Supplementary Materials

Efficient and low-cost error removal in DNA synthesis by a high-durability MutS

Jia Zhang[1,6,7], Yefei Wang[2], Baihui Chai[1,6,7], Jichao Wang[3,7], Lulu Li[4], Min Liu[1,6,7], Guang Zhao[3,7], Lishan Yao[2,7], Xiaolian Gao[4,5], Yifeng Yin[1], Jian Xu[1,6,7,*]

[1]Single-Cell Center, CAS Key Laboratory of Biofuels and Shandong Key Laboratory of Energy Genetics, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101, China

[2]CAS Key Laboratory of Biofuels and Shandong Key Laboratory of Synthetic Biology, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101, China

[3]CAS Key Laboratory of Biobased Materials, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101, China

[4]LC-BIO Technologies CO., LTD., Hangzhou, 310018, China

[5]Department of Biology and Biochemistry, University of Houston, Houston, TX 77004-5001, USA

[6]Laboratory of Marine Biology and Biotechnology, Qingdao National Laboratory of Marine Science and Technology, Qingdao, Shandong, China

[7]University of Chinese Academy of Sciences, Beijing, China

[*]Address correspondence to: xujian@qibebt.ac.cn

Contents

Supplementary Methods

The plasmids constructed in this study

All the plasmids used in this study are summarized in **Table S2**. Briefly, pMBP-eMutS which was constructed for MBP–eMutS–CBM fusion protein expression was constructed by ligation the eMutS–CBM fragment and the pMAL-c5X plasmid with the restriction sites of *BamH* I and *Not* I. pZJQIBEBT001 was constructed for the expression of eMutS-E38C-G70C gene which is eMutS gene's mutation. pZJQIBEBT002 was constructed for the expression of eMutS-S120C-S151C gene which is eMutS gene's mutation. pZJQIBEBT003 was constructed for the expression of eMutS-L157C-G233C gene which is eMutS gene's mutation. pZJQIBEBT004 was constructed for the expression of eMutS-E177C-G195C gene which is eMutS gene's mutation. pZJQIBEBT005 was constructed for the expression of eMutS-E451C-V465C gene which is eMutS gene's mutation. pZJQIBEBT006 was constructed for the expression of eMutS-Y474C-R500C gene which is eMutS gene's mutation. pZJQIBEBT007 was constructed for the expression of eMutS-T581C-K644C gene which is eMutS gene's mutation. pZJQIBEBT008 was constructed for the expression of eMutS-H585C-Q626C gene which is eMutS gene's mutation. pZJQIBEBT009 was constructed for the expression of eMutS-I597C-H760C gene which is eMutS gene's mutation. pZJQIBEBT010 was constructed for the expression of eMutS-M609C-T723C gene which is eMutS gene's mutation. pZJQIBEBT011 was constructed for the expression of eMutS-S120C-S151C-L157C-G233C gene which is eMutS gene's mutation. pZJQIBEBT012 was constructed for the expression of eMutS-S120C-S151C-L157C-G233C-M609C-T723C-E451C-V465C gene which is eMutS gene's mutation. pZJQIBEBT013 was constructed for the expression of MBP-eMutS-S120C-S151C-L157C-G233C gene which is eMutS gene's mutation with MBP fusion

protein. pZJQIBEBT014 was constructed for the expression of MBP-eMutS-S120C-S151C-L157C-G233C-M609C-T723C-E451C-V465C gene which is eMutS gene's mutation with MBP fusion protein. All the plasmids which have eMutS mutations were constructed with the method of site-directed mutagenesis just as the Fig S1 showed[1] (**Fig. S1**) and all the primers were list in **Table S1**.

The strains constructed in this study

The above plasmids were transformed into *Escherichia coli* BL21 by the heat shock method. The transformants were isolated on Luria-Bertani (LB) medium supplemented with ampicillin and validated by PCR. After the plasmids pMBP-eMutS, pZJQIBEBT001, pZJQIBEBT002, pZJQIBEBT003, pZJQIBEBT004, pZJQIBEBT005, pZJQIBEBT006, pZJQIBEBT007, pZJQIBEBT008, pZJQIBEBT009, pZJQIBEBT010, pZJQIBEBT011, pZJQIBEBT012, pZJQIBEBT013 and pZJQIBEBT014 were transformed into *Escherichia coli* BL21, respectively, to obtain eMutS expressing strains MBP-eMutS, eMutS-E38C-G70C, eMutS-S120C-S151C, eMutS-L157C-G233C, eMutS-E177C-G195C, eMutS-E451C-V465C, eMutS-Y474C-R500C, eMutS-T581C-K644C, eMutS-H585C-Q626C, eMutS-I597C-H760C, eMutS-M609C-T723C, Mutant2, Mutant4, MBP-Mutant2 and MBP-Mutant4 (**Table S3**).

Supplementary Tables and Figures

Table S1. The primers used in this study.

Primer names	Primer sequences
MlyI-1F	CACAGGAGTCCTCAC
MlyI-2F	CCAAGGAGTCGCTAG
MlyI-3F	CAGAGGAGTCCTGAG
MlyI-4F	CCTACGAGTCGCAAC
MlyI-5F	CATTCGAGTCGTCTG
MlyI-6F	CGTACGAGTCCCTTC
MlyI-7F	CTGTCGAGTCTGTAT
MlyI-8F	GGTACGAGTCCGATC
MlyI-9F	CTTACGAGTCGTCTC
MlyI-10F	GGATCGAGTCTCTAC
MlyI-11F	GTGTGGAGTCATATG
MlyI-12F	GGATCGAGTCAATTC
MlyI-13F	CGTTCGAGTCATCCA
MlyI-14F	GCATGGAGTCGGATG
xMlyI-1F	GTATAGAGTCAGTCG
xMlyI-2F	GTCACGAGTCATGGC
xMlyI-3F	GATAAGAGTCTCAGC
xMlyI-4F	GAGTGGAGTCTACCG
LMlyI-1F	CGATCGAGTCGCATC
LMlyI-2F	CACTGGAGTCCGTAC
LMlyI-3F	GCTAGGAGTCCATGC
LMlyI-4F	GCACTGAGTCGTCAG
2BbsI-1F	ACGCTCTGAAGACCC
2BbsI-2F	CGAGATAGAAGACAG
E38C-G70C-F1	GGATGGGTGATTTTTATTGCCTGTTTTATGACGACG
E38C-G70C-R1	CGTCGTCATAAAACAGGCAATAAAAATCACCCATCC
E38C-G70C-F2	GCCGATCCCGATGGCGTGCATTCCCTACCATGCGG
E38C-G70C-R2	CCGCATGGTAGGGAATGCACGCCATCGGGATCGGC
S120C-S151C-F1	TACGCCAGGCACCATCTGCGATGAAGCCCTGTTGC
S120C-S151C-R1	GCAACAGGGCTTCATCGCAGATGGTGCCTGGCGTA
S120C-S151C-F2	GCGACGCTGGATATCTGCTCCGGGCGTTTTCGC
S120C-S151C-R2	GCGAAAACGCCCGGAGCAGATATCCAGCGTCGC
L157C-G233C-F1	TTCCGGGCGTTTTCGCTGCAGCGAACCGGCTGACC
L157C-G233C-R1	GGTCAGCCGGTTCGCTGCAGCGAAAACGCCCGGAA
L157C-G233C-F2	CGAGAACGCGCCGCGCTGCCTTTGTGCTGCCGGTTG
L157C-G233C-R2	CAACCGGCAGCACAAAGGCAGCGCGGCGCGTTCTCG
E177C-G195C-F1	ACGCACTAATCCTGCGTGCCTGCTGTATGCAGAAG
E177C-G195C-R1	CTTCTGCATACAGCAGGCACGCAGGATTAGTGCGT
E177C-G195C-F2	ATTGAAGGCCGTCGCTGCCTGCGCCGTCGCCCG

E177C-G195C-R2	CGGGCGACGGCGCAGGCAGCGACGGCCTTCAAT
E451C-V465C-F1	TATCTGGAGCGTCTGTGCGTCCGCGAGCGTGAAC
E451C-V465C-R1	GTTCACGCTCGCGGACGCACAGACGCTCCAGATA
E451C-V465C-F2	CTGGACACGCTGAAATGCGGCTTTAATGCGGTG
E451C-V465C-R2	CACCGCATTAAAGCCGCATTTCAGCGTGTCCAG
Y474C-R500C-F1	GCGGTGCACGGCTACTGCATTCAAATCAGCCGT
Y474C-R500C-R1	ACGGCTGATTTGAATGCAGTAGCCGTGCACCGC
Y474C-R500C-F2	CTGAAAAACGCCGAGTGCTACATCATTCCAGAG
Y474C-R500C-R2	CTCTGGAATGATGTAGCACTCGGCGTTTTTCAG
T581C-K644C-F1	CCGGGCATTCGCATTTGCGAAGGTCGCCATCCG
T581C-K644C-R1	CGGATGGCGACCTTCGCAAATGCGAATGCCCGG
T581C-K644C-F2	TATGTACCGGCACAATGCGTCGAGATTGGACCT
T581C-K644C-R2	AGGTCCAATCTCGACGCATTGTGCCGGTACATA
H585C-Q626C-F1	ATTACCGAAGGTCGCTGCCCGGTAGTTGAACAA
H585C-Q626C-R1	TTGTTCAACTACCGGGCAGCGACCTTCGGTAAT
H585C-Q626C-F2	AGTACCTATATGCGCTGCACCGCACTGATTGCG
H585C-Q626C-R2	CGCAATCAGTGCGGTGCAGCGCATATAGGTACT
I597C-H760C-F1	CTGAATGAGCCATTTTGCGCCAACCCGCTGAAT
I597C-H760C-R1	ATTCAGCGGGTTGGCGCAAAATGGCTCATTCAG
I597C-H760C-F2	ACCATTGCCTTTATGTGCAGCGTGCAGGATGGC
I597C-H760C-R2	GCCATCCTGCACGCTGCACATAAAGGCAATGGT
M609C-T723C-F1	TCGCCGCAGCGCCGCTGCTTGATCATCACCGGTC
M609C-T723C-R1	GACCGGTGATGATCAAGCAGCGGCGCTGCGGCGA
M609C-T723C-F2	TAAGATTAAGGCATTGTGCTTATTTGCTACCCACTA
M609C-T723C-R2	TAGTGGGTAGCAAATAAGCACAATGCCTTAATCTTA

Table S2. The plasmids used in this study.

Plasmids	Selection marker and description	Reference
peMutS-CBM3-EGFP	AMP, eMutS	(Wan et al., 2014)
pMBP-eMutS-CBM3	AMP, MBP-eMutS-CBM3	This study
pZJQIBEBT001	AMP, eMutS-E38C-G70C-CBM3-EGFP	This study
pZJQIBEBT002	AMP, eMutS-S120C-S151C-CBM3-EGFP	This study
pZJQIBEBT003	AMP, eMutS-L157C-G233C-CBM3-EGFP	This study
pZJQIBEBT004	AMP, eMutS-E177C-G195C-CBM3-EGFP	This study
pZJQIBEBT005	AMP, eMutS-E451C-V465C-CBM3-EGFP	This study
pZJQIBEBT006	AMP, eMutS-Y474C-R500C-CBM3-EGFP	This study
pZJQIBEBT007	AMP, eMutS-T581C-K644C-CBM3-EGFP	This study
pZJQIBEBT008	AMP, eMutS-H585C-Q626C-CBM3-EGFP	This study
pZJQIBEBT009	AMP, eMutS-I597C-H760C-CBM3-EGFP	This study
pZJQIBEBT010	AMP, eMutS-M609C-T723C-CBM3-EGFP	This study
pZJQIBEBT011	AMP, eMutS-S120C-S151C-L157C-G233C-CBM3-EGFP	This study
pZJQIBEBT012	AMP, eMutS-S120C-S151C-L157C-G233C-M609C-T723C-E451C-V465C-CBM3-EGFP	This study
pZJQIBEBT013	AMP, MBP-eMutS-S120C-S151C-L157C-G233C-CBM3	This study
pZJQIBEBT014	AMP, MBP-eMutS-S120C-S151C-L157C-G233C-M609C-T723C-E451C-V465C-CBM3	This study

Table S3. The strains constructed in this study.

Strains	Selection marker and plasmids	Competent cell
MBP-eMutS	AMP, pMBP-eMutS-CBM3	*Escherichia coli* BL21
eMutS-E38C-G70C	AMP, pZJQIBEBT001	*Escherichia coli* BL21
eMutS-S120C-S151C	AMP, pZJQIBEBT002	*Escherichia coli* BL21
eMutS-L157C-G233C	AMP, pZJQIBEBT003	*Escherichia coli* BL21
eMutS-E177C-G195C	AMP, pZJQIBEBT004	*Escherichia coli* BL21
eMutS-E451C-V465C	AMP, pZJQIBEBT005	*Escherichia coli* BL21
eMutS-Y474C-R500C	AMP, pZJQIBEBT006	*Escherichia coli* BL21
eMutS-T581C-K644C	AMP, pZJQIBEBT007	*Escherichia coli* BL21
eMutS-H585C-Q626C	AMP, pZJQIBEBT008	*Escherichia coli* BL21
eMutS-I597C-H760C	AMP, pZJQIBEBT009	*Escherichia coli* BL21
eMutS-M609C-T723C	AMP, pZJQIBEBT010	*Escherichia coli* BL21
Mutant2	AMP, pZJQIBEBT011	*Escherichia coli* BL21
Mutant4	AMP, pZJQIBEBT012	*Escherichia coli* BL21
MBP-Mutant2	AMP, pZJQIBEBT013	*Escherichia coli* BL21
MBP-Mutant4	AMP, pZJQIBEBT014	*Escherichia coli* BL21

Table S4. The disulfide bonds constructed in this study.

Plasmid	Amino Acid Mutation	Gene Mutation
pZJQIBEBT001	E38C-G70C	GAA114TGC – GGG210TGC
pZJQIBEBT002	S120C-S151C	AGC360TGC – AGT453TGC
pZJQIBEBT003	L157C-G233C	CTG471TGC – GGA699TGC
pZJQIBEBT004	E177C-G195C	GAA531TGC – GGC585TGC
pZJQIBEBT005	E451C-V465C	GAA1353TGC – GTT1395TGC
pZJQIBEBT006	Y474C-R500C	TAC1422TGC – CGC1500TGC
pZJQIBEBT007	T581C-K644C	ACC1743TGC – AAA1932TGC
pZJQIBEBT008	H585C-Q626C	CAT1755TGC – CAG1878TGC
pZJQIBEBT009	I597C-H760C	ATC1791TGC – CAC2280TGC
pZJQIBEBT010	M609C-T723C	ATG1827TGC – ACG2169TGC
pZJQIBEBT011	S120C-S151C, L157C-S151C	AGC360TGC – AGT453TGC, CTG471TGC – GGA699TGC
pZJQIBEBT012	S120C-S151C, L157C-S151C, M609C-T723C, E451C-V465C	AGC360TGC – AGT453TGC, CTG471TGC – GGA699TGC, ATG1827TGC – ACG2169TGC, GAA1353TGC – GTT1395TGC
pZJQIBEBT013	S120C-S151C, L157C-S151C	AGC360TGC – AGT453TGC, CTG471TGC – GGA699TGC
pZJQIBEBT014	S120C-S151C, L157C-S151C, M609C-T723C, E451C-V465C	AGC360TGC – AGT453TGC, CTG471TGC – GGA699TGC, ATG1827TGC – ACG2169TGC, GAA1353TGC – GTT1395TGC

Table S5. List of oligos synthesized for 59-bp heteroduplex or homoduplex DNA construction for assaying the enzymatic activity of proteins.

Oligos names	Oligos sequences
mutsgene-F	GCGGACTATTTAACACAGCTTTAGGCGCTGGACGAGGTACTATGAATCGGCCTTGCTCC
mutsgene-R	GGAGCAAGGCCGATTCATAGTACCTCGTCCAGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-QS	GGAGCAAGGCCGATTCATAGTACCTCGTCAGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-CR	GGAGCAAGGCCGATTCATAGTACCTCGTCCTAGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-CT	GGAGCAAGGCCGATTCATAGTACCTCGTCTAGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-CA	GGAGCAAGGCCGATTCATAGTACCTCGTCAAGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-CG	GGAGCAAGGCCGATTCATAGTACCTCGTCGAGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-GT	GGAGCAAGGCCGATTCATAGTACCTCGTCCATCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-GA	GGAGCAAGGCCGATTCATAGTACCTCGTCCAACGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-GC	GGAGCAAGGCCGATTCATAGTACCTCGTCCACCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-AC	GGAGCAAGGCCGATTCATAGTACCTCGTCCCGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-AG	GGAGCAAGGCCGATTCATAGTACCTCGTCCGGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-AT	GGAGCAAGGCCGATTCATAGTACCTCGTCCTGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-TC	GGAGCAAGGCCGATTCATAGTACCTCGCCCAGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-TG	GGAGCAAGGCCGATTCATAGTACCTCGGCCAGCGCCTAAAGCTGTGTTAAATAGTCCGC
mutsgene-R-TH-TA	GGAGCAAGGCCGATTCATAGTACCTCGACCAGCGCCTAAAGCTGTGTTAAATAGTCCGC

Table S6. List of synthesized double-stranded DNA samples for assaying the enzymatic activity of proteins.

Double-stranded DNA name	Single-stranded DNA name
Perfect match	mutsgene-F and mutsgene-R
Deletion	mutsgene-F and mutsgene-R-QS
Insertion	mutsgene-F and mutsgene-R-CR
Substitution-CT	mutsgene-F and mutsgene-R-TH-CT
Substitution-CA	mutsgene-F and mutsgene-R-TH-CA
Substitution-CG	mutsgene-F and mutsgene-R-TH-CG
Substitution-GT	mutsgene-F and mutsgene-R-TH-GT
Substitution-GA	mutsgene-F and mutsgene-R-TH-GA
Substitution-GC	mutsgene-F and mutsgene-R-TH-GC
Substitution-AC	mutsgene-F and mutsgene-R-TH-AC
Substitution-AG	mutsgene-F and mutsgene-R-TH-AG
Substitution-AT	mutsgene-F and mutsgene-R-TH-AT
Substitution-TC	mutsgene-F and mutsgene-R-TH-TC
Substitution-TG	mutsgene-F and mutsgene-R-TH-TG
Substitution-TA	mutsgene-F and mutsgene-R-TH-TA

Single-stranded DNA were synthesized from company. Double-stranded DNA were formed by annealing from the single-stranded DNA.

Table S7. The cost to prepare an iMICC system.

Compose	Amount	iMICC Cost
MutS protein	1.2 pmol	$0.013
Empty column	1	$0.261
Cellulose slurry	500μL	$0.100
Total		$0.374

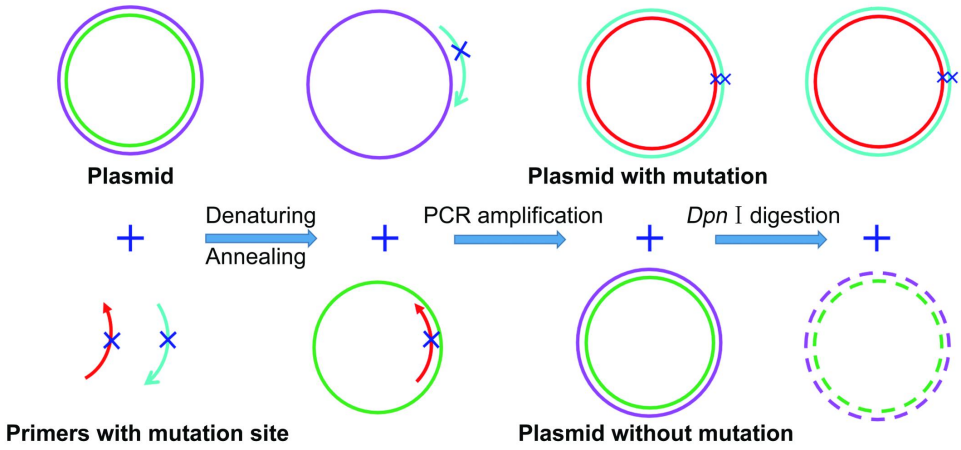

Figure S1. Flow chart for site-directed mutagenesis[1]**.** In order to generate PCR products for plasmids with mutation sites, we used plasmids as template and primers with mutation site, numerous **reco**mbinant plasmids containing mutation sites and a small amount of non-mutate plasmids were obtained. After amplification, *Dpn* I was used to specifically digest the non-mutate plasmids, the high purity products plasmids with mutation sites could be obtained.

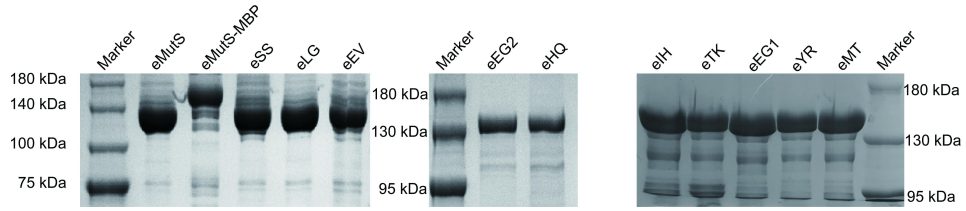

Figure S2. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis result of all kinds of purified proteins. All the purified proteins were loaded with 15 μl on 8% SDS-PAGE, stained with Coomassie Brilliant Blue R250. The molecular weight of all the eMutS mutant proteins is 147.8 kDa. The molecular weight of the MBP-eMutS fusion proteins is 165.3 kDa. Protein Marker (PageRulerTM Prestained Protein Ladder, #c510010 and #c620014 from Fermentas). eEG1, eMutS-E38C-G70C. eSS, eMutS-S120C-S151C. eLG, eMutS-L157C-G233C. eEV, eMutS-E451C-V465C. eTK, eMutS-T581C-K644C. eMT, eMutS-M609C-T723C. eIH, eMutS-I597C-H760C. eEG2, eMutS-E177C-G195C. eYR, eMutS-Y474C-R500C. eHQ, eMutS-H585C-Q626C.

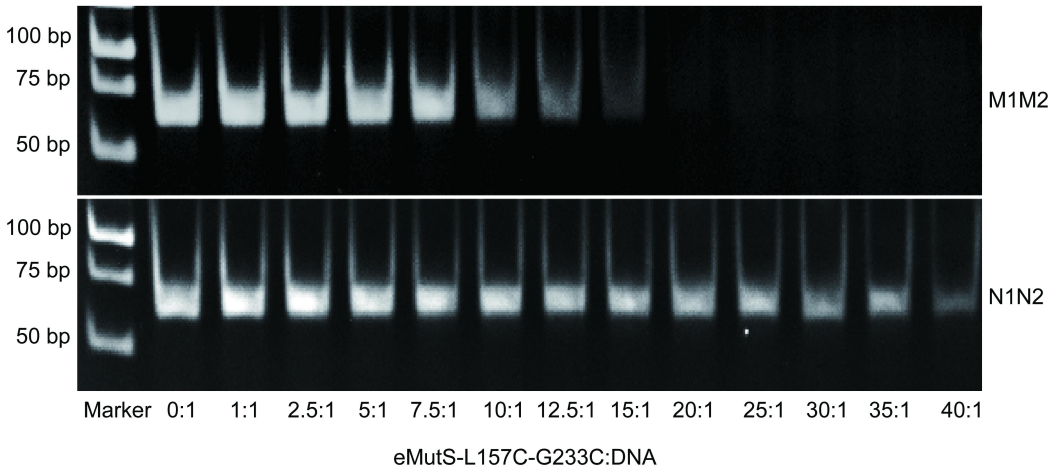

Figure S3. Band-shift assay of the eMutS-L157C-G233C protein bound to heteroduplexes containing a single-bp mismatch and perfectly matched homoduplexes. The 59 bp homoduplex (N1N2) and the 59 bp heteroduplexes (M1M2) were incubated with eMutS-L157C-G233C at various molar ratios (eMutS-L157C-G233C:duplexes from 0:1 to 40:1) for 10 min at 25 °C.

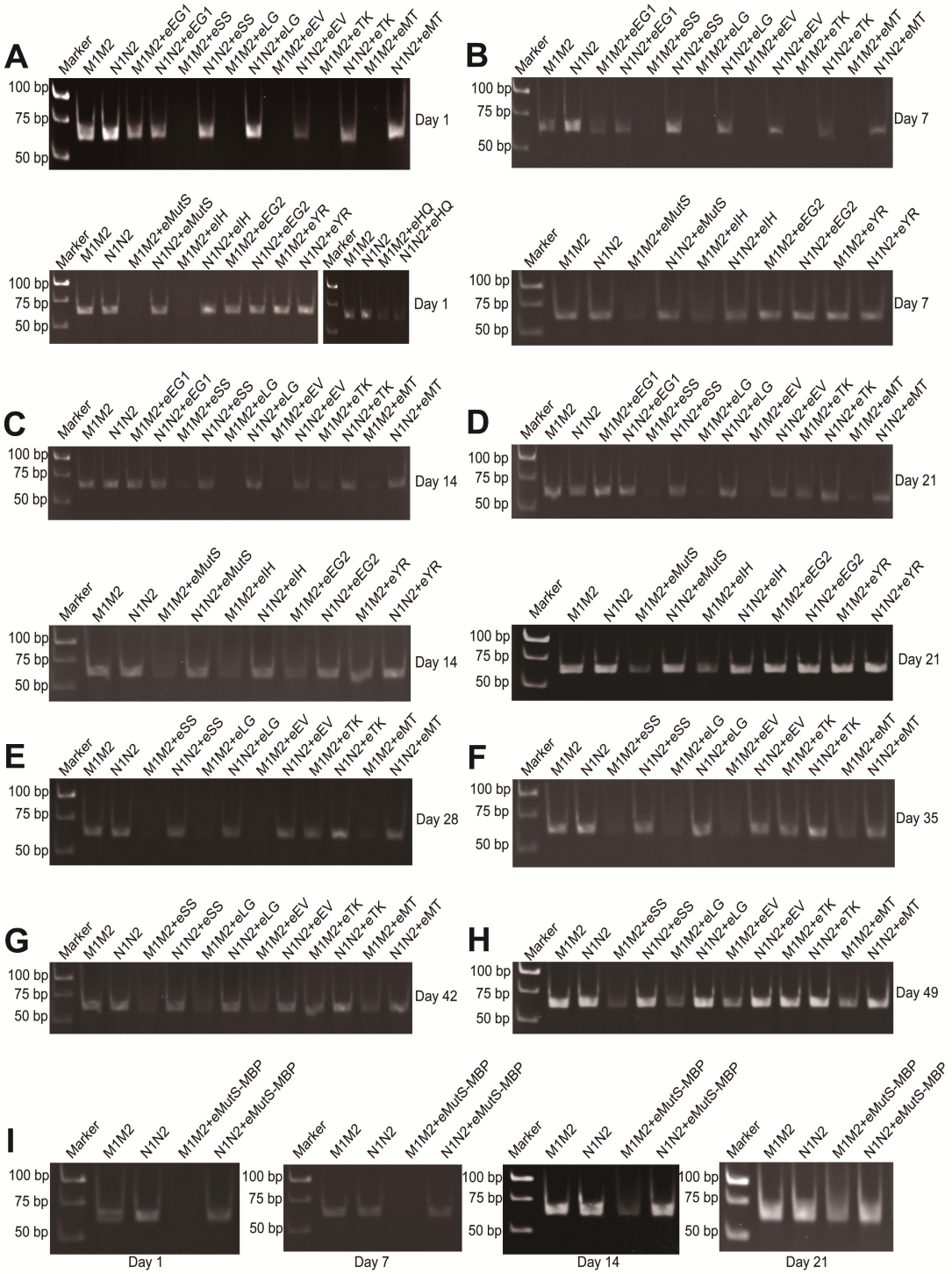

Figure S4. Band-shift assay to measure the binding of different proteins to heteroduplexes containing a single-bp mismatch and perfectly matched homoduplexes. The 59 bp homoduplexes (N1N2) and the 59 bp heteroduplexes (M1M2) were incubated with different proteins at the molar ratios (proteins:duplexes 20:1) for 10 min at 25 °C. This band-shift assay was performed to analyze the nonspecific binding properties of different proteins to mismatched DNA

and perfect DNA. The band-shift assay results of all proteins on the 1st (**A**), 7th (**B**), 14th (**C**), 21st (**D**), 28th (**E**), 35th (**F**), 42nd (**G**) and 49th (**H**) day were shown. (**I**) The band-shift assay results of MBP–eMutS fusion protein from the 1st day to the 21st day were shown. eEG1, eMutS-E38C-G70C. eSS, eMutS-S120C-S151C. eLG, eMutS-L157C-G233C. eEV, eMutS-E451C-V465C. eTK, eMutS-T581C-K644C. eMT, eMutS-M609C-T723C. eIH, eMutS-I597C-H760C. eEG2, eMutS-E177C-G195C. eYR, eMutS-Y474C-R500C. eHQ, eMutS-H585C-Q626C.

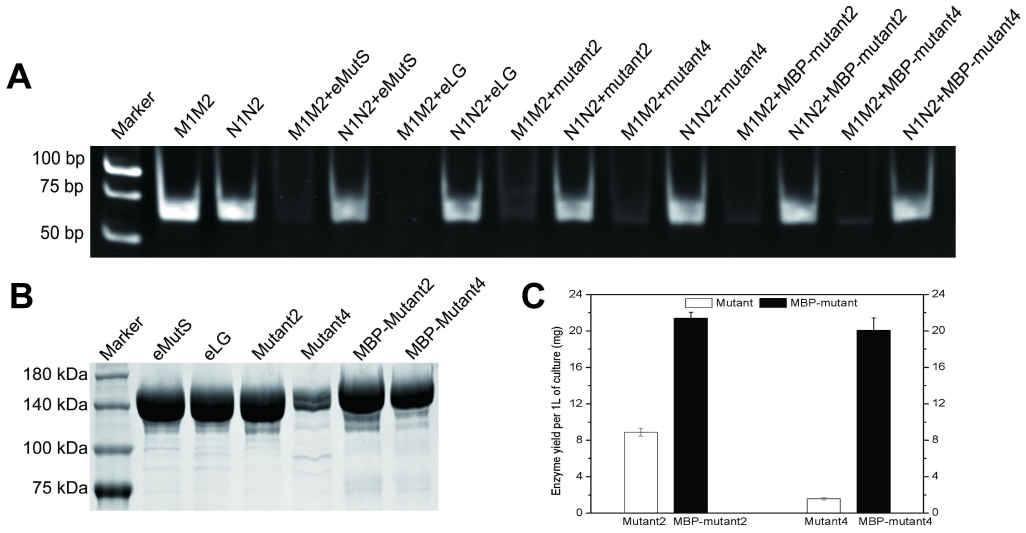

Figure S5. Combination two or four pairs of disulfide bonds to explore whether these stability improvement effect could be additive. (**A**) Band-shift assay to measure the error correction ability of proteins with multiple pairs of disulfide bonds. The 59 bp homoduplexes (N1N2) and 59 bp heteroduplexes (M1M2) were incubated with different proteins which were stored at 30 °C for 4 h at the molar ratios (proteins:duplexes 20:1) for 10 min at 25 °C. eLG, eMutS-L157C-G233C. (**B**) Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis result of all kinds of purified proteins on 8% SDS-PAGE. Protein Marker (PageRulerTM Prestained Protein Ladder, #c620014 from Fermentas). (**C**) The expression of different strains. Mutant2, eMutS-S120C-S151C-L157C-S151C. Mutant4, eMutS-S120C-S151C-L157C-S151C-E451C-V465C-M609C-T723C.

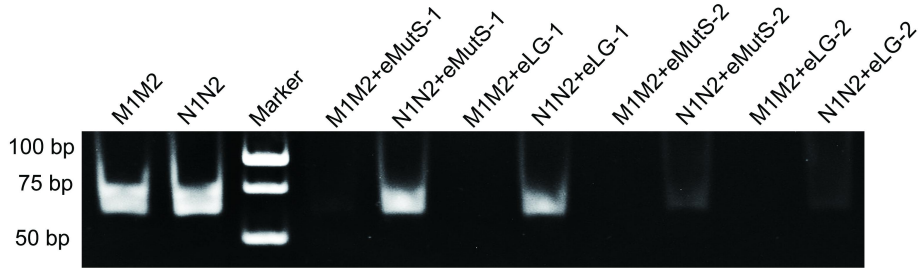

Figure S6. Evaluation of the error-removal ability of different proteins in different storage conditions. 1, Proteins were stored in liquid at 4 ℃. 2, Proteins were stored in glycerin at -20 ℃. eLG, eMutS-L157C-G233C.

Supplementary References

(1) Xia, Y.; Xun, L. Revised Mechanism and Improved Efficiency of the Quik Change Site-Directed Mutagenesis Method. *Methods Mol. Biol*. 2017, 25, 367–374.