

Applied Microbiology and Biotechnology

A novel galactolipase from a green microalga *Chlorella kessleri*: purification, characterization, molecular cloning and heterologous expression

Shuhei HASHIRO¹, Koyu FUJIUCHI², Daisuke SUGIMORI^{2*}, and Hisashi YASUEDA^{1*}

¹ *Institute for Innovation, Ajinomoto Co., Inc., 1-1 Suzuki-cho, Kawasaki-ku, Kawasaki 210-8681, Japan*

² *Department of Symbiotic Systems Science and Technology, Graduate School of Symbiotic Systems Science and Technology, Fukushima University, 1 Kanayagawa, Fukushima 960-1296, Japan*

*To whom correspondence should be addressed:

Hisashi YASUEDA¹, Tel.: +81-44-245-8554; Fax: +81-44-244-9617; E-mail: hisashi_yasueda@ajinomoto.com

Daisuke SUGIMORI², Tel.: +81-24-548-8206; Fax: +81-24-548-8206; E-mail: sugimori@sss.fukushima-u.ac.jp

Supplementary figures

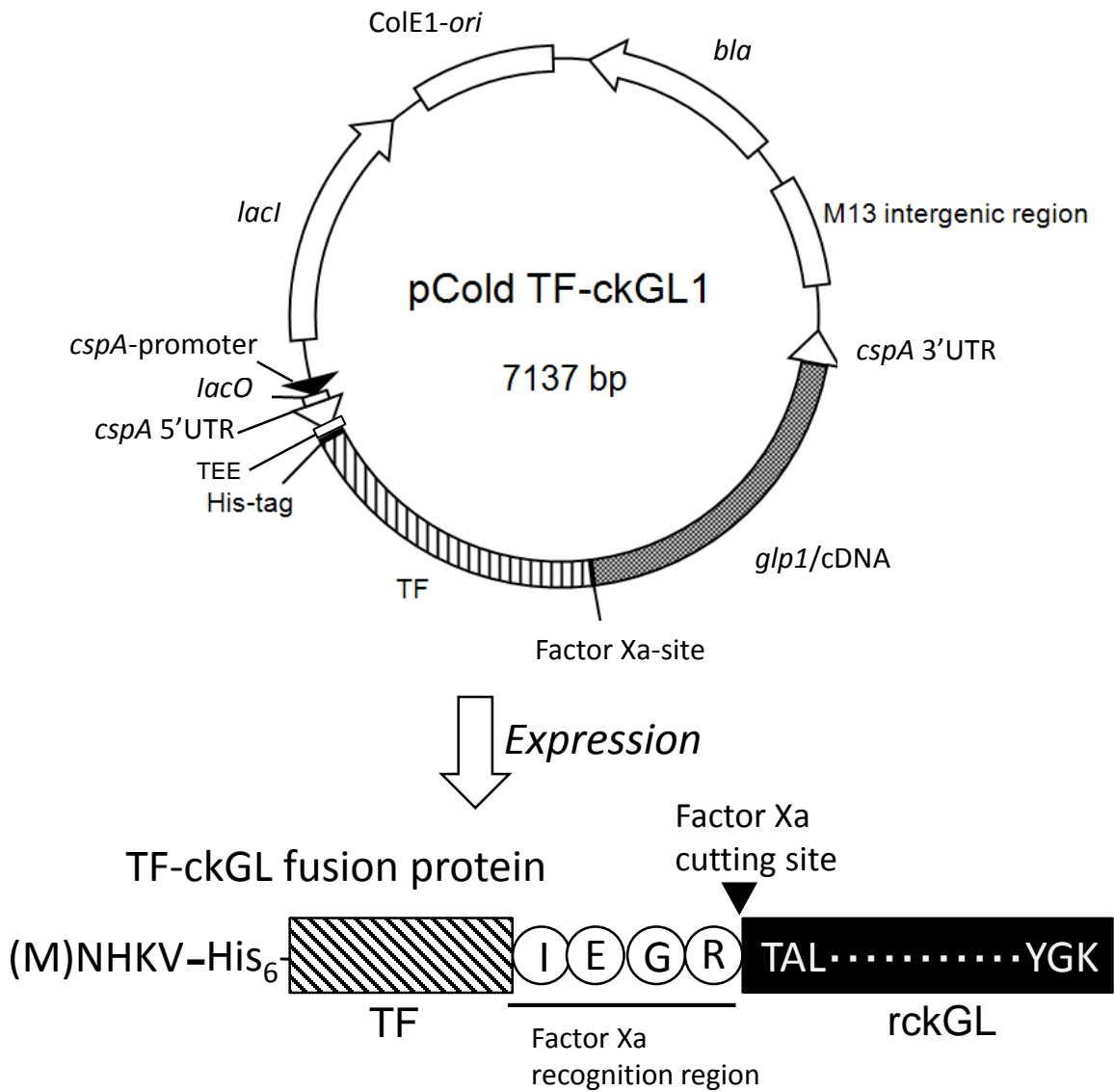


Figure S1 Structure of the rckGL expression plasmid pCold TF-ckGL1 and the TF-ckGL fusion protein. The fusion protein was produced using the promoter of the *cspA* gene, which encodes a cold shock protein. TEE, translation enhancing element; His₆, His-tag; TF, trigger factor; UTR, untranslated region. The recognition region and cutting site of factor Xa are indicated as "I-E-G-R" and a black down-pointing triangle, respectively.

10 MAHPKALLQL QTGALLQAHV RYGGAMRRAS LATSCLALQR SRRIGTAVEA RRLSAAAPVLA TALREVAATKS EVSEARQLEL 80
 90 90 100 110 120 130 140 150 160
 RGSNDWKGLL PTKENDTIDP DLRAIILMYG EMAEVTYDSF IGDPSRRFFG RNREKPSDL YPDIVLPQSK ARSYELLPNV
 170 180 190 200 210 220 230 240
 DEAYMFATEG AYKLSDDSKG GNPSPDRIAD PQKNWIGYI AVSKEAEDGM RDIAMFRGT ITGGEWKINI FGDRLVAWDP
 250 260 270 280 290 300 310 320
 RTPGLGGKVH EGFRATYWET GDQDLDSPPRA TVRRALEAIL RKHGTSIGSV TVTGHSLGGA LASVAAYDVG LRLKSTPEFGS
 330 340 350 360 370 380 390 400
 IRDKLTNEGK KIRDKIHVSA IIFASPRVGN PDFKHAVKQA GVKMLRVVNY RDVVPQVPGF LSTAVTKSLI LAGIDARAPG
 410 420 430 440 450 460 470 480
 DGVGIVLKFL NWGLSKVPVL GDWSYEHTDP EWTVNTDDIT YVEGPAANNAK RPRMVLERKD VGPRHNLEVY LYMVATLSAQ
 490 500 510 520 530 540 550 560
 KGGAIERDPV LLNKGDCWCE SKWLEEHKRP DTPKAWFQPQ VCETFSQWKK DTYGK

Figure S2 Amino acid sequence of pre-mature ckGL.

The pre-mature ckGL was composed of 535 amino acids (aa) and a pre-sequence (60 aa) containing a chloroplast transit peptide (cTP, shaded in gray). The initiator Met of pre-mature ckGL was assigned as position +1 of the aa sequence, and the *N*-terminal Thr of mature ckGL at position 61 is indicated with a dot. The *N*-terminal aa sequence of purified ckGL was detected experimentally (bold underline). Amino acid sequences detected by LC-MS/MS analysis of ckGL were all assigned to corresponding sequences (indicated by double underlines), although the MS-analysis could not distinguish between Leu and Ile residues since they have identical mass. The one negatively charged aa (Glu) and Arg residues in the cTP region are indicated by an open circle and open triangles, respectively. Two Cys residues are indicated by closed triangles.

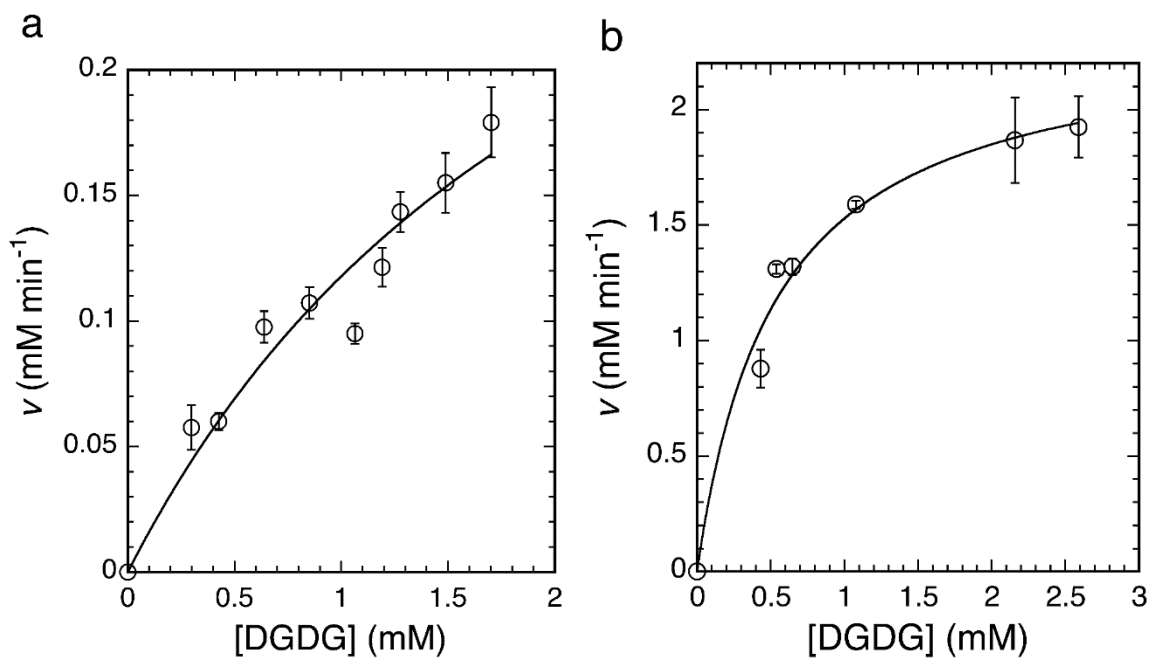


Figure S3 Steady-state kinetic analyses of (a) the purified native ckGL and (b) rckGL. Each data point represents the mean of three independent determinations with standard errors indicated.

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
 GACTGAGC ACAGTTGAG TCGTTGCT AATAGCGAC CCCGATC ACCTACTCET CAGTTGTCG TCCACCCCTT CCTGTAATG CTGCGAGCA CATAGCTTGT AAGGTCACCC CAGAGTGGG AATGTCACAC
 ATGCGGCGCG CCGTCTGAGCGG TCTCTGTC AGGCCCAATG CCGCGTGSSG GTGSCCAITG CCGTGTGTC TTGGCAACTI TCCAGTGTG CIGTTCAGCG ATCGGCGCG CAGSTGATG TSCACAGAG
 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
 CACTTITTC CTTTCTCA ATTIGATCG CTTGAGCCTG AATAGTCTT GATACTCCT AGGCGTATC CAITAGAGT GCGCTACTI GCGTCTAGS GGCTAGCAI CTGCGACTG AACACGCTG SCATCTITCA
 GAGCTGTGAA CGAGTGTCC TGTGACTCC ACTGTGAANA CDTGGCCCC GCTACTGACC CCGGTGTAG ATACTGTGT ACTGATFVG AAGCCGGAG CTTGAGCTG TGGCACTGC CDTGGCGAG GTTGCGGCA
 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450
 AGTCTGAGT CAGTGAAGC CGGAGCTT AGCTGCGGC CACGACGAC TGGAAAGTA GATGTCCTT TAGATATAT GTGAGGACA TTGGCTGCG TCTGTGCGG GGGCTTCGT CTAATGTTG GSTATGACC CTAGGAGTAA
 TGTAAITTC AAGCTTCCA CCCCAGCTT GACTCTCCT AATGTTTAC CAGGCTGAG TACTGTGA TGGCTGTAC CCAGTGTTC TGTGCTTTC ATGACAGCC CAGCGACTT CCATCTCCG ASSETCTGT
 CCAACCAAG AGAATGAC CACTGAGCC GACTTGGGG CCACCAITC GATGATGSS GAGATGGGS AGGTACCTA CGACAGTTC ATCGGAGCC CTTCTACG CTTGTAGT TCTGTGTC AAGTGGGT CCGTAGCGG
 460 470 480 490 500 510 520 530 540 550 560 570 580 590 600
 CTGCTTTCA CTGTGCTAG CCGACAGGCG AGCTGSCCT TTTTATGA GCGGGGCA TAGGCTGCA TGGCCAMGS CACCAITGA TGGGCCATC TAACTGTG GAGGCCCTI GGAGCGGCTI SCCACTGCT GATTGACCC
 CTTGATGGA CATAAGTA ACTGCAACC GACTGAGAG TCGAGAGCA GCGCCTTAGS GAACTGCAC AUCTGCACG CCGCAACCA GTGCGCCTI GGCGGCTG TGTGCTGCA GAAAGGACT GTTTCGCA CAIAGGCTI
 610 620 630 640 650 660 670 680 690 700 710 720 730 740 750
 GCAGTTTGG GAGGACCCG TTTGAGAGC CFTGTGACT TACCAGAC ATCGTGTG CTATCTGTA GCGCGGAGS TCTGTTGCG TGTGTAACC GGGGCTTTI GGTTCATCA TTGCAATG GTTGTGTT AAGCTAGCGG
 GCAATGCGAC ACTACACTG CTRACAGG GACCTGCA ITGGTGAAT GCAITGSGC AITCGGTTC AATCGGCTT TGCAGGTCC TGCAGAGA ACCAAGAT TGTGAGGAA ATTCGCTG CTGTGGTTC
 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900
 TTTGACTTG TTAGGCACC CATACTTC TCCAAITGG TCGAGAGG GTCCTAGATC ACAGAGAG ATGGTGGCA GCTTGGTGT CCTGATGTT GCGGTCTGT GGGATCGCT GGGATCGCT CCCCTGTCA
 910 920 930 940 950 960 970 980 990 1000 1010 1020 1030 1040 1050
 AGAAATGTA AGCACTTCCG TTTCCGTTG AGTCTGTTA AGCTGTTGT GACCCAGAC TTAGCCTGC GAITCTGTA TTCTCTCCAC ACAGCTAGA SCTCTCTCC AATGTCAGC AGSCCTACAI GTTTCGAGC GAGGCTGCTI
 1060 1070 1080 1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 1960 1970 1980 1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
 ACHARCTGAG CGAGGATCC AAGGAGGCA ACCCTCTCCC CGACCGGAT CCGCAACCCG ATGGATAGC ATGGATAGC TACTATCCAG TGTCAAAGA GGCAGGTCTAG TGCCCGCACA GACCCGCGCC AGGGSCCTGC TGCRAAGTTI
 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200 2210 2220 2230 2240 2250
 GCTGCGAAGC TTTGTTGCA GATTTTCA CAACTGCATA AAGTTGAGC AAGGCGAGS CAAGTGAIC TTGAGAGTGT AITGGCAITG CTTGGGTTI GAGCAGGAG GTAGTGAIG TCTCTGAMGG GECTCCAGG ATCCCTGTTI
 GAGGATCTG GATGAGAT CGCACCCAC GCTCTTTGCG TCGAGAGGAI GGAATGGAG ACAITGGCAI CAITTTCCGS GGCACATCA CCGGCGCGGA GTGGAAGCA ACACTTITG GCGACAGTC TGTCTAGS ITTGGAGAC
 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520 2530 2540 2550
 CAGBAGCAC ATGCTTTGCA CTGACATGT GCTGCTGAG TCGACACTI ACCCTTAAG ACAGCCATG AATTAGAT GTCACACT GTAATCGC AITCGGGGA TCCGSGGGA TGTGGTCA AATGCTGTI GACCTCTAI
 2560 2570 2580 2590 2600 2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
 GCCCAAAIT GCTGTAGCAG CFTGCGCCG CTTTCTGAGC GTTTCATCA GAAATCTGC ACTGCACAC AGGGCTGAC TCAAGGTTG TGACTACAI GACACCGG AACACCAITA CTTCTGTG CAGGCTGTTI
 2710 2720 2730 2740 2750 2760 2770 2780 2790 2800 2810 2820 2830 2840 2850
 GCTGGGACC CCGGACCCC CCGGCTGGG GCGAGATGTC ATGAGGCTC CAGGCGACC TACTGGGA CCGGGGTA CTTCTCTG GCGTGTGT AITGAGCTI GCAATGCTI TTAGCAGG CTTCTGTTI
 2860 2870 2880 2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
 GCAGCAGTAT CTTTGTGCA CAAAGGCGA CACTGTGACA TGTATTTC CRAAGAGT AITGGGAAC GATTCCTACC CAAAGCTTGT ACTGGATGC CTTGACCTC CAITGCTTGT TCCACCAICI CTAITGCGCTG CCGAGCTTGC

Figure S4 DNA sequence of *glp1* gene.
 The sequence of *glp1* gene contained 10 introns and 11 exons (regions with underlines). The ATG for initiation codon and the ACT codon for the N-terminal amino acid (Thr) of mature cKGL are indicated with bold bars. The 5'- and 3'-untranslated regions are underlined with wavy lines.

3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120 3130 3140 3150
CGGACAGC CTGTACTG CTGTAATGTA CGGTACAGTG TGGACAGGC CACAGAGCA CTACTTATG GCAATGTG CTAGAGCTG GGGTCTTCA CCGCTTACC CCGTGTGTC TGGACCTCAI CTGTGCTI CTTCAGAG
3160 3170 3180 3190 3200 3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
AGCCAGAG TTAGCTAG TCCCAAGI CTTCCTGGT TCGACCTC TGTGGCTGT TCGCCCTAG ACTGCAGAC GTCCCGGSG CTGAGAGC CATTCTGGC AATCTAGCA CTTCAGTGS CTCGCTACA GTGAGGCC
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400 3410 3420 3430 3440 3450
ACAGCTGGS GCGTGGCT GCGAGGTGS CAGCTATG TGTGGCTG TCGCTAGAT CACAGCAAC TGGCAGTC AGGCAAC AGGACAGT AGGACAGT ACGGCAACA AGATCCAGT CTCAGCAATA ACCTTTGGT
3460 3470 3480 3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
CGCCGTGT GGCACACCC GACTTCAAG ATCTGTGAA GCAGCGGSG GTAGAGTGT TGGAGTCT CAACTACAG GAGTGTGTC CAAAGTGTCC CCGGAGAGC CCGGAGAGC TTAGCTGTS GGGTGGTII GTACTGCT
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720 3730 3740 3750
GCAGGTGC GCAGCGGT GCTGTGTA CCGAGTGT GCAGCCCAA TGGATAGTA CAATAGGCT TTAGGCTT ASGCACTC CTATTCT AGTCCACT AGCCCTGSG TTGCTTGT ATTGCTCT TCAATAGCG
3760 3770 3780 3790 3800 3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
CAACAACA GAGTGTGCA GTTGGCAG CAGCAATAG AGCAACCG CAGCAAC CTAAGACAG AGCCCTAA TCAATAGT CTAGAGTCT TTATAGT TGGAGTCT TTAGAGTCT CTGAGCAGS CTGTGACAA
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000 4010 4020 4030 4040 4050
GAGCTCAC TGGCGGCA TTAGTCCAG GCTCCAGT GAGGTGTS GCAGTGTCT CAGTCTTC ACTGSSAG TCCAGAGT TCGGCTGC CTAGAGTCT GATAGAGCA CACCGACCA CAGGTGCGA GAGGTGTS
4060 4070 4080 4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
CCGGTAGS CTCAGTGC TGTCTCAG ACCTGCTCT GTTGTGTC TCTTCTC TCGGTGGG CCGAGSCA CATTAAAG TGGTTCAT GATGTCCAG CAGGSSGTC TCGTACCC ACGCCCTGS
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320 4330 4340 4350
CTGTGCTT GTGTTTCC CAGGGACA CTCAGGCAC GCGCAGGCT GTGCTTGA GGTCTGGSG TATTATTA TTTATTAI TGTCTTCT CTTGGCAGC ACAGCTGAG TGGTACAGC AGCATGCT TCTTCTAGT
4360 4370 4380 4390 4400 4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
GCAGTCCA GATGCACT GACTTCCAG CTGGGCCA GAGCTGSG CTACAGCC GGTACTAC CCGCTAAA CTTGCTAC TCGCTGTC TGGTCAAT GAGGATAA CAGGAGAT ATCATAAT TGGAGSGCC
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600 4610 4620 4630 4640 4650
GCCAACAI GCAGAGGC CATGCCGT GCTGGAGGC AAGGATGSG GCGCCGCA CACTGAGG GTACAGTCT TACCCAC TGGCAGCC CTTGACTT TGGACTTT TACAGTCT GCGCTCAGI GGGAGGCTG
4660 4670 4680 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
CCGACTTG TCTTCTG GAGCCCTG CAGCCTTG CAGACTTG ACTGTGAG CAGTCTCT CCAAGTCA AGGCAACT CCGACTCT GAGTGTGC GAGTGTGC AGAGTGGC CTTCTGAGA GCGCATGGA TGGCCCTCC
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920 4930 4940 4950
TTGACCTT TCACTCTC GAGCTTCT CAGCTTAC CAGCTTAC TCGCTTAC CAGCTTAC TCGCTTAC TCGCTTAC TCGCTTAC TCGCTTAC TCGCTTAC TCGCTTAC TCGCTTAC TCGCTTAC TCGCTTAC
4960 4970 4980 4990 5000 5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
ACTGTGTT GGAGTCCAG TGGCTGGAG AGCAGAGC CCGCAGC GTCACAGT GTCACAGT GTCACAGT GTCACAGT GTCACAGT GTCACAGT GTCACAGT GTCACAGT GTCACAGT GTCACAGT GTCACAGT
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200 5210 5220 5230 5240 5250
GAAAAAI ATGGGGTCA GCAATGCG GCGAGGCG CTGGACTT TCCCTGGS ATTTTAT AGAAGGAT GAGGTAGT GCGAGAGS AGCATGCT ATTGTGCT TTGTAACA GCGCCCAT TATCCACA
5260 5270 5280 5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
AGTATGSC CAGAGAGS GGTCTGAA CAGCTTGS GGGTCAAT ACAAGTCT TACTGATC TGAACATC CCGGACT AGCACTCC TCTCATIT CCGTCTGAG TTGACTTG TTGCTCC TTGACTTG
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500 5510 5520 5530 5540 5550
CTCGPACA CTGACACA AATTTTCT TATTTTCT GAGTGTGA CAITGCTT TGAATTT TCGTCTT TCGTCTT TCGTCTT TCGTCTT TCGTCTT TCGTCTT TCGTCTT TCGTCTT TCGTCTT
5560 5570 5580 5590 5600 5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
GTACAGAT AGTGTGAG ACTGCTTA TATGAGCA GTGCTGAT GAGTCTGC TCGTCTGC TCGTCTGC TCGTCTGC TCGTCTGC TCGTCTGC TCGTCTGC TCGTCTGC TCGTCTGC TCGTCTGC TCGTCTGC
5710 5720 5730 5740 5750 5760 5770 5780 5790 5800 5810 5820 5830 5840 5850

Figure S4 (continued)

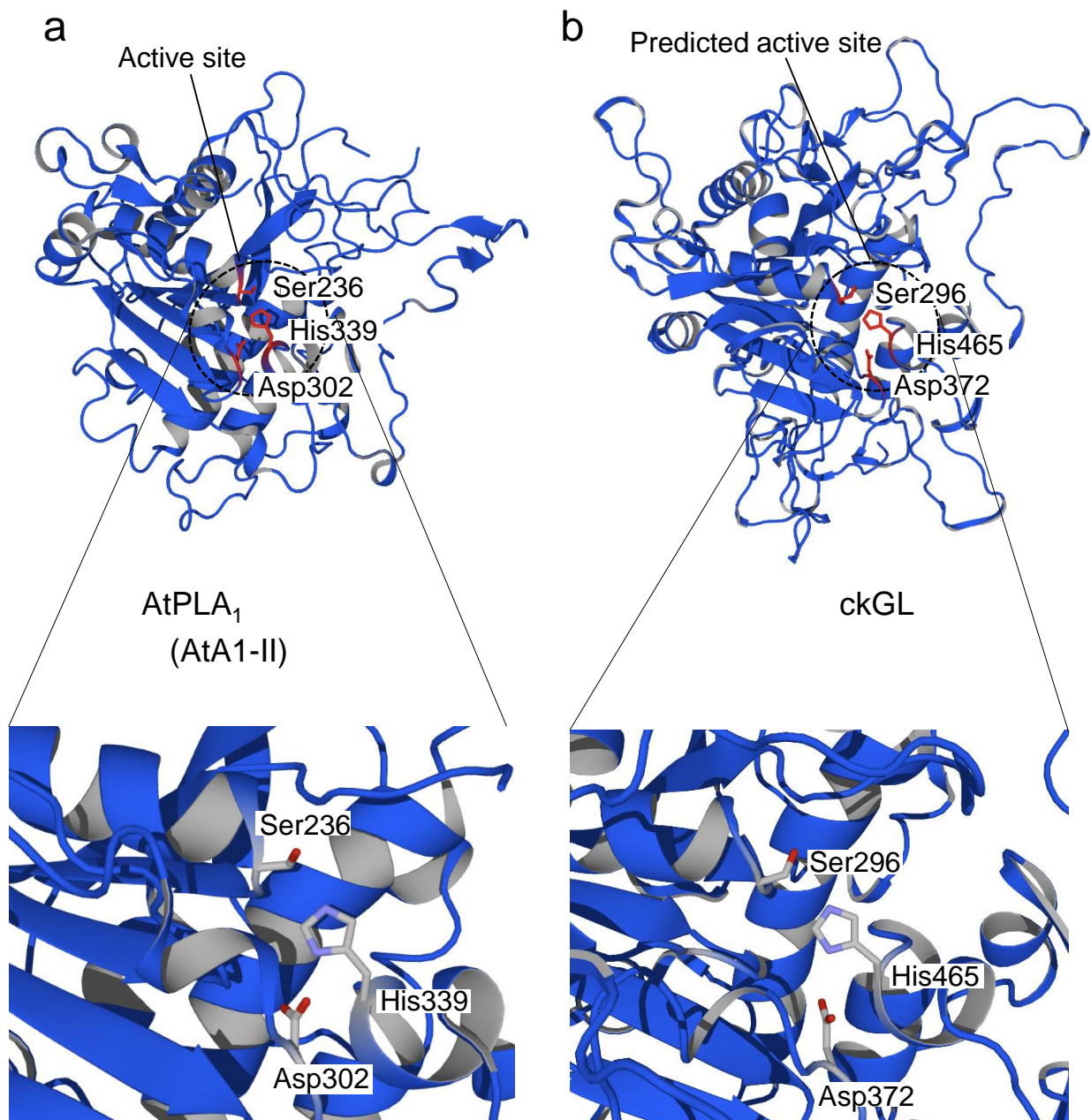


Figure S5 Crystal and model structures of AtPLA₁ and ckGL. (a) PLA₁ from *A. thaliana* (PDB ID, 2YIJ) as crystal structure; (b) ckGL as homology-modeled structure. The structures are represented in ribbon models as monomer from the same viewpoint. The putative catalytic triad (Ser, Asp, and His) in the area within black dotted circle is shown in red. In the lower panels, close-up views of the catalytic triads, with oxygen and nitrogen atoms in red and cyan, respectively, are shown as stick models.