Supporting Information

Improved Bst DNA polymerase variants derived via a machine-learning approach

Inyup Paik[1,2], Phuoc H. T. Ngo[1,2,3], Raghav Shroff[1,2,4], Daniel J. Diaz[2,3], Andre C. Maranhao[1,2], David J.F. Walker[1,2], Sanchita Bhadra[1,2] and Andrew D. Ellington[1,2,]*

1 Department of Molecular Biosciences, College of Natural Sciences, The University of Texas at Austin, Austin, TX 78712, USA

2 Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, TX 78712, USA

3 Department of Chemistry, College of Natural Sciences, The University of Texas at Austin, Austin, TX 78712, USA

4 CCDC Army Research Lab-South, Austin, TX 78712, USA

* To whom correspondence should be addressed. Tel: +1-512-232-3424, +1-512-471-6445; Fax: +1-512-471-7014; Email: ellingtonlab@gmail.com

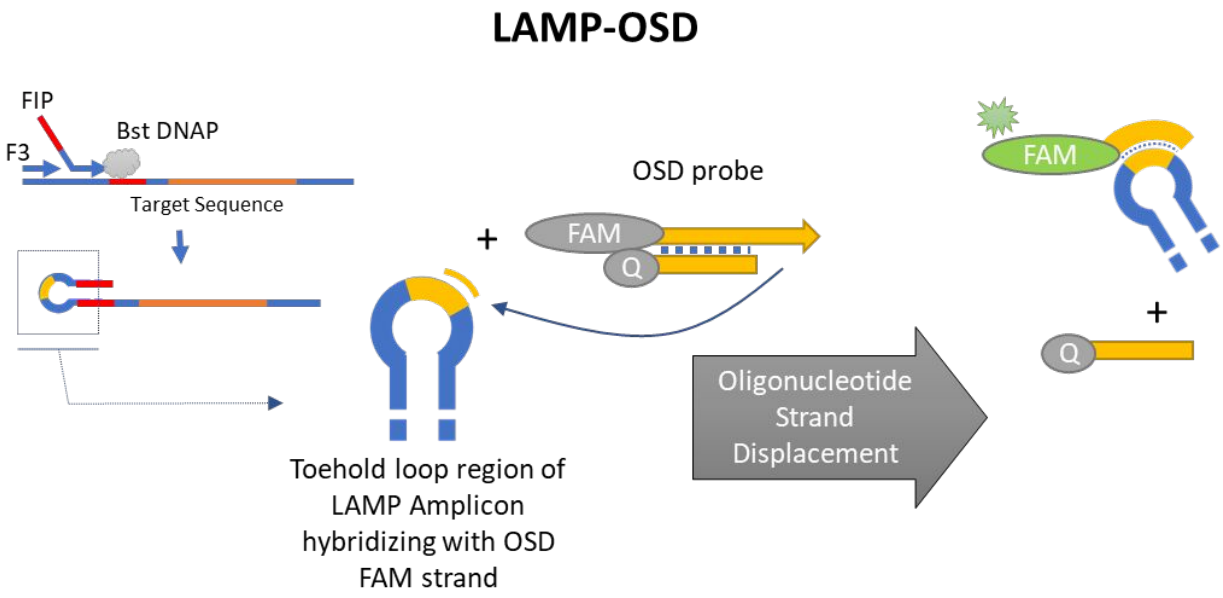

Figure S1. Schematic diagram of LAMP-OSD (Oligonucleotide Strand Displacement) FAM and Q represent 6-Carboxyfluorescein (6-FAM) fluorophore and quencher, respectively.

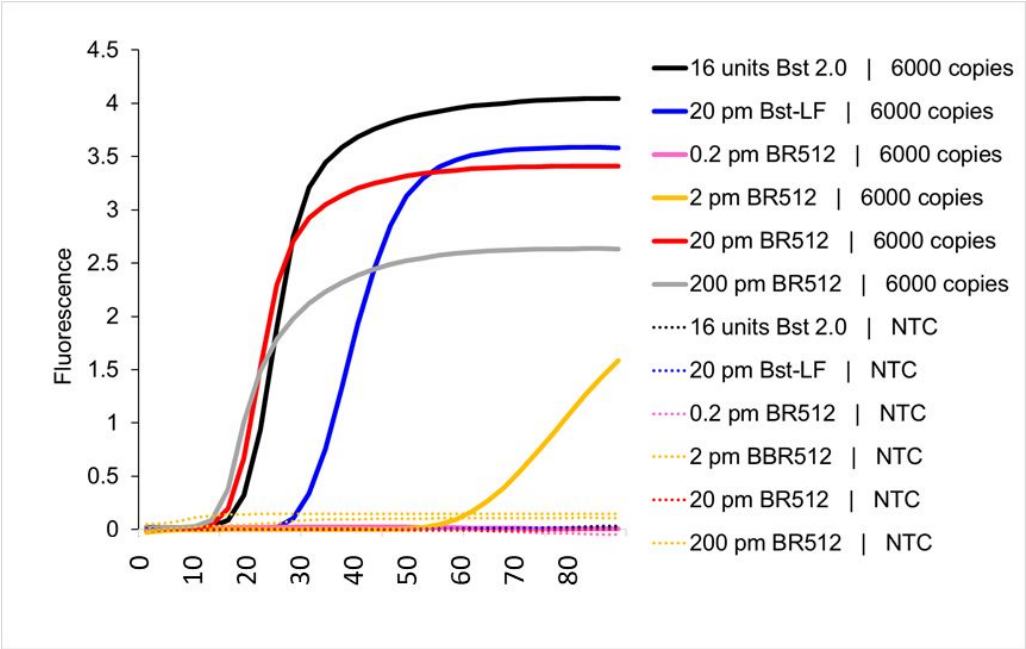

Figure S2. Effect of varying amounts of Br512 on LAMP-OSD of DNA templates. Indicated amounts of Br512 were compared with indicated amounts of in-house purified Bst-LF and commercially sourced Bst 2.0 in human *GAPDH* gene-specific LAMP-OSD assays operated in 1X isothermal buffer (NEB). Reactions were seeded with either 6000 copies of *GAPDH* plasmid template or with no specific templates (NTC). Amplification curves generated by real-time measurement of OSD fluorescence at 65 ˚C are depicted.

.

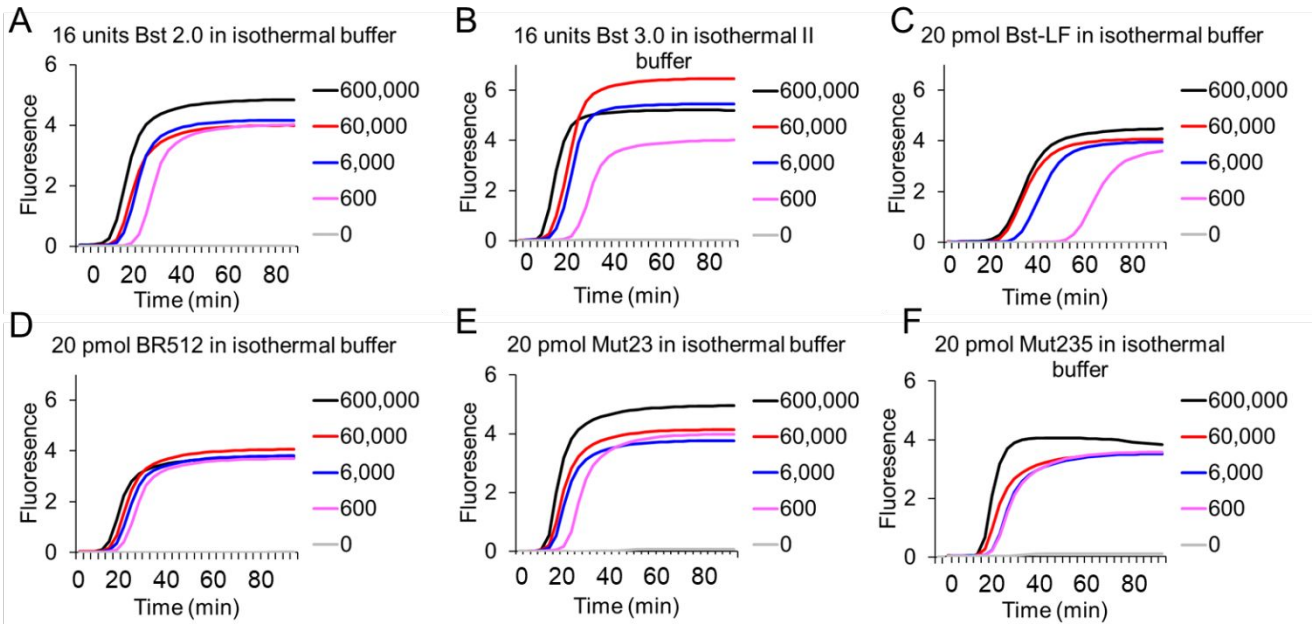
A
16 units Bst 2.0 in isothermal buffer
B
16 units Bst 3.0 in isothermal II buffer
C
20 pmol Bst-LF in isothermal buffer
D
20 pmol BR512 in isothermal buffer
E
20 pmol Mut23 in isothermal buffer
F
20 pmol Mut235 in isothermal buffer
Fluoresence
Time (min)
0 20 40 60 80
0 2 4 6
600,000
60,000
6,000
600
0

Figure S3. Comparison of Br512, Mut23, Mut235, Bst-LF, Bst2.0, and Bst3.0 in LAMP-OSD assays of DNA templates. LAMP-OSD assays for the human *gapd* gene were carried out with 16 units of commercially sourced Bst 2.0 (panel A), 16 units of commercially sourced Bst 3.0 (panel B), 20 pm of in-house purified Bst-LF (panel C), 20 pm of in-house purified Br512 (panel D), 20 pm of in-house purified Mut23 (panel E), or with 20 pm of in-house purified Mut235 (panel F) in the indicated reaction buffer. Amplification curves were observed in real-time at 65 ˚C by measuring OSD fluorescence in reactions seeded with 600,000 (black traces), 60,000 (red traces), 6,000 (blue traces), 600 (pink traces), and 0 (gray traces) copies of *GAPDH* plasmid templates.

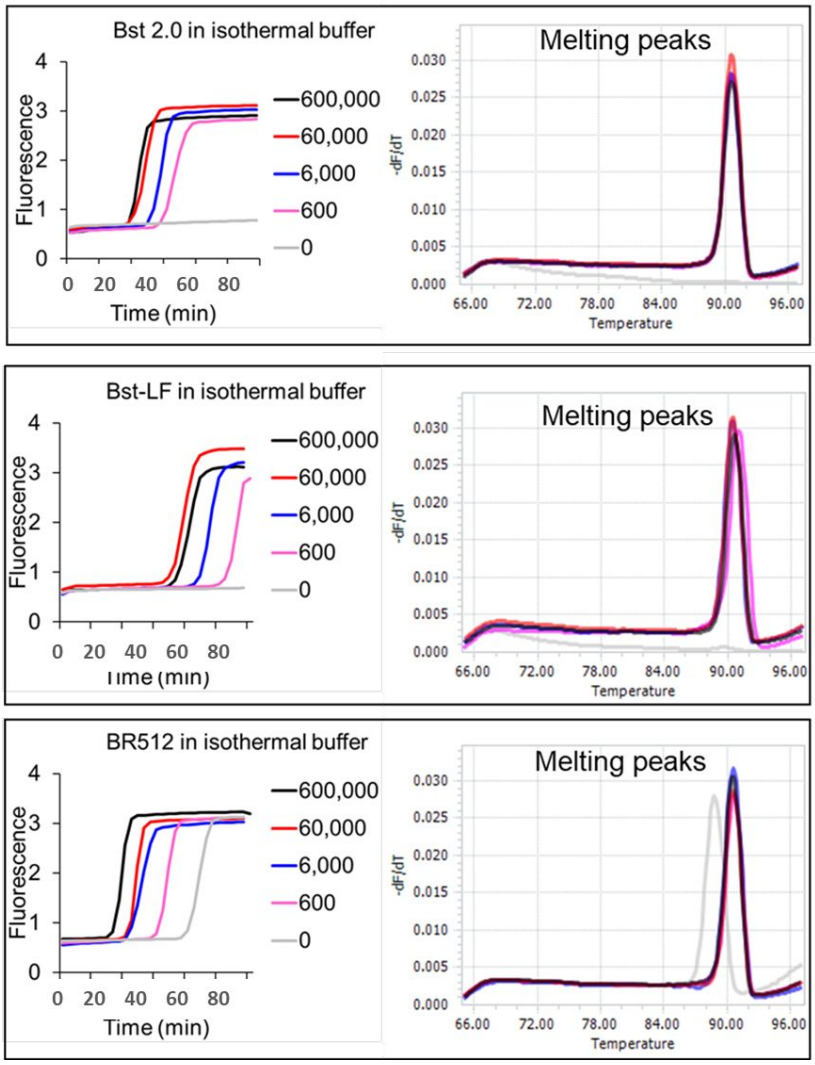

Figure S4. Comparison of Br512, Bst-LF, and Bst 2.0 in LAMP assays of DNA templates read using EvaGreen intercalating dye. LAMP assays for human *GAPDH* gene were operated using Bst 2.0, Bst-LF, or BR512 in the indicated reaction buffer. Amplification curves observed in real-time at 65 ˚C by measuring EvaGreen fluorescence in reactions seeded with 600,000 (black traces), 60,000 (red traces), 6,000 (blue traces), 600 (pink traces), and 0 (gray traces) copies of *GAPDH*

plasmid templates are depicted. LAMP amplicons were analyzed using the 'melt curve analysis' on LightCycler 96 real-time PCR machine and resulting melting peaks are indicated in the corresponding colored traces.

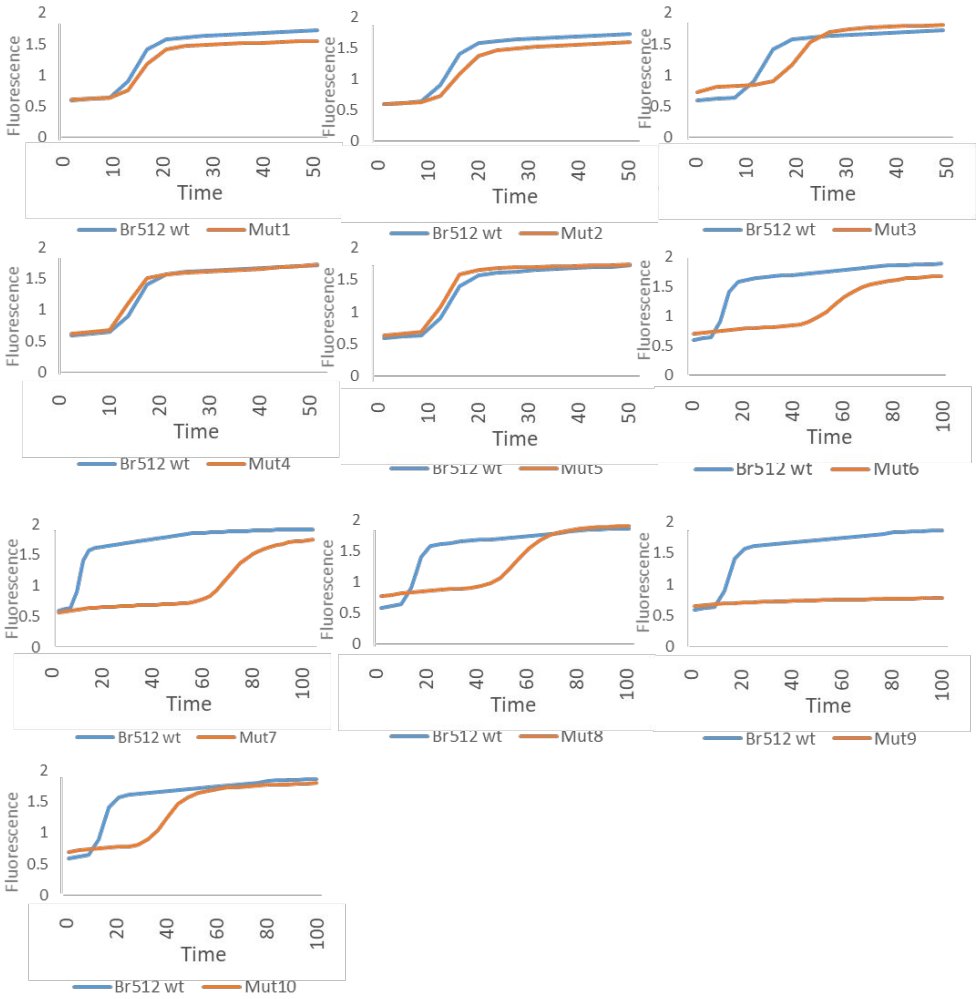

Figure S5. Initial evaluation of computationally predicted substitutions on Br512 (Bst-LF) activity. LAMP assays were carried out with a 20 pg ($6x10^7$ copies) of *GAPDH* DNA template to

assess the effect of the individual mutations suggested by Mutcompute on Br512 activity. Amplification was observed by EvaGreen dye fluorescence change (Y-axis) over time of incubation in minutes (X-axis) at 65 °C. Blue traces indicates Br512 wild type and burnt orange traces are individual mutations (Mut1-Mut10).

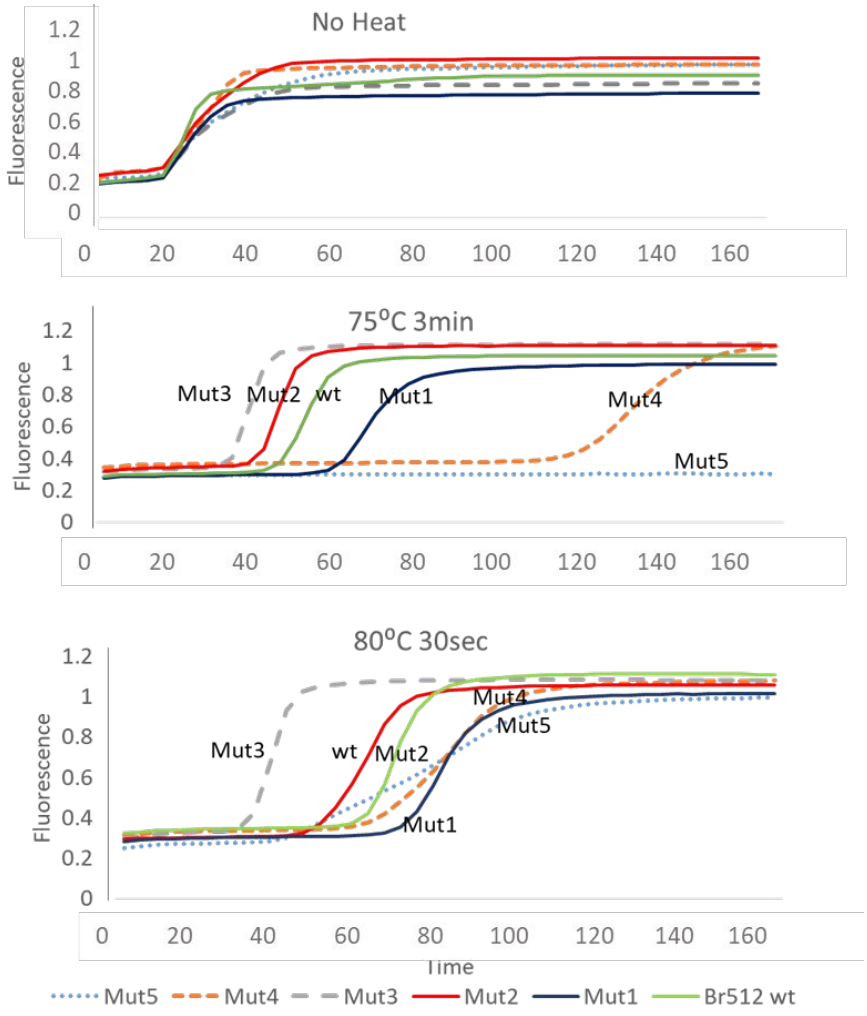

Figure S6. Heat challenge LAMP assay with computationally predicted single amino acid substitutions. LAMP assays assembled with wildtype Br512 (wt) or Mutcompute calculated Br512 variants (Mut1 to 5) were subjected to indicated thermal challenges (top panel: no thermal challenge; middle panel: 3 min at 75 ° C; lower panel: 30 sec at 80 ° C) prior to real time measurement of DNA amplification during continuous incubation at 65°C. Amplification kinetics was determined by measuring EvaGreen fluorescence (Y-axis) over incubation time in minutes (X-axis; hh:mm:ss). Green: Br512 wt (wild type), Dark blue: Mut1, Red: Mut2, Dotted gray: Mut3, Dotted orange: Mut4, Dotted blue: Mut5

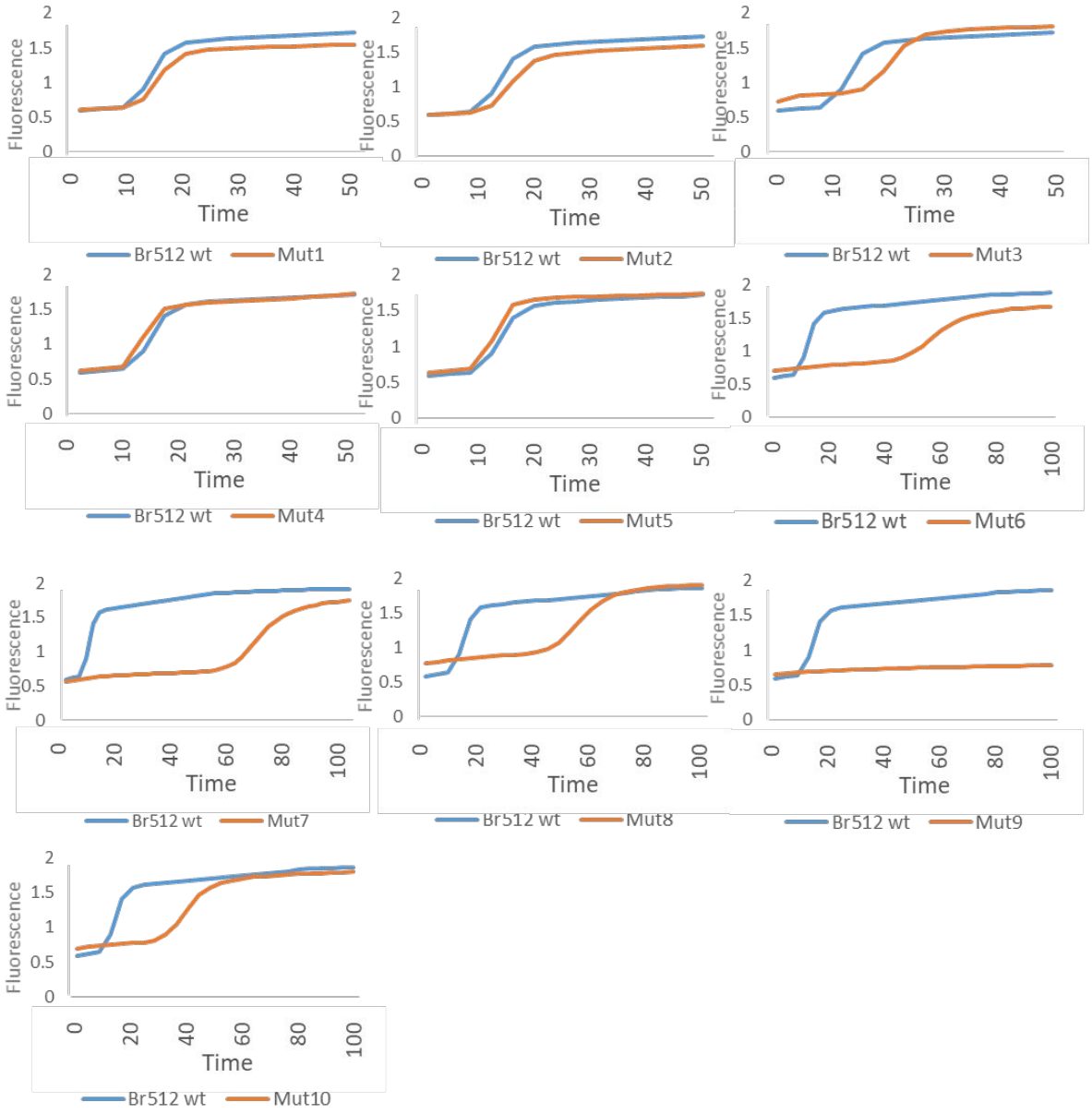

Figure S7. Heat challenge LAMP assay with double mutation Br512 variants.

Activities of wild type (blue traces) and the various double mutant Mutcompute Br512 variants (orange traces) were compared in identical LAMP assays containing 20 pg ($6x10^7$ copies) of

GAPDH DNA templates that were subjected to indicated thermal challenges (top panel: no thermal challenge; middle panel: 3 min at 75 °C; lower panel: 30 sec at 80 °C) prior to real time measurement of DNA amplification at 65°C. Representative amplification curves determined by measuring EvaGreen fluorescence (Y-axis) over incubation time (X-axis; minutes) are depicted.

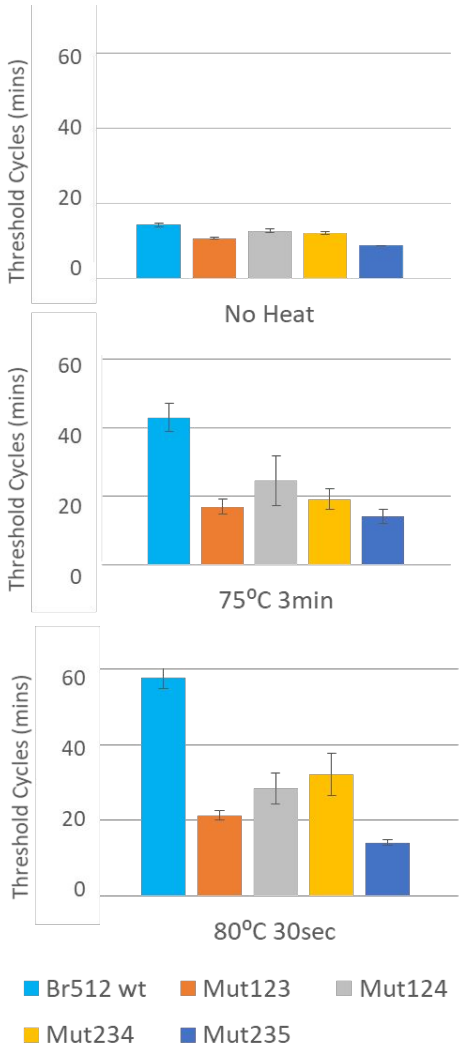
Threshold Cycles (mins)
60
40
20
0
No Heat
Threshold Cycles (mins)
60
40
20
0
75°C 3min
Threshold Cycles (mins)
60
40
20
0
80°C 30sec
Br512 wt
Mut123
Mut124
Mut234
Mut235

Figure S8. Threshold cycle (Ct) analysis of triple Mutcompute variants. *GAPDH* LAMP assay results shown in Figure 3 b-d were further quantified with Ct values in minutes. Threshold cycles for amplification of 20 pg ($6x10^7$ copies) *GAPDH* DNA templates were calculated using the Lightcycler96 software (Roche). Lower Ct indicates faster amplification. Upper panel: Ct values for No Heat LAMP, Middle panel: Ct values for 75^0C 3min heat challenge LAMP, Lower panel: Ct values for 80^0C 30sec heat challenge LAMP. , Error bar=S.D., n=2 for No Heat, n=3 for 75^0C and 80^0C heat challenge LAMP (Y-axis: Ct in minutes).

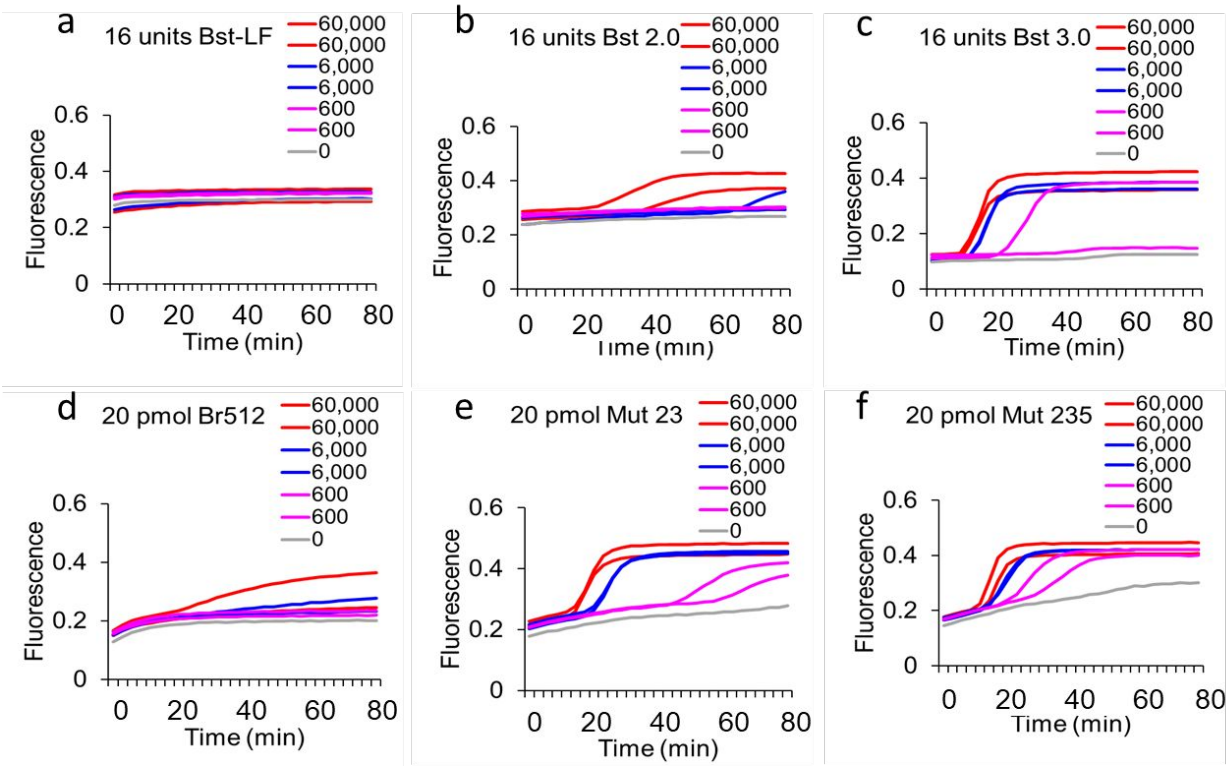

Figure S9. Comparison of Br512, Mut23, Mut235, Bst-LF, Bst2.0, and Bst3.0 in LAMP-OSD assays executed at 73 °C. LAMP-OSD assays for the human *GAPDH* gene were carried out with indicated amounts of commercially sourced Bst-LF, Bst 2.0, and Bst 3.0 and in-house purified Br512, Mut23, and Mut235. Amplification curves were observed in real-time at 73 °C by measuring OSD fluorescence in reactions seeded with 60,000 (red traces), 6,000 (blue traces), 600 (pink traces), and 0 (gray traces) copies of *GAPDH* plasmid templates.

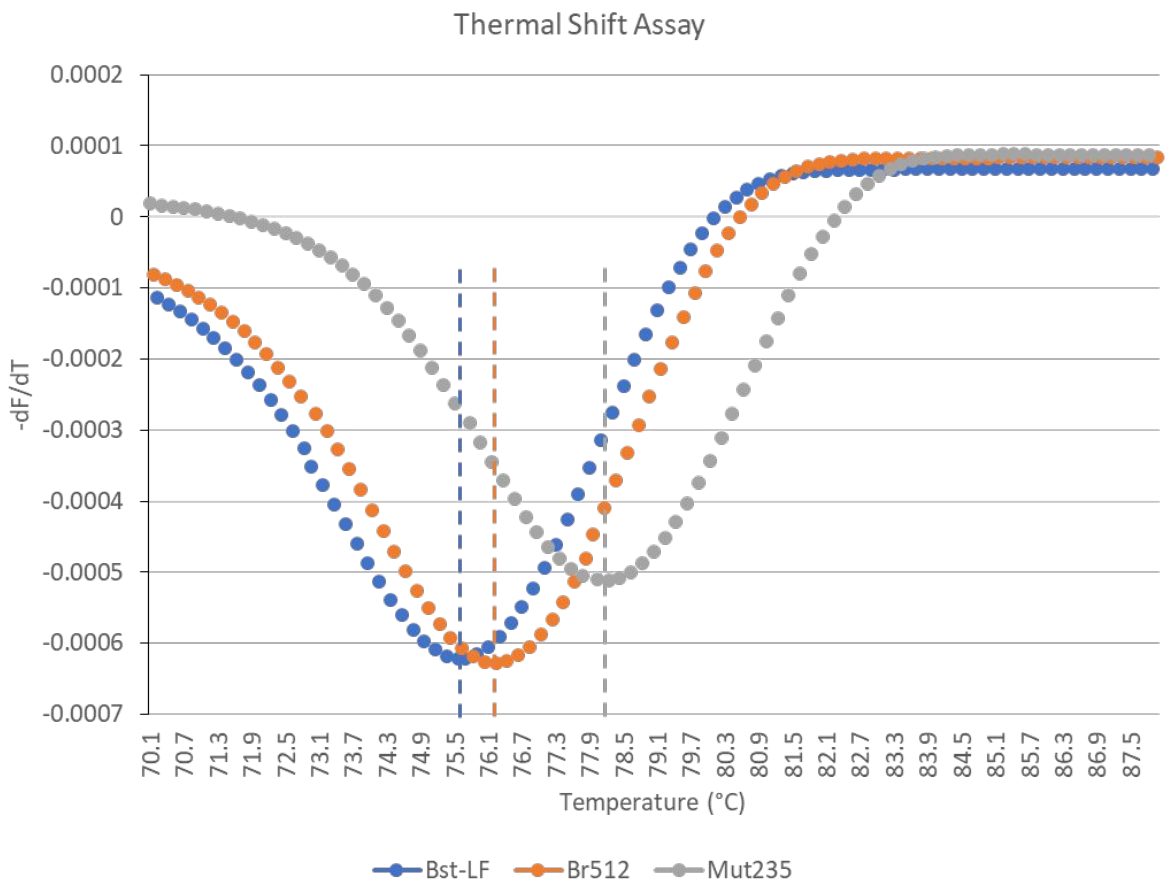

Figure S10. Protein Thermal Shift Assay for engineered Bst-LF variants.

Thermal stability of the parental (Bst-LF) and engineered enzyme variants were analyzed using Protein Thermal Shift™ (Thermo Fisher; Catalog Number: 4461146), a dye-based protein thermal shift assay, according to the manufacturer's instructions. The enzymes (5µg) were incubated in a Lightcyler 96 (Roche) real-time PCR machine programmed to ramp temperature from 37 ˚C to 95 ˚C at the rate of 0.1 ˚C/sec while continuously measuring changes in red fluorescence. Melt curves

generated by plotting change in fluorescence (dF) as a function of changing temperature (dT) are depicted.

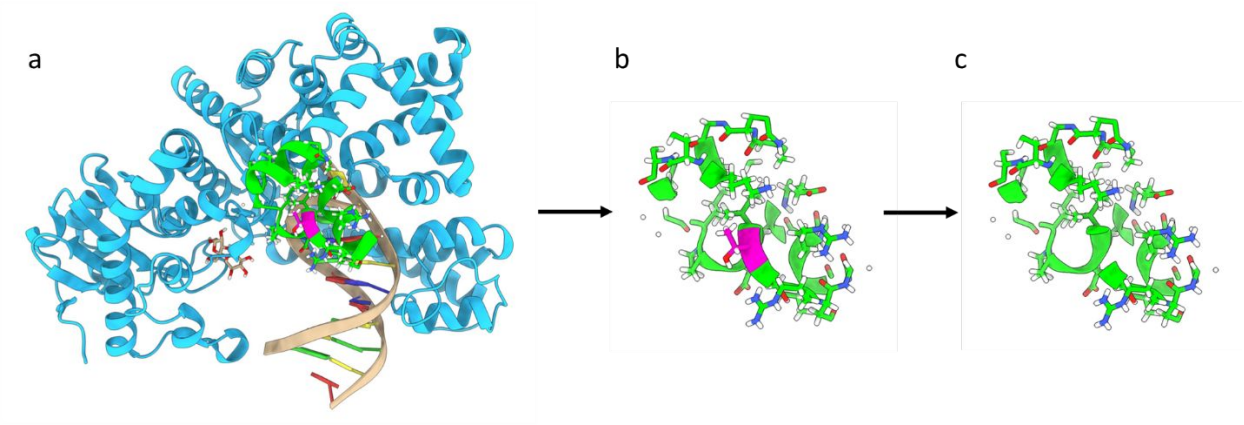

Figure S11. Generating a microenvironment and a label for self-supervised learning. (a) Select a focal residue (pink; T493; Mut2) and filter all atoms within a 10-angstrom cube of the alpha carbon (green). The cube orientation is determined by normalizing to the protein backbone and aligning the side chain with the +z axis. (b) Delete remaining protein atoms (blue). (c) Mask the focal residue (pink; T493) by deleting it to generate the microenvironment. Now we can utilize the microenvironment as input to a CNN model and the masked focal residue as a label to conduct

Supplementary Table 1. Oligonucleotide and template sequences used in the study		
Name	**Sequence**	**Use**
gapdLAMP.F3	GCCACCCAGAAGACTGTG	gapd LAMP-OSD
gapdLAMP.B3	TGGCAGGTTTTTCTAGACGG	
gapdLAMP.FIP	CGCCAGTAGAGGCAGGGATGAGGGAAACTGTGGCGTGAT	
gapdLAMP.BIP	GGTCATCCCTGAGCTGAACGGTCAGGTCCACCACTGACAC	
gapdLAMP.LR	TGTTCTGGAGAGCCCCGCGGCC	
gapdOSD.F	/56-FAM/CTCACTGGCATGGCCTTCCGTGTCCCCACTGCCAAC/3InvdT/	
gapdOSD.Q	GGACACGGAAGGCCATGCCAGTGAG/3IABkFQ/	
gapd template	CTAGTAACGGCCGCCAGTGTGCTGGAATTCCCACAGTCCATGCCATCAC TGCCACCCAGAAGACTGTGGATGGCCCCTCCGGGAAACTGTGGCGTGA TGGCCGCGGGGCTCTCCAGAACATCATCCCTGCCTCTACTGGCGCTGC CAAGGCTGTGGGCAAGGTCATCCCTGAGCTGAACGGGAAGCTCACTGG CATGGCCTTCCGTGTCCCCACTGCCAACGTGTCAGTGGTGGACCTGAC CTGCCGTCTAGAAAAACCTGCCAAATATGATGACATCAAGAAGGTGGTG AAGCAGGCGTCGGAGGGCCCCCTCAAGGGCATCCTGGGCTACACTGA GCACCAGGTGGTCTCCTCTGACTTCAACAGCGACACCCACTCCTCCACC TTTGACGCTGGGGCTGGCATTGCCCTCAACGACCACTTTGTCAAGCTCA TTTCCTGGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGC	
V191L F	CTGCTTTTAGAGTTAGAACAGCCTC	Mut1
V191L R	GCGGTCCTGCTCGTTGCGAC	
T493N F	GATATCAACAGCCGTAACTTCAATGTAC	Mut2
T493N R	AGGAAGGTAGCGACGACGGTG	
A552G F	CTTGAGGGCCCCAAGGAAGAGATG	Mut3

self-supervised learning. MutCompute predicted an asparagine as the most probable amino acid to belong in the center of this microenvironment

A552G R	GATCAATTCGTCATGGACTTGCAGT	
R562V F	CTTTGCGTTCTGGTGCCGGAAGTA	Mut4
R562V R	ACGCTCCATCTCTTCCTTGGG	
S371D F	GATTACGATCAGATCGAGCTTCGCG	Mut5
S371D R	TGCTGCGAAGATCAGCCAGTCC	
N528E F	GACTTACAGGCTCGTCTGAAGGAAG	Mut6
N528E R	GATCATAGCTTTCTTGATGATATCTG	
T510F F	ATGAATTTCCCCATCCAGGGGTCAG	Mut7
T510F R	CGCCATGCGTTCGGCGAAG	
I304V F	TCTACCTACGTTGAGGGTCTGTTAAAGGTC	Mut8
I304V R	CTGCAGCTTGCCCAGCTGGC	
Y303H F	TCTACCCACATTGAGGGTCTGTTAAAGGTC	Mut9
Y303H R	CTGCAGCTTGCCCAGCTGGC	
V572A F	GGCCGCGACGCTGCGCGTAC	Mut10
V572A R	TGTTCCATTACTTCCGGCACCAG	

Supplementary Table 2. Full sequence of pKAR2-Br512; 6218 bp. Br512 is highlighted.

Detailed annotation of Br512

4285-4311 : 8x His

4312-4452 : HP47

4453-4476 : 2(GS)3(A)P

4476-6216 : Bst-LF

ATCCTAAGCTTAATTTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGT
TTTTTGAGCTGAGCTTGGACTCCTGTTGATAGATCCGGCCGGTAATGACCTCAGAA
CTCCATCTGCCTAATGGAGTGATTCTAGCCGGTCGTTACACGTGGAACGGGAACT
GCCAGACATCAAATAAAACAAAAGGCTCAGTCGGAAGACTGGGCCTTTTGTTTTAT
CTGTTGTTTGTCGGTGAACACTCTCCCGGAAAACTCACGTTAAGGGATTTTGGTCA
TGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAACTAGTGAAGTTACCATC
ACGGAAAAAGGTTATGCTGCTTTTAAGACCCACTTTCACATTTAAGTTGTTTTTCTAA
TCCGCAAATGATCAATTCAAGGCCGAATAAGAAGGCTGGCTCTGCACCTTGGTGAT
CAAATAATTCGATAGCTTGTCGTAATAATGGCGGCATACTATCAGTAGTAGGTGTTT
CCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGCAACCTAAAGTAA
AATGCCCCACTGCACTGAGTGCATATAATGCATTCTCTAGTGAAAAACCTTGTTGG
CATAAAAAGGCTAATTGATTTTCGAGAGTTTCATACTGTTTTTCTGTAGGCCGTGTA
CCTAAATGTACTTTTGCTCCATCGCGATGACTTAGTAAAGCACATCTAAAACTTTTA
GCGTTATTACGTAAAAAATCTTGCCAGCTTTCCCCTTCTAAAGGGCAAAAGTGAGT
ATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAAAGCCCGCTTATTTT

TTACATGCCAATACAATGTAGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGG
GTTGTTAAACCTTCGATTCCGACCTCATTAAGCAGCTCTAATGCGCTGTTAATCACT
TTACTTTTATCTAAACGAGACATCATTAACCTCCTCAAGAGGATCGAATAGTTATTA
CCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCAT
AGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCT
GGCCCCAGTGCTGCAATGATACCGCGAGAGCCACGCTCACCGGCTCCAGATTTAT
CAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTT
ATCCGCCTCCATCCAGTCTATCAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGC
CAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGC
TCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTG
TCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAAT
TCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACC
AAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAAT
ACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAAC
GTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATG
TAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCT
GGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
CGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG
GTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATTTTTTAAGGCA
GTTATTGGTGCCGCTTAAACGCCTGGGGTAATGACTCTCTAGCTTGAGGCATCAAA
TAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCG
GTGAACGCTCTCCTGAGTAGGACAAATCCGCCCTCTAGATTACGTGCAGTCGATG
ATAAGCTGTCAAACGGAATTTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCC
GGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTC
CGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGT
ATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCA
GGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT
TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG
TAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAA
CCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAG
CAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTAC
GGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGA
TCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAG
GGATTTTGGTCATGGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTG
CAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAAT
GAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGT
CTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAA
TAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGC

AAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCA
TCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAA
GACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAAC
CGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTC
TTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCAT
CATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAG
CCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATG
TTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCAC
CTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGT
TGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACA
CCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTGTAATCGTTAATCCGC
AAATAACGTAAAAACCCGCTTCGGCGGGTTTTTTTATGGGGGGAGTTTAGGGAAAG
AGCATTTGTCATCATGACCATGACATTAACCTATAAAAATAGGCGTATCACGAGGC
CCTTTCCCTAGGGTCTTCACACTCTATCATTGATAGAGTTAATACGACTCACTATAG
GGTCCCTATCAGTGATAGAGAGAATTCGTACTGAGCACAGCTGTCACCGGATGTG
CTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCCTCTACAAATAATTTTGTTTA
AACTAGTTAGATAAGGAGGTTACATATGCACCATCATCACGGTCATCACCACCATC
CGCGTGGTGTTGACCCGAGCCGTAAGGAGAACCACCTGTCTGACGAAGACTTCAA
GGCGGTGTTCGGTATGACCCGTTCTGCGTTCGCGAACCTGCCGCTGTGGAAACAA
CAGAACCTGAAGAAGGAGAAAGGTCTGTTCGGTTCTGGAAGCGCAGCAGCACCTA
AGATGGCATTCACATTGGCCGATCGTGTCACCGAAGAGATGCTGGCAGACAAGGC
AGCCTTGGTCGTGGAGGTAGTTGAGGAGAACTATCACGACGCACCGATTGTTGGA
ATCGCCGTGGTCAATGAACATGGTCGCTTCTTCTTGCGCCCTGAGACTGCGTTGG
CCGACCCACAATTCGTGGCCTGGTTAGGAGATGAAACGAAGAAGAAGTCAATGTT
CGACAGCAAACGCGCAGCCGTAGCTCTGAAGTGGAAAGGAATTGAGCTGTGTGGT
GTGAGTTTCGACCTTCTCTTAGCAGCGTACTTGCTTGATCCCGCTCAAGGCGTCGA
CGACGTGGCAGCCGCTGCCAAGATGAAGCAATATGAAGCGGTGCGTCCGGATGA
GGCTGTGTACGGGAAGGGAGCTAAACGCGCGGTGCCTGATGAACCCGTGCTTGC
TGAGCACTTGGTACGCAAGGCTGCGGCTATCTGGGAGCTGGAGCGTCCCTTCCTG
GATGAGTTGCGTCGCAACGAGCAGGACCGCCTGCTTGTAGAGTTAGAACAGCCTC
TTAGCTCTATTCTTGCCGAGATGGAGTTCGCTGGTGTCAAAGTAGATACCAAGCGC
CTTGAGCAAATGGGTAAGGAGTTGGCTGAACAACTGGGCACAGTGGAACAGCGTA
TCTACGAACTGGCCGGTCAGGAGTTCAACATCAACAGCCCCAAGCAGCTGGGAGT
GATCCTGTTCGAGAAGTTGCAGCTGCCAGTATTGAAGAAGACTAAGACTGGCTACA
GTACCTCGGCTGACGTACTGGAGAAGCTGGCTCCTTACCATGAGATCGTGGAGAA
CATCTTGCACTACCGCCAGCTGGGCAAGCTGCAGTCTACCTACATTGAGGGTCTG
TTAAAGGTCGTGCGTCCAGACACGAAGAAGGTGCATACGATCTTCAATCAGGCGC
TGACCCAAACTGGTCGTTTGTCGTCCACAGAGCCCAATCTTCAGAATATCCCTATT
CGTCTTGAGGAAGGCCGCAAGATTCGCCAGGCCTTCGTTCCTTCGGAATCGGACT
GGCTGATCTTCGCAGCAGATTACTCACAGATCGAGCTTCGCGTGTTGGCACATATC
GCGGAGGATGACAACTTAATGGAGGCGTTCCGCCGCGATCTGGATATCCATACTA
AGACCGCGATGGATATCTTCCAAGTGTCAGAAGACGAGGTAACACCGAACATGCG
ACGCCAGGCGAAAGCGGTTAACTTCGGCATCGTCTACGGCATCAGCGACTATGGC
CTGGCCCAGAACTTGAACATCAGCCGCAAGGAGGCAGCCGAGTTCATCGAGCGCT
ACTTCGAGAGTTTCCCAGGTGTGAAGCGTTATATGGAGAATATCGTACAAGAGGCG

AAGCAGAAAGGCTACGTGACCACGCTGTTACACCGTCGTCGCTACCTTCCTGATAT
CACTAGCCGTAACTTCAATGTACGTTCCTTCGCCGAACGCATGGCGATGAATACCC
CCATCCAGGGGTCAGCTGCAGATATCATCAAGAAAGCTATGATCGACTTAAACGCT
CGTCTGAAGGAAGAACGCTTACAGGCGCACCTCTTACTGCAAGTCCATGACGAATT
GATCCTTGAGGCGCCCAAGGAAGAGATGGAGCGTCTTTGCCGTCTGGTGCCGGA
AGTAATGGAACAGGCCGTCACGCTGCGCGTACCTCTGAAAGTCGATTACCACTAC
GGCTCCACCTGGTATGACGCCAAGTAAGG