21	-3.27
2ocj	201	L	P	-0.35	0.10	-0.31
2ocj	203	V	A	-0.49	0.59	-0.90
2ocj	206	L	S	0.1	1.75	-1.54
2ocj	220	Y	C	3.98	1.28	-2.13
2ocj	228	D	E	-0.05	0.35	0.23
2ocj	232	I	T	3.19	3.42	-2.90
2ocj	236	Y	F	-0.27	1.32	-0.07
2ocj	237	M	I	3.18	1.46	-1.41
2ocj	239	N	Y	-1.49	0.54	-0.39
2ocj	242	C	S	3.07	2.40	-2.29
2ocj	245	G	S	1.21	0.32	-0.24
2ocj	248	R	Q	1.87	0.63	-1.02
2ocj	249	R	S	1.92	2.06	-2.37
2ocj	255	I	F	3.29	0.98	-0.92
2ocj	260	S	P	0.32	0.96	-1.03
2ocj	268	N	D	-1.21	0.70	-0.70
2ocj	270	F	C	4.54	3.22	-3.18

2ocj	273	R	H	0.45	0.66	-1.08
2ocj	282	R	W	3.3	1.25	-1.33

Table S4. Predicted ΔΔG values of our model PMSPcnn on the test set myoglobin, related to Figure 6.

	Mutation sites			Experimental	Predicted ΔΔG (kcal/mol)	
PDB ID	Position	Wild	Mutant	ΔΔG (kcal/mol)	Direct	Reverse
1bz6	4	E	A	0.35	0.38	-0.37
1bz6	7	W	F	0.9	1.31	-1.29
1bz6	8	Q	A	0.89	-0.58	0.61
1bz6	8	Q	G	-0.5	0.63	-0.76
1bz6	9	L	A	0.41	0.74	-0.82
1bz6	11	L	A	0.52	0.90	-0.98
1bz6	13	V	A	0.34	1.21	-1.33
1bz6	14	W	F	1.1	1.12	-1.12
1bz6	15	A	L	-0.1	0.14	0.01
1bz6	18	E	A	0.95	0.59	-0.44
1bz6	20	D	A	0.5	0.83	-0.78
1bz6	23	G	A	1.12	0.57	-0.54
1bz6	24	H	V	0.52	0.37	-0.54
1bz6	28	I	A	1.33	1.42	-1.31
1bz6	28	I	F	0.28	1.21	-1.08
1bz6	28	I	L	0.55	0.54	-0.42
1bz6	28	I	M	0.56	0.62	-0.66
1bz6	28	I	V	-0.04	0.01	0.04
1bz6	29	L	A	0.39	1.09	-1.07
1bz6	29	L	F	0.13	0.82	-1.20
1bz6	29	L	I	1.12	0.98	-0.91
1bz6	29	L	M	-0.12	0.40	-0.27
1bz6	29	L	N	5.59	1.92	-2.12
1bz6	29	L	V	1.72	1.61	-1.53
1bz6	30	I	A	1.9	1.50	-1.78
1bz6	32	L	A	2.04	1.95	-1.93
1bz6	32	L	F	0.07	1.73	-1.79
1bz6	32	L	W	0.53	1.98	-2.98
1bz6	36	H	Q	1.3	0.50	-0.66
1bz6	43	F	I	0.89	1.15	-1.50
1bz6	43	F	V	2.66	1.28	-1.74
1bz6	44	D	A	-0.25	0.46	-0.58
1bz6	48	H	Q	0.62	0.51	-0.52
1bz6	49	L	I	0.8	0.88	-0.85
1bz6	51	T	A	1.41	1.36	-1.41
1bz6	56	K	A	0.35	0.12	-0.19
1bz6	60	D	A	0.15	0.14	-0.24
1bz6	64	H	A	-0.58	0.23	-0.63

1bz6	64	H	F	-1.51	-1.05	-0.21
1bz6	64	H	G	0.04	1.51	-1.88
1bz6	64	H	I	-0.38	0.32	-1.72
1bz6	64	H	L	-1.05	0.35	-1.83
1bz6	64	H	N	0.33	0.51	-0.97
1bz6	64	H	Q	0.54	0.60	-0.78
1bz6	64	H	R	0.08	0.19	-0.70
1bz6	64	H	T	-0.23	0.82	-1.55
1bz6	64	H	V	-0.36	0.59	-1.63
1bz6	64	H	W	-0.47	-0.08	-1.04
1bz6	64	H	Y	-0.52	-1.55	0.29
1bz6	66	V	A	-0.75	-0.48	0.33
1bz6	67	T	A	0.26	0.28	-0.22
1bz6	67	T	F	-0.19	-0.64	-0.05
1bz6	67	T	P	0.85	1.53	-1.88
1bz6	67	T	Q	-0.23	0.77	-1.10
1bz6	68	V	A	0.93	1.65	-1.47
1bz6	68	V	F	-0.82	1.73	-1.38
1bz6	68	V	I	-0.35	0.35	-0.58
1bz6	68	V	L	-0.42	0.27	-0.60
1bz6	68	V	N	3.81	2.48	-2.89
1bz6	68	V	Q	2.42	1.50	-1.84
1bz6	68	V	S	2.99	3.31	-3.37
1bz6	68	V	T	0.6	1.15	-0.88
1bz6	68	V	W	0.07	1.64	-1.94
1bz6	68	V	Y	0.17	2.84	-2.78
1bz6	69	L	A	1.18	1.37	-1.47
1bz6	69	L	I	0.02	0.38	-0.33
1bz6	69	L	M	0	0.12	-0.19
1bz6	69	L	V	0.1	0.52	-0.61
1bz6	77	K	A	-0.2	0.34	-0.31
1bz6	82	H	Q	0.05	0.31	-0.45
1bz6	88	P	A	-0.59	-0.46	0.36
1bz6	93	H	G	-0.04	0.22	-0.39
1bz6	97	H	Q	0.11	0.37	-0.18
1bz6	106	F	A	0.7	0.27	-0.33
1bz6	107	I	A	0.43	3.10	-3.47
1bz6	107	I	F	-0.58	1.46	-1.63
1bz6	107	I	T	2.23	2.91	-1.98
1bz6	107	I	V	1.09	0.28	-0.01
1bz6	109	E	A	0.17	-0.27	0.17
1bz6	109	E	G	-0.89	1.04	-1.00
1bz6	111	I	A	1.84	2.05	-2.14
1bz6	111	I	L	0.64	0.54	-0.39
1bz6	111	I	M	1.14	1.23	-1.01

1bz6	113	H	Q	0.26	0.15	-0.26
1bz6	114	V	A	1.45	1.44	-1.48
1bz6	115	L	A	1.4	1.58	-1.66
1bz6	116	H	A	-0.16	-0.10	0.07
1bz6	117	S	A	0.26	0.07	-0.12
1bz6	118	R	A	0.65	1.45	-0.74
1bz6	119	H	F	0.68	0.04	-0.34
1bz6	122	D	A	0.1	-0.01	-0.09
1bz6	123	F	A	1.1	0.79	-0.78
1bz6	123	F	K	2.1	2.09	-2.09
1bz6	123	F	T	3.5	3.39	-3.28
1bz6	125	A	L	-0.6	-0.16	0.15
1bz6	129	G	A	-1.05	-0.83	0.83
1bz6	130	A	K	3.7	1.91	-1.79
1bz6	130	A	L	2.3	1.37	-1.30
1bz6	131	M	A	2.2	2.22	-1.98
1bz6	133	K	A	0.05	-0.22	0.24
1bz6	135	L	A	1.7	1.34	-1.39
1bz6	135	L	I	1.54	1.46	-1.40
1bz6	135	L	M	0.79	0.70	-0.72
1bz6	135	L	V	2.25	2.19	-2.14
1bz6	137	L	A	1.78	2.19	-1.82
1bz6	139	R	A	0.45	1.52	-0.55
1bz6	140	K	A	0.35	-0.16	0.16
1bz6	142	I	A	1.92	1.62	-1.80
1bz6	142	I	L	-0.63	-0.28	0.20
1bz6	142	I	M	-0.93	-0.31	0.35
1bz6	142	I	V	0.12	0.33	-0.51
1bz6	144	A	L	-0.4	0.05	-0.07
1bz6	149	L	A	1.6	1.52	-1.43

Table S5. The detailed information on the datasets used in this work, related to STAR Methods.

Dataset	Reference	Num. of mutations (Proteins)	Num. of stabilizing mutations (Proteins)	Num. of destabilizing mutations (Proteins)	Usage for
Q6422	Our work	6422 (147)	3211 (147)	3211 (147)	Training
M223	Kroncke et al.[24]	223 (7)	42 (6)	181 (7)	Testing
S^{sym}	Pucci et al.[25]	684 (30)	342 (15)	342 (15)	Testing
p53	Pires et al.[13]	42 (1)	11 (1)	31 (1)	Testing
myoglobin	Kepp et al.[26]	134 (1)	38 (1)	96 (1)	Testing

REFERENCES

1. Rose, P.W., Prlic, A., Altunkaya, A., Bi, C., Bradley, A.R., Christie, C.H., Costanzo, L.D., Duarte, J.M., Dutta, S., and Feng, Z., et al. (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res 45, D271-D281. 10.1093/nar/gkw1000.
2. Bateman, A., Martin, M., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., and Bursteinas, B., et al. (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49, D480-D489. 10.1093/nar/gkaa1100.
3. Delano W. The pymol molecular graphics system, Proteins Structure Function and Bioinformatics 2002;30:442-454.
4. Singh, J., Paliwal, K., Litfin, T., Singh, J., and Zhou, Y. (2022). Reaching alignment-profile-based accuracy in predicting protein secondary and tertiary structural properties without alignment. Sci Rep-Uk 12, 7607. 10.1038/s41598-022-11684-w.
5. Li, B., Yang, Y.T., Capra, J.A., and Gerstein, M.B. (2020). Predicting changes in protein thermodynamic stability upon point mutation with deep 3D convolutional neural networks. Plos Comput Biol 16, e1008291. 10.1371/journal.pcbi.1008291.
6. Montanucci, L., Capriotti, E., Frank, Y., Ben-Tal, N., and Fariselli, P. (2019). DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. Bmc Bioinformatics 20, 335. 10.1186/s12859-019-2923-1.
7. Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P., and Rooman, M. (2009). Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. Bioinformatics 25, 2537-2543. 10.1093/bioinformatics/btp445.
8. Guerois, R., Nielsen, J.E., and Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320, 369-387. 10.1016/S0022-2836(02)00442-4.
9. Kellogg, E.H., Leaver-Fay, A., and Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins 79, 830-838. 10.1002/prot.22921.
10. Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S., and Lackner, P. (2015). MAESTRO--multi agent stability prediction upon point mutations. Bmc Bioinformatics 16, 116. 10.1186/s12859-015-0548-6.
11. Worth, C.L., Preissner, R., and Blundell, T.L. (2011). SDM--a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res 39, W215-W222. 10.1093/nar/gkr363.
12. Dehouck, Y., Kwasigroch, J.M., Gilis, D., and Rooman, M. (2011). PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. Bmc Bioinformatics 12, 151. 10.1186/1471-2105-12-151.
13. Pires, D.E., Ascher, D.B., and Blundell, T.L. (2014). mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics 30, 335-342. 10.1093/bioinformatics/btt691.
14. Pires, D.E., Ascher, D.B., and Blundell, T.L. (2014). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res 42, W314-W319. 10.1093/nar/gku411.

15. Cheng, J., Randall, A., and Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62, 1125-1132. 10.1002/prot.20810.
16. Parthiban, V., Gromiha, M.M., and Schomburg, D. (2006). CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res 34, W239-W242. 10.1093/nar/gkl190.
17. Giollo, M., Martin, A.J., Walsh, I., Ferrari, C., and Tosatto, S.C. (2014). NeEMO: a method using residue interaction networks to improve prediction of protein stability upon mutation. Bmc Genomics 15, S7. 10.1186/1471-2164-15-S4-S7.
18. Masso, M., and Vaisman, I.I. (2008). Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. Bioinformatics 24, 2002-2009. 10.1093/bioinformatics/btn353.
19. Capriotti, E., Fariselli, P., and Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33, W306-W310. 10.1093/nar/gki375.
20. Chen, C.W., Lin, J., and Chu, Y.W. (2013). iStable: off-the-shelf predictor integration for predicting protein stability changes. Bmc Bioinformatics 14 Suppl 2, S5. 10.1186/1471-2105-14-S2-S5.
21. Quan, L., Lv, Q., and Zhang, Y. (2016). STRUM: structure-based prediction of protein stability changes upon single-point mutation. Bioinformatics 32, 2936-2946. 10.1093/bioinformatics/btw361.
22. Pancotti, C., Benevenuta, S., Repetto, V., Birolo, G., Capriotti, E., Sanavia, T., and Fariselli, P. (2021). A Deep-Learning Sequence-Based Method to Predict Protein Stability Changes Upon Genetic Variations. Genes-Basel 12, 911. 10.3390/genes12060911.
23. Li, G., Panday, S.K., and Alexov, E. (2021). SAAFEC-SEQ: A Sequence-Based Method for Predicting the Effect of Single Point Mutations on Protein Thermodynamic Stability. Int J Mol Sci 22. 10.3390/ijms22020606.
24. Kroncke, B.M., Duran, A.M., Mendenhall, J.L., Meiler, J., Blume, J.D., and Sanders, C.R. (2016). Documentation of an Imperative To Improve Methods for Predicting Membrane Protein Stability. Biochemistry-Us 55, 5002-5009. 10.1021/acs.biochem.6b00537.
25. Pucci, F., Bernaerts, K.V., Kwasigroch, J.M., and Rooman, M. (2018). Quantification of biases in predictions of protein stability changes upon mutations. Bioinformatics 34, 3659-3665. 10.1093/bioinformatics/bty348.
26. Kepp, K.P. (2015). Towards a "Golden Standard" for computing globin stability: Stability and structure sensitivity of myoglobin mutants. Biochim Biophys Acta 1854, 1239-1248. 10.1016/j.bbapap.2015.06.002.