Electronic Supplementary Material (ESI) for New Journal of Chemistry.

Supplementary Information

Enhanced Thermostability of *Streptomyces mobaraenesis* Transglutaminase via Computation-aided Site-directed Mutations and Structural Analysis

Yongzhen Li[1], Banghao Wu[1], Yumeng Zhang[1],
Lanxuan Liu[2], Linquan Bai[1], Ting Shi[1,*]

[1]State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

[2]Digital Innovation of AI, WuXi Biologics, Shanghai 201400, China

***Corresponding author:**

Ting Shi, Email: tshi@sjtu.edu.cn.

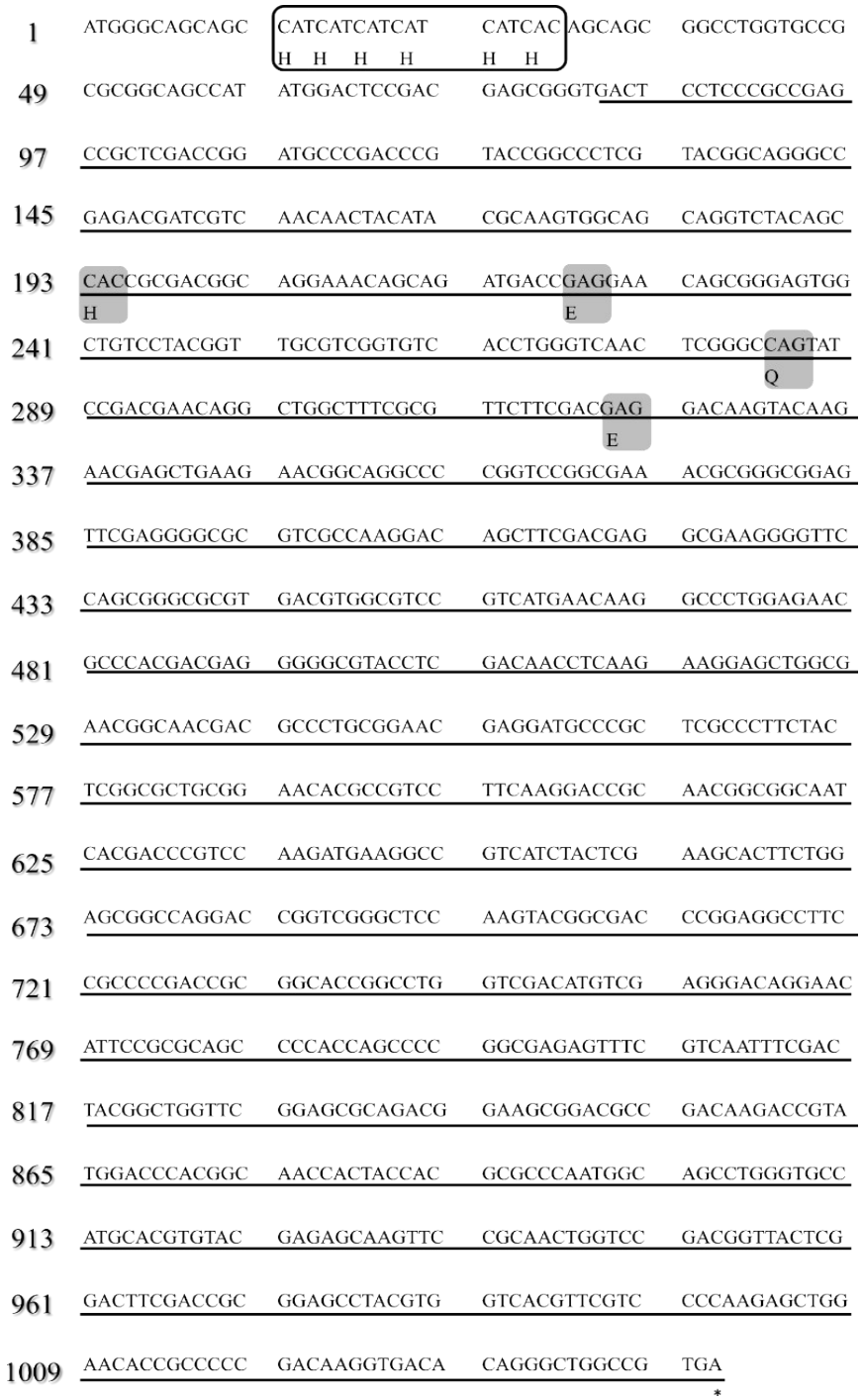

Fig. S1 The sequence analysis of smTG. Numbers (left side): the number of nucleotide sequences; Single underline: the region of mature smTG; Gray background: the mutation sites of the wild-type smTG; Box: His-tag.

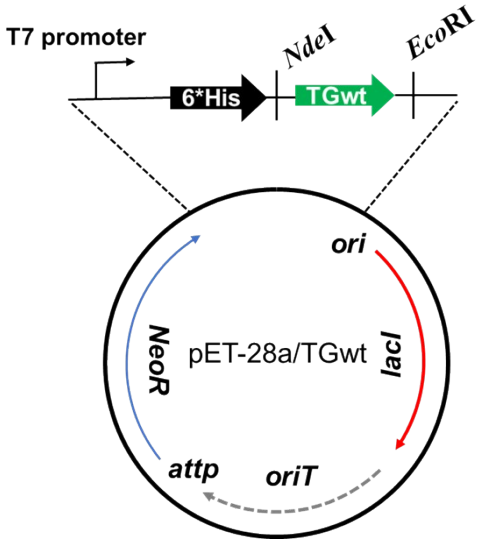

Fig. S2 Schematic diagram of plasmid pET28a/TGwt.

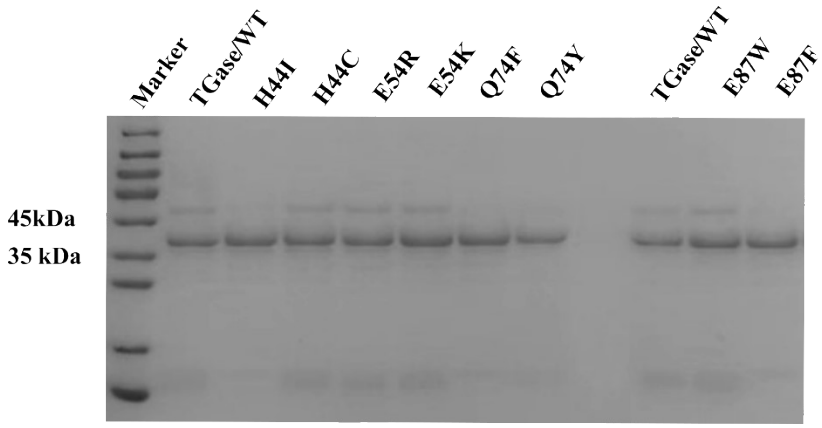

Fig. S3 SDS-PAGE analysis of the TGm1 and its variants purified by affinity chromatography. The Mu-TGase and mutants were intracellularly expressed in *E. coli* BL21 (DE3) pGro7 using pET-28a/TGwt and its derivatives.

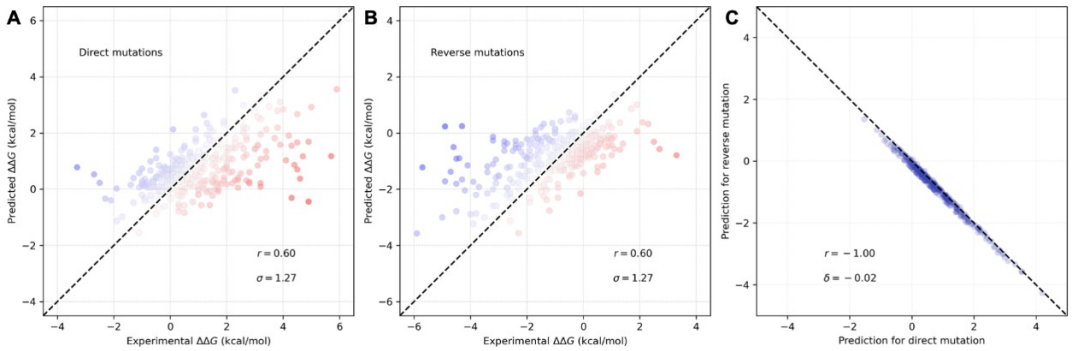

Fig. S4 Model performance on S^{sym} dataset. (A) Model performance on predicting ΔΔG for direct mutations; (B) Model performance on predicting ΔΔG for reverse mutations; (C) Direct versus reverse ΔΔG values of all the mutations in the test set. The color of points in A and B represents the difference between experimental and predicted ΔΔG (Red: $\Delta\Delta G_{exp} > \Delta\Delta G_{pred}$, Blue: $\Delta\Delta G_{exp} < \Delta\Delta G_{pred}$).

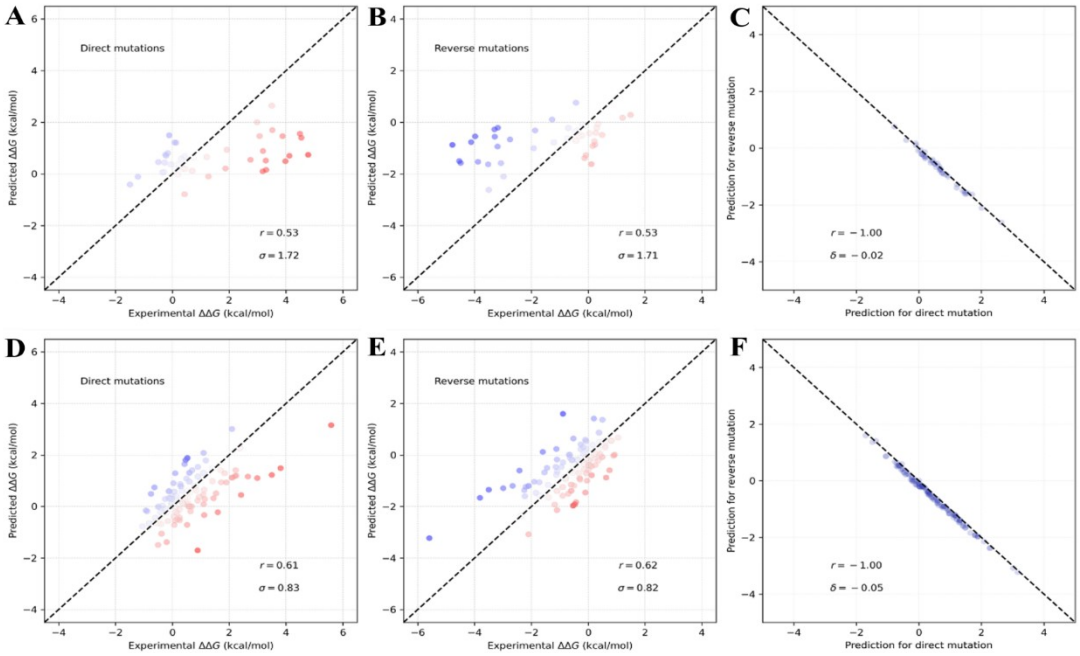

Fig. S5 Model predicts the mutation effects on protein stability well in p53 protein and myoglobin. A, D. Model performance on predicting ΔΔG for direct mutations; B, E. Model performance on predicting ΔΔG for reverse mutations; C, F. Direct versus reverse ΔΔG values of all the mutations in the test set.

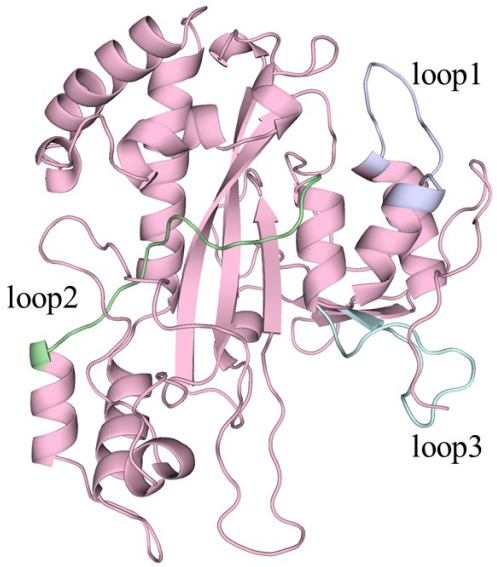

Fig. S6 The crystal structure of smTG (PDB ID: 6gmg). The residue regions used for prediction are shown in different colors. Loop1 contains residues 42-54 in purple, loop2 contains residues 72-87 in green, and loop3 contains residues 274-289 in cyan.

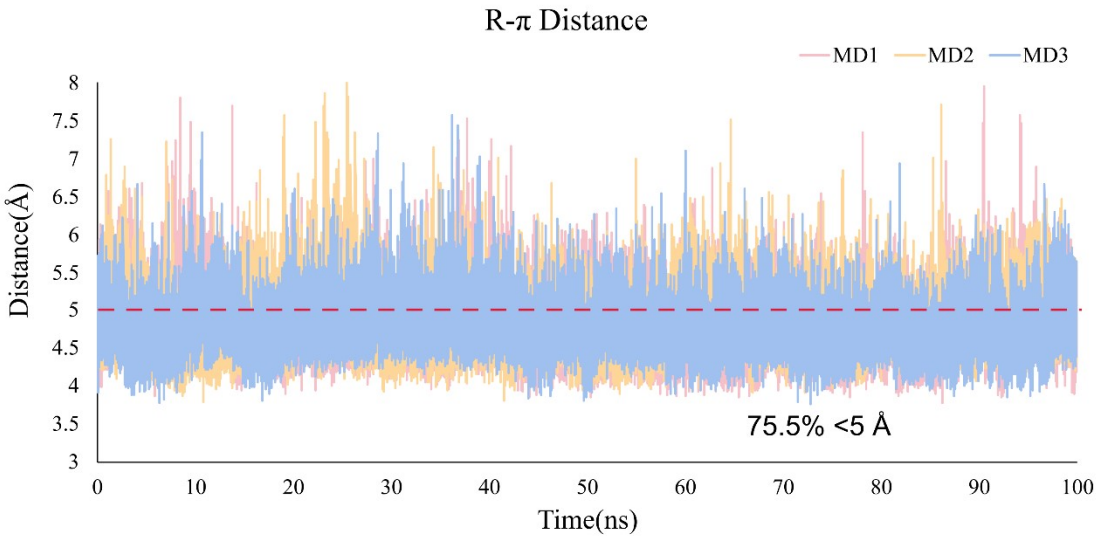

Fig. S7 Statistics of the ring-to-ring distance between F74 and P177. More than 75% of the time in 100 ns is within 5 Å distance.

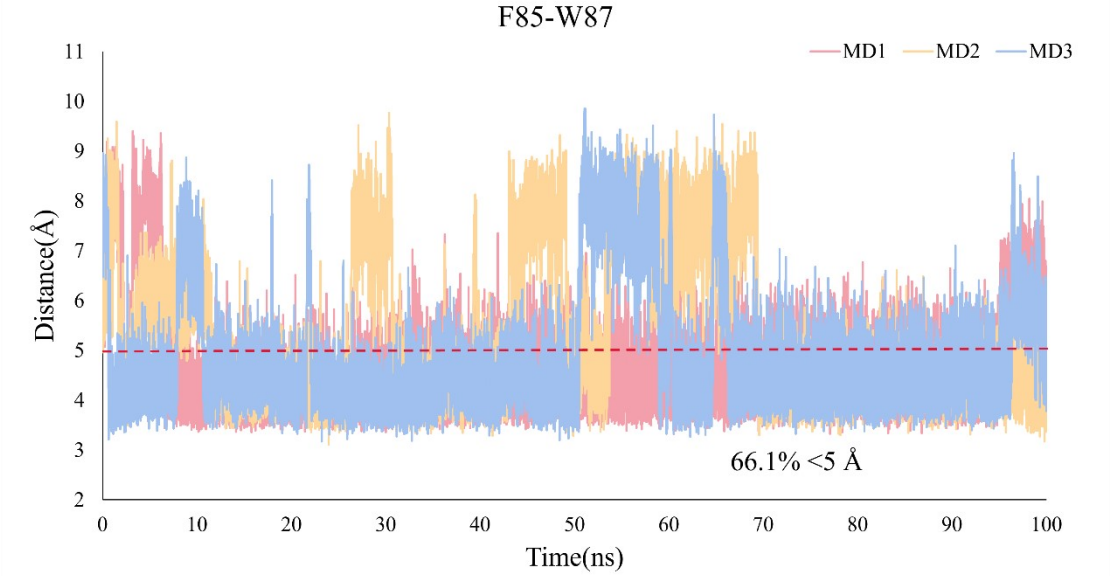

Fig. S8 Statistical results of side-chain distances for F85 and W87. More than 66.1% of the time in 100 ns is within 5 Å distance.

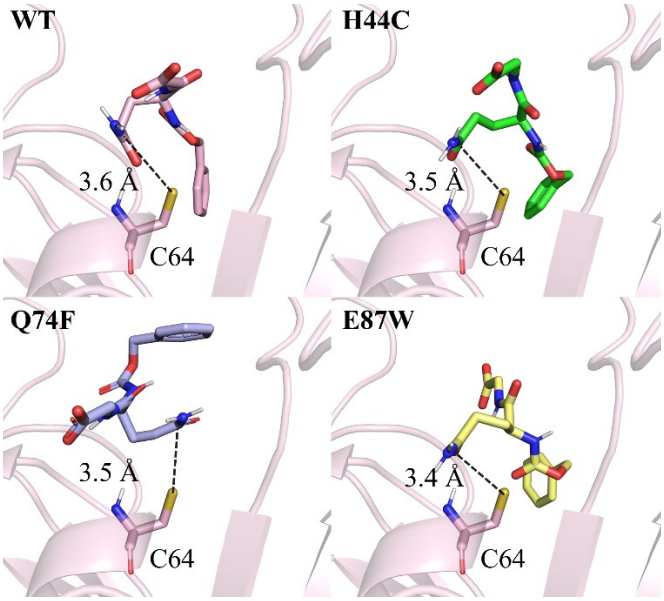

Fig. S9 Docking poses of CBZ-Gln-Gly in different systems.

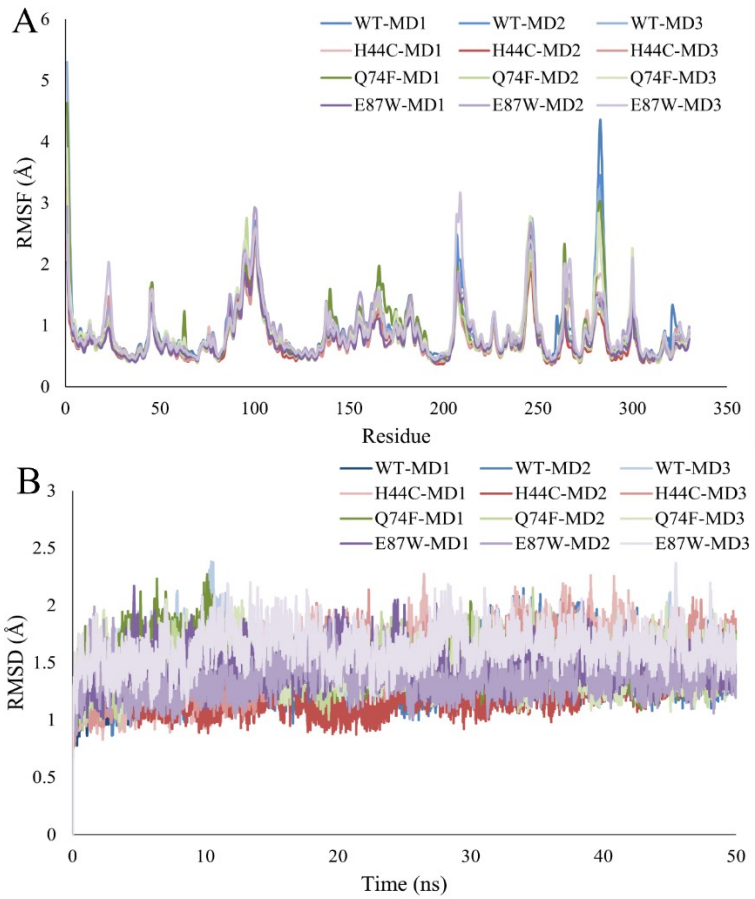

Fig. S10 RMSF and RMSD of wild-type and mutant smTG with substrate.

Table. S1 Primers used in this study.

Primers	Sequences (5′-3′)	Mutations
MTGase-F	CCGGAATTCTCACGGCCAGCCCTGTGTCAC	
MTGase-R	CGCCATATGGACTCCGACGAGCGGGTGAC	
Mutation1-F	GCAGGAAACAGCAGATGACCAAGGAACAG CGGGAGTGGCTGTC	E54K
Mutation2-F	GCAGGAAACAGCAGATGACCCGCGAACAG CGGGAGTGGCTGTC	E54R
Mutation1-R	GGTCATCTGCTGTTTCCTGCCGTCGCGGTG GCTGTAGACC	E54K, E54R
Mutation3-F	TCACCTGGGTCAACTCGGGCTACTATCCG ACGAACAGGCTGGC	Q74Y
Mutation4-F	TCACCTGGGTCAACTCGGGCTTCTATCCG ACGAACAGGCTGGC	Q74F
Mutation2-R	GCCCGAGTTGACCCAGGTGACACCGACG CAACCGTAGGAC	Q74Y, Q74F
Mutation5-F	AGTGGCAGCAGGTCTACAGCTGCCGCGA CGGCAGGAAACAGCA	H44C
Mutation6-F	AGTGGCAGCAGGTCTACAGCATACGCGA CGGCAGGAAACAGCA	H44I
Mutation3-R	GCTGTAGACCTGCTGCCACTTGCGTATGT AGTTGTTGACG	H44C, H44I
Mutation7-F	TGGCTTTCGCGTTCTTCGACTGGGACAAG TACAAGAACGAGCT	E87W
Mutation8-F	TGGCTTTCGCGTTCTTCGACTTCGACAAG TACAAGAACGAGCT	E87F
Mutation4-R	GTCGAAGAACGCGAAAGCCAGCCTGTTC GTCGGATACTGG	E87F, E87W

The mutation sites were underlined.

Table. S2 The ablation study on the selection of the GNN operator, radius of mutation neighborhood, and node features.

GNN Operator	σ_{dir}	r_{dir}	σ_{rev}	r_{rev}
GCN	0.53	0.54	1.40	1.40
GIN	0.58	0.57	1.37	1.37
GAT	**0.60**	**0.60**	**1.27**	**1.27**
Neighborhood radius (Å)	σ_{dir}	r_{dir}	σ_{rev}	r_{rev}
10	0.59	0.59	1.30	1.30
12	**0.60**	**0.60**	**1.27**	**1.27**
14	0.55	0.54	1.33	1.33
16	0.52	0.53	1.38	1.37
Feature encoding	σ_{dir}	r_{dir}	σ_{rev}	r_{rev}
w/o Evolutionary	0.55	0.56	1.34	1.33
w/o Energy	0.36	0.35	1.64	1.65
w/o Amino acid	0.54	0.54	1.34	1.34
All	**0.60**	**0.60**	**1.27**	**1.27**

Table. S3 Performance comparison on balanced test set S^{sym}.

Method	σ_{dir}	r_{dir}	σ_{rev}	r_{rev}	$r_{dir\text{-}rev}$	$\langle\delta\rangle$
FoldX	1.56	0.63	2.13	0.39	-0.38	-0.47
Rosetta	2.31	0.69	2.61	0.43	-0.41	-0.69
CUPSAT	1.71	0.39	2.88	0.05	-0.54	-0.72
PoPMuSiC v2.1	1.21	0.63	2.18	0.25	-0.29	-0.71
STRUM	1.05	0.75	2.51	-0.15	0.34	-0.87
MAESTRO	1.36	0.52	2.09	0.32	-0.34	-0.58
mCSM	1.23	0.61	2.43	0.14	-0.26	-0.91
MUPRO	0.94	0.79	2.51	0.07	-0.02	-0.97
INPS	0.51	0.50	1.42	1.44	-0.99	-0.04
I-Mutant 3.0	1.23	0.62	2.32	-0.04	0.02	-0.68
DDGun3D	1.42	0.56	1.46	0.53	-0.99	-0.02
iSTABLE	1.10	0.72	2.28	-0.08	0.02	-0.68
DUET	1.20	0.63	2.38	0.13	-0.21	-0.84
NeEMO	1.08	0.72	2.35	0.02	0.09	-0.60
ThermoNet	1.56	0.47	1.55	0.47	-0.96	-0.01
ACDC-NN	1.45	0.57	1.45	0.57	-1.00	-0.00
Ours	**1.27**	**0.60**	**1.27**	**0.60**	**-1.00**	**-0.02**

σ_{dir} and r_{dir} are the root mean square error and the Pearson correlation coefficient between the predicted and experimental ΔΔG values for the direct mutations. σ_{rev} and r_{rev} are the root mean square error and the Pearson correlation coefficient for the reverse mutations. The antisymmetry between direct and reverse mutations is assessed using the correlation coefficient $r_{dir\text{-}rev}$ and the average bias $\langle\delta\rangle$. A perfectly unbiased predictor should have $r_{dir\text{-}rev}$= -1 and $\langle\delta\rangle$ = 0 kcal/mol.

Table. S4 Prediction results on p53 protein.

Method	σ_{dir}	r_{dir}	σ_{rev}	r_{rev}	$r_{dir\text{-}rev}$	$\langle\delta\rangle$
ThermoNet	2.01	0.45	1.92	0.56	-0.93	-0.04
ACDC-NN	1.69	0.61	1.69	0.61	-1.00	-0.00
Ours	1.72	0.53	1.71	0.53	-1.00	-0.02

The TP53 tumor suppressor gene encodes the p53 transcription factor that is mutated in about 45% of all human cancers. Single amino acid substitutions in p53 often affect DNA binding or disrupt the conformation and stability of p53 protein.[1] The TP53 tumor suppressor gene encodes the p53 transcription factor that is mutated in about 45% of all human cancers. Single amino acid substitutions in p53 often affect DNA binding or disrupt the conformation and stability of p53 protein. With the same data-processing pipeline, the model capacity of estimating the thermodynamic effects derived from mutations in p53. The results were compared with ThermoNet and ACDC-NN.

Table. S5 Prediction results on myoglobin.

Method	σ_{dir}	r_{dir}	σ_{rev}	r_{rev}	$r_{dir\text{-}rev}$	$\langle\delta\rangle$
ThermoNet	1.16	0.38	1.18	0.37	-0.97	-0.02
ACDC-NN	0.89	0.58	0.89	0.58	-1.00	-0.00
Ours	0.83	0.61	0.82	0.62	-1.00	-0.05

Myoglobin regulates cellular oxygen concentration in cardiac myocytes and oxidative skeletal muscle fibers by reversible binding of oxygen.[2] The myoglobin dataset includes 134 mutations throughout the chain (PDB ID: 1BZ6). With the same data-processing pipeline, the model capacity of estimating the thermodynamic effects derived from mutations in myoglobin protein was examined. The results were compared with ThermoNet and ACDC-NN.

Table. S6 The prediction of mutations with ΔΔG value below -0.5 kcal/mol.

PDB	Position	WT	MUT	ΔΔG (kcal/mol)	PDB	Position	WT	MUT	ΔΔG (kcal/mol)
1iu4A	43	S	F	-0.87	1iu4A	74	Q	I	-0.92
1iu4A	44	H	V	-0.95	1iu4A	74	Q	C	-0.84
1iu4A	44	H	C	-0.75	1iu4A	74	Q	L	-0.8
1iu4A	44	H	I	-0.73	1iu4A	74	Q	M	-0.63
1iu4A	44	H	F	-0.72	1iu4A	74	Q	T	-0.61
1iu4A	44	H	Y	-0.72	1iu4A	74	Q	V	-0.59
1iu4A	44	H	W	-0.66	1iu4A	76	P	Y	-1.23
1iu4A	44	H	T	-0.65	1iu4A	78	N	M	-1.19
1iu4A	44	H	P	-0.5	1iu4A	78	N	C	-0.76
1iu4A	48	R	P	-0.98	1iu4A	79	R	I	-0.86
1iu4A	49	K	Y	-0.57	1iu4A	79	R	T	-0.81
1iu4A	49	K	F	-0.51	1iu4A	79	R	V	-0.73
1iu4A	49	K	P	-0.51	1iu4A	83	A	W	-0.66
1iu4A	54	E	K	-0.87	1iu4A	84	S	C	-0.9
1iu4A	54	E	L	-0.81	1iu4A	84	S	F	-0.88
1iu4A	54	E	R	-0.62	1iu4A	84	S	Y	-0.81
1iu4A	54	E	I	-0.6	1iu4A	84	S	W	-0.55
1iu4A	54	E	C	-0.59	1iu4A	84	S	I	-0.54
1iu4A	54	E	P	-0.58	1iu4A	85	F	W	-0.65
1iu4A	54	E	T	-0.58	1iu4A	86	D	L	-0.75
1iu4A	54	E	F	-0.56	1iu4A	86	D	I	-0.57
1iu4A	54	E	Y	-0.55	1iu4A	87	E	Y	-2.03
1iu4A	72	S	L	-0.71	1iu4A	87	E	F	-1.79
1iu4A	73	G	M	-0.69	1iu4A	87	E	W	-1.18
1iu4A	73	G	C	-0.65	1iu4A	87	E	L	-0.91
1iu4A	74	Q	Y	-1.25	1iu4A	87	E	H	-0.75
1iu4A	74	Q	F	-1.13	1iu4A	87	E	C	-0.7
1iu4A	74	Q	P	-1.1	1iu4A	87	E	V	-0.52

Table. S7 Residual enzyme activities of wide-type and mutant smTG with different incubation time at 60 ℃.

Incubation time at 60 ℃	0 min	10 min	20 min	30 min
TGase (U/mg)	27.68	1.34	0.40	0.17
H44C (U/mg)	25.64	2.01	0.92	0.21
H44I (U/mg)	32.81	2.17	0.41	0.26
E54K (U/mg)	27.50	0.40	0.21	0.17
E54R (U/mg)	31.97	1.16	0.25	0.11
Q74Y (U/mg)	30.44	0.78	0.29	0.31
Q74F (U/mg)	27.80	1.84	3.07	0.62
E87W (U/mg)	21.93	1.08	2.33	0.45
E87F (U/mg)	28.63	1.16	0.71	0.43

Table. S8 Hydrogen bond interactions more than 80%.

Sys	Acceptor	Donor	Frac	Sys	Acceptor	Donor	Frac
WT	E58@O	S61@OG	0.9741	**Q74F**	E58@O	S61@OG	0.9652
	S293@O	W272@N	0.9411		S293@O	W272@N	0.9329
	A267@O	T270@OG1	0.9000		N78@OD1	Y198@OH	0.9051
	Y198@O	V311@N	0.8963		Y34@O	W38@N	0.8841
	N78@O	Y198@OH	0.8810		N33@OD1	D18@N	0.8797
	Y34@O	W38@N	0.8737		V112@O	S116@OG	0.8748
	V271@O	Q262@NE2	0.8723		Y198@O	V311@N	0.8686
	N33@O	D18@N	0.8678		V271@O	Q262@NE2	0.8674
	D306@O	R307@NH1	0.8618		D306@O	R307@NH1	0.8593
	T29@O	V290@N	0.8598		D306@OD2	S204@OG	0.8470
	Y310@O	F82@N	0.8564		D255@O	HIP201@N	0.8387
	V112@O	S116@OG	0.8487		HID274@O	Y291@N	0.8386
	D306@O	S204@N	0.8312		T29@O	V290@N	0.8357
	D306@O	S204@OG	0.8306		Y310@O	F82@N	0.8355
	I315@O	K194@N	0.8216		R192@O	K317@N	0.8264
					Q39@OE1	M52@N	0.8154
					D306@OD1	S204@N	0.8153
					A173@O	T177@OG1	0.8024
H44C	E58@O	S61@OG	0.9830	**E87W**	E58@O	S61@OG	0.9842
	S293@O	W272@N	0.9189		S293@O	W272@N	0.9337
	N78@OD1	Y198@OH	0.9088		N78@OD1	Y198@OH	0.8932
	Y198@O	V311@N	0.8856		Y198@O	V311@N	0.8875
	V271@O	Q262@NE2	0.8852		Y34@O	W38@N	0.8740
	N33@OD1	D18@N	0.8794		N33@OD1	D18@N	0.8721
	Y34@O	W38@N	0.8753		V271@O	Q262@NE2	0.8634
	A267@O	T270@OG1	0.8713		D306@O	R307@NH1	0.8621
	D306@O	R307@NH1	0.8604		Y310@O	F82@N	0.8598
	V112@O	S116@OG	0.8539		V112@O	S116@OG	0.8559
	Y310@O	F82@N	0.8482		D306@OD2	S204@OG	0.8446
	T29@O	V290@N	0.8349		D306@OD1	S204@N	0.8406
	D255@O	HIP201@N	0.8341		HID274@O	Y291@N	0.8398
	R192@O	K317@N	0.8298		T29@O	V290@N	0.8349
	T68@O	S72@OG	0.8248		D255@O	HIP201@N	0.8267
	D306@OD2	S204@OG	0.8234		R192@O	K317@N	0.8244
	D306@OD1	S204@N	0.8230		I315@O	K194@N	0.8058
	Q39@OE1	M52@N	0.8225		Q39@OE1	M52@N	0.8030
	I315@O	K194@N	0.8112				

Table. S9 DSSP analysis of wild protein and Q74F about residue 177,178,179,180.

Sys	State	Res	Para	Anti	3-10 Helix	Alpha	Pi	Turn	Bend
WT	DSSP1	177	0	0	0.1752	0.5765	0	0.1935	0.0548
		178	0	0	0.1988	0.5765	0	0.2206	0.0041
		179	0	0	0.1988	0.5765	0	0.1888	0.0360
		180	0	0	0.0917	0.5765	0	0.1929	0.1389
	DSSP2	177	0	0	0.1923	0.5692	0	0.1929	0.0456
		178	0	0	0.2168	0.5692	0	0.2120	0.0020
		179	0	0	0.2168	0.5692	0	0.1924	0.0216
		180	0	0	0.1013	0.5692	0	0.1900	0.1394
	DSSP3	177	0	0	0.2189	0.5328	0	0.2054	0.0429
		178	0	0	0.2407	0.5328	0	0.2240	0.0025
		179	0	0	0.2407	0.5328	0	0.1903	0.0363
		180	0	0	0.1166	0.5328	0	0.1970	0.1536
	AVE	177	0	0	0.1955	**0.5595**	0	0.1973	0.0478
		178	0	0	0.2188	**0.5595**	0	0.2189	0.0029
		179	0	0	0.2188	**0.5595**	0	0.1905	0.0313
		180	0	0	0.1032	**0.5595**	0	0.1933	0.1440
Q74F	DSSP1	177	0	0	0.1444	0.6749	0	0.1454	0.0353
		178	0	0	0.1612	0.6749	0	0.1627	0.0011
		179	0	0	0.1612	0.6749	0	0.1452	0.0186
		180	0	0	0.0631	0.6749	0	0.1558	0.1061
	DSSP2	177	0	0	0.1513	0.6501	0	0.1505	0.0481
		178	0	0	0.1755	0.6501	0	0.1728	0.0015
		179	0	0	0.1755	0.6501	0	0.1640	0.0103
		180	0	0	0.0773	0.6501	0	0.1741	0.0984
	DSSP3	177	0	0	0.1433	0.6546	0	0.1581	0.0440
		178	0	0	0.1662	0.6546	0	0.1776	0.0017
		179	0	0	0.1662	0.6546	0	0.1697	0.0096
		180	0	0	0.0764	0.6546	0	0.1642	0.1048
	AVE	177	0	0	0.1463	**0.6599**	0	0.1513	0.0425
		178	0	0	0.1676	**0.6599**	0	0.1710	0.0014
		179	0	0	0.1676	**0.6599**	0	0.1596	0.0128
		180	0	0	0.0723	**0.6599**	0	0.1647	0.1031

Table. S10 MM-GBSA in smTG and its mutants.

System	MM-GBSA(kcal/mol)			Ave.	Std.err.of
	MD1	MD2	MD3	(Sys)	mean
WT	-21.30	-19.30	-20.02	-20.21	0.83
H44C	-26.40	-26.90	-26.50	-26.60	0.22
Q74F	-21.14	-21.69	-20.96	-21.26	0.31
E87W	-17.50	-17.29	-16.08	-16.96	0.63

REFERENCES IN SUPPLEMENTARY INFORMATION

(1) Olivier, M.; Eeles, R.; Hollstein, M.; Khan, M. A.; Harris, C. C.; Hainaut, P. The IARC TP53 Database: New Online Mutation Analysis and Recommendations to Users. *Hum. Mutat.* **2002**, *19* (6), 607–614.

(2) Ordway, G. A.; Garry, D. J. Myoglobin: An Essential Hemoprotein in Striated Muscle. *J. Exp. Biol.* **2004**, *207* (20), 3441–3446.