Supplementary Information for

Engineered protein cages with phosphate-dependent reversible opening

V. Vishal Malolan[1,2], Lukasz Koziej[1], Jonathan G. Heddle,[1,3] Agnieszka Gawin[1], and Yusuke Azuma[1*]

1. Malopolska Centre of Biotechnology, Jagiellonian University, 30-387 Krakow, Poland

2. Doctoral School of Exact and Natural Sciences, Jagiellonian University, 30-348 Krakow, Poland

3. Centre for Programmable Biological Matter, School of Biological and Biomedical Sciences, Durham University, Durham DH1 3LE, U.K.

*E-mail: yusuke.azuma@uj.edu.pl

Supplementary Tables 1-3

Supplementary Table 1 | Proteins used in this study

Name	Sequence	MW[a]
AaLS-wt (Strep)	MEIYEGKLTA EGLRFGIVAS RFNHALVDRL VEGAIDCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAIERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18102.74
AaLS-POF (R21C, E32H, D36H, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18103.75
AaLS (R21C, **E32S**, D36H, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VSGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18053.68
AaLS (R21C, E32H, **D36S**, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VHGAISCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18053.68
AaLS (R21C, E32H, D36H, **I125S**) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQASERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18053.68
AaLS (R21C, E32S, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VSGAIDCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18031.63
AaLS (R21C, D36S, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VEGAISCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18045.66
AaLS (R21C, E32H, D36H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAIERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18079.77
AaLS (**R21S**, E32H, D36H, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS SFNHALVDRL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18087.69
AaLS (E32H, D36H, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS RFNHALVDRL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18156.80

AaLS (R21C, E32H, D36H, **T86A**, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGAAPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18073.72
AaLS (R21C, E32H, D36H, I125H, **R127A**) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDRL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHEAAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18018.64
AaLS (R21C, **R29A**, E32H, D36H, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNHALVDAL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18018.64
AaLS (R21C, **H24S**, E32H, D36H, I125H) (Strep)	MEIYEGKLTA EGLRFGIVAS CFNSALVDRL VHGAIHCIVR HGGREEDITL VRVPGSWEIP VAAGELARKE DIDAVIAIGV LIRGATPHFD YIASEVSKGL ANLSLELRKP ITFGVITADT LEQAHERAGT KHGNKGWEAA LSAIEMANLF KSLRLEGGWS HPQFEK*	18053.68
cpAaLS(119)-tev*-msfGFP	MTLEQAIERA GTKHGNKGWE AALSAIEMAN LFKSLRGTGG SGSSMEIYEG KLTAEGLRFG IVASRFNHAL VDRLVEGAID CIVRHGGREE DITLVRVPGS WEIPVAAGEL ARKEDIDAVI AIGVLIRGAT PHFDYIASEV SKGLANLSLE LRKPITFGVI TADAGGAGGS ENLYFQSGGS GSKGEELFTG VVPILVELDG DVNGHKFSVR GEGEGDATNG KLTLKFICTT GKLPVPWPTL VTTLTYGVQC FSRYPDHMKR HDFFKSAMPE GYVQERTISF KDDGTYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYNFNSHNV YITADKQKNG IKANFKIRHN VEDGSVQLAD HYQQNTPIGD GPVLLPDNHY LSTQSKLSKD PNEKRDHMVL LEFVTAAGIT HGMDELYK*	45686.62
cpAaLS(119)-srt*	MTLEQAIERA GTKHGNKGWE AALSAIEMAN LFKSLRGTGG SGSSMEIYEG KLTAEGLRFG IVASRFNHAL VDRLVEGAID CIVRHGGREE DITLVRVPGS WEIPVAAGEL ARKEDIDAVI AIGVLIRGAT PHFDYIASEV SKGLANLSLE LRKPITFGVI TADAGGAGGS LPSTGG*	18397.96
Strep-sTEVp-R5	GWSHPQFEKG GSGESLFKGP RDYNPISSTI VHLTNESDGH TTSLYGIGFG PFIITNKHLF RRNNGTLVVQ SLHGVFKVKN TTTLQQHLID GRDMIIIRMP KDFPPFPQKL KFREPQREER IVLVTTNFQT KSMSSMVSDT SSTFPSGDGI FWKHWIQTKD GQCGSPLVST RDGFIVGIHS ASNFTNTNNY FTSVPKNFME LLTNQEAQQW VSGWRLNADS VLWGGHKVFM DKPEEPFQPV KEATQLMNRR RRR*	28846.67
SrtA*	MGSSHHHHHH GGSGMQAKPQ IPKDKSKVAG YIEIPDADIK EPVYPGPATR EQLNRGVSFA KENQSLDDQN ISIAGHTFID RPNYQFTNLK AAKKGSMVYF KVGNETRKYK MTSIRNVKPT AVEVLDEQKG KDKQLTLITC DDYNEETGVW ETRKIFVATE VKLE*	18471.84

a. calculated by Expasy ProtParam (https://web.expasy.org/protparam/)

14 **Supplementary Table 2 | Plasmids used in this study**

name	gene	N-tag	C-tag	promoter /operator[a]	Ori	marker[b]
pET_AaLS-wt-S	lumazine synthase from *Aquifex aeolicus* (AaLS)	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_E32H_ D36H_I125H-StreptII	AaLS mutant containing R21C, E32H, D36H, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_E32S_ D36H_I125H-StreptII	AaLS mutant containing R21C, E32S, D36H, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_E32H_ D36S_I125H-StreptII	AaLS mutant containing R21C, E32H, D36S, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_E32H_ D36H_I125S-StreptII	AaLS mutant containing R21C, E32H, D36H, & I125S	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_E32S_ I125H-StreptII	AaLS mutant containing R21C, E32S, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_D36S_ I125H-StreptII	AaLS mutant containing R21C, D36S, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_E32H_ D36H-StreptII	AaLS mutant containing R21C, E32H, & D36H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21S_E32H_ D36H_I125H-StreptII	AaLS mutant containing R21S, E32H, D36H, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_E32H_D36H_ I125H-StreptII	AaLS mutant containing E32H, D36H, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_E32H_ D36H_T86A_I125H-StreptII	AaLS mutant containing R21C, E32H, D36H, T86A & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_E32H_ D36H_I125H_R127A-StreptII	AaLS mutant containing R21C, E32H, D36H, I125H, & R127A	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_R29A_ E32H_D36H_I125H-StreptII	AaLS mutant containing R21C, R29A, E32H, D36H, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R
pET21_AaLS_R21C_H24S_ E32H_D36H_I125H-StreptII	AaLS mutant containing R21C, H24S, E32H, D36H, & I125H	-	Strep II	P_{T7} / *lac*O	ColE1	Amp^R

pACYC_Ptet_cpAaLS(119)-tev*-GFP	circularly permuted variant of AaLS (cpAaLS(119)) fused to GFP via a TEV protease recognition site	-	-	P_{tet} / *tet*O	p15A	Cm^R
pACYC_Ptet_cpAaLS(119)-srt*	cpAaLS(119) fused to a sortase A recognition site	-	-	P_{tet} / *tet*O	p15A	Cm^R
pMAL_S-sTEVp	engineered variant of protease from tobacco etch virus (super TEV protease)	Strep II	-	P_{tac} / *lac*O	ColE1	Amp^R
pET30_H-SrtA*	engineered variant of sortase A	His6	-	P_{T7} / *lac*O	ColE1	Kan^R

a. P_{T7}/*lac*O, T7 promoter combined with lactose operator and the LacI repressor (*lacI*); P_{tet}/*tet*O, the tetracycline promoter combined with tetracycline operator; *P*tac/*lac*O, tac promoter combined with lactose operator.

b. Amp^R, β-lactamase; Cm^R, chloramphenicol acetyltransferase; Kan^R, aminoglycoside 3'-phosphotransferase

Supplementary Table 3 | Oligonucleotides used in this study

Name	Sequence
FW_XhoI_StreptII_SpeI	TCGAGGGCGGCTGGAGCCATCCGCAGTTcGAAAAgTAA
RV_SpeI_StreptII_XhoI	CTAGTTAcTTTTCgAACTGCGGATGGCTCCAGCCGCCC
Fw_AaLS_E32SD36H	GCTCTTGTCGACCGTCTGGTGTCTGGTGCAATTCACTGCATAGTCCG
Rv_AaLS_E32SD36H	CGGACTATGCAGTGAATTGCACCAGACACCAGACGGTCGACAAGAGC
FW_AaLS_E32HD36S	CCGTCTGGTGCATGGTGCAATTTCTTGCATAGTCCGTCATGGCGG
RV_AaLS_E32HD36S	CCGCCATGACGGACTATGCAAGAAATTGCACCATGCACCAGACGG
Fw_AaLS_I125S	GCTGACACCTTGGAACAGGCTAgCGAGCGCGCCGGC
Rv_AaLS_I125S	GCCGGCGCGCTCGcTAGCCTGTTCCAAGGTGTCAGC
FW_AaLS_D36S	CCGTCTGGTGGAGGGTGCAATTTCTTGCATAGTCCGTCATGGCGG
RV_AaLS_D36S	CCGCCATGACGGACTATGCAAGAAATTGCACCCTCCACCAGACGG
Fw_AaLS_E32S	GCTCTTGTCGACCGTCTGGTGAGCGGTGCAATTGATTGCATAGTCCGTC
Rv_AaLS_E32S	GACGGACTATGCAATCAATTGCACCGCTCACCAGACGGTCGACAAGAGC
FW_AaLS_I125	GCTGACACCTTGGAACAGGCTatcGAGCGCGCCGGCAC
RV_AaLS_I125	GTGCCGGCGCGCTCGATAGCCTGTTCCAAGGTGTCAGC
FW_AaLS_R21S	CTTCGTTTCGGTATCGTAGCATCAAGCTTTAATCATGCTCTTGTCGACCGTC
RV_AaLS_R21S	GACGGTCGACAAGAGCATGATTAAAGCTTGATGCTACGATACCGAAACGAAG
Fw_AaLS_R21	CTTCGTTTCGGTATCGTAGCATCAcgtTTTAATCATGCTCTTGTCGACCGTC
Rv_AaLS_R21	GACGGTCGACAAGAGCATGATTAAAACGTGATGCTACGATACCGAAACGAAG
Fw_AaLS_T86A	CGTTCTCATCAGAGGCGCAGCGCCACATTTCGATTATATCGCCTCTG
Rv_AaLS_T86A	CAGAGGCGATATAATCGAAATGTGGCGCTGCGCCTCTGATGAGAACG
Fw_AaLS_I125H_R127A	CACCTTGGAACAGGCTCACGAGGCGGCCGGCACAAAACACGGCAAC
Rv_AaLS_I125H_R127A	GTTGCCGTGTTTTGTGCCGGCCGCCTCGTGAGCCTGTTCCAAGGTG
Fw_AaLS_R21C_R29A	CATGCTTTAATCATGCTCTTGTCGACGCGCTGGTGCATGGTGCAATTCACTGC
Rv_AaLS_R21C_R29A	GCAGTGAATTGCACCATGCACCAGCGCGTCGACAAGAGCATGATTAAAGCATG
Fw_AaLS_H24S	CGGTATCGTAGCATCATGCTTTAATAGCGCTCTTGTCGACCGTCTGGTG
Rv_AaLS_H24S	CACCAGACGGTCGACAAGAGCGCTATTAAAGCATGATGCTACGATACCG
FW_TetR	CTAATGCGCTGTTAATCACTTTAC
RV_XhoI_stop_GFP	GAGCCTCGAGTCATTATTTGTACAGTTCATCCATGCCATG
RV_XhoI_srt_GGS	GAGCCTCGAGTTAGCCGCCGGTAGACGGCAGGGATCCGCCCGCCCCACC
$dsDNA_{14}$_FW	$(Gly)_3$-GGACGTTGACGTTC
$dsDNA_{14}$_RV	FITC-GAACGTCAACGTCC
$dsDNA_{20}$_FW	$(Gly)_3$-TCCATGACGTTCCTGACGTT
$dsDNA_{20}$_RV	FITC-AACGTCAGGAACGTCATGGA
$dsDNA_{30}$_FW	$(Gly)_3$-AAATCGTCGTTTTGTCGTTTTGTCGTTTTT
$dsDNA_{30}$_RV	FITC-AAAAACGACAAAACGACAAAACGACGATTT