Electronic Supplementary Material (ESI) for ChemComm.
This journal is © The Royal Society of Chemistry 2023

Supporting Information for

Photo-regulated genetic encoding of dibenzo[*c,g*][1,2]diazocine on Proteins via Configuration Switching

Tingting Zheng, Jielin Fu, Qin Xiong, Xin Shen, Baolin Li, Xiaohu Zhao and Zhipeng Yu*

Key Laboratory of Green Chemistry and Technology of Ministry of Education, College of Chemistry, Sichuan University, 29 Wangjiang Road, Chengdu (610064), P. R. China,

E-mail: zhipengy@scu.edu.cn

Table of Content

Supplemental Figures and Tables

General Information

All chemicals and reagents were commercially obtained and used without any further purification. Commercially available chemicals were obtained from Adamas, Acros Organics, Aldrich Chemical Co., Alfa Aesar and TCI and used as received unless otherwise stated. Anhydrous solvents DMF were purchased from Acros Organics.

Exact ESI mass spectra were scanned by a Thermo Q-Exactive™ mass spectrometry instrument. LC/ESI-MS data were obtained on a Thermo LTQ-XL mass spectrometer. The ^{1}H, ^{13}C spectra were scanned on a Brüker Avance spectrometer (^{1}H: 400 MHz or 600 MHz, ^{13}C: 101 MHz). Chemical shifts (δ) for ^{1}H and ^{13}C NMR spectra are given in ppm relative to TMS. The residual solvent signals were used as references for ^{1}H and ^{13}C NMR spectra and the chemical shifts converted to the TMS scale ($CDCl_3$, 7.26 ppm for ^{1}H NMR and 77.16 ppm for ^{13}C NMR; DMSO-d_6, 2.50 ppm for ^{1}H-NMR and 39.5 ppm for ^{13}C-NMR; D_2O, 4.79 ppm for ^{1}H-NMR). Shifts multiplicity was reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, brs. = broad. The photo-irradiation power density of 405 nm LED array (13.4 $mW \cdot cm^{-2}$), 405 nm laser (500 $mW \cdot cm^{-2}$) and 532 nm laser (100 $mW \cdot cm^{-2}$) were measured via an optical power meter purchased from Thorlabs.

All primer sequences used for library construction or site-directed mutation (SDM) are listed in **Table S1**. All amino acid sequences of proteins mentioned in this work are presented in **Supplementary Notes**. Plasmid *p*BAD-sfGFP-Q204TAG, *p*ET-sfGFP-Q204TAG-His_6 and *p*ET-sfGFP-N149TAG-His_6 are reported in our previous work.[1]

PCR was carried out by using AccuPrime™ Pfx DNA polymerse (Thermo Scientific). After SDM-PCR, the reaction system was digested by DpnI to remove the template plasmid. In OE-PCR, the product was digested by BcuI and SalI to create complementary sticky ends, follow by dealing with T4 DNA ligase. Restriction enzymes and T4 DNA ligase were purchased from Thermo Scientific or NEB. E.Z.N.A.® Plasmid Mini Kit I (from Omega Bio-tek®) was used for purifying restriction digested DNA and gel purified DNA according to the manufacturer protocol. *E. coli* DH5α competent cells and *E. coli* BL21 (DE3) competent cells (TransGen Biotech)

were obtained from commercial source for chemical transformation. The synthesis of oligonucleotide primers, gene fragments and DNA sequencing were serviced and conducted by Tsingke Biotechnology. The enzymatic activity of GST mutants was measured by Glutathione S-transferase (GST) Activity Assay Kit, which was purchased from Solarbio. Antibodies for Western-Blot analysis were purchased from Beyotime Biotechnology or ZEN-BIOSCIENCE.

Supplemental Notes: The amino acid sequence of proteins

Note S1. The amino acid sequence of the evolved *Mmc*DBDAARS:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRS
SRTARALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTK
KAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSIS
TGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGK
PFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPL
EYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEI
GPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVG
DSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKV
KHDFKNIKRAARSESYYNGISTNL*.

Where * is the “ochre” stop codon, TAA.

Note S2. The amino acid sequence of the Opti-*Mmc*DBDAARS:

MDKKPLNTLISATGLWMSHTGTIHKIKHREVSRSKIYIEMACGDHLVVNNSRS
SRTARALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTK
KAMPKSVARAPKPLENSEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSIS
TGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGK
PFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPL
EYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEI
GPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVG

DSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKV
KHDFKNIKRAARSESYYNGISTNL*.
Where * is the “ochre” stop codon, TAA.

Note S3. The amino acid sequence of the evolved *Mmt*DBDAARS:
MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRS
SRTARALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTK
KAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSIS
TGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGK
PFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPL
EYIERMGIDNDTELSKQIFRVDKNFCLRPMLTPNLYNYLRKLDRALPDPIKIFEI
GPCYRKESDGKEHLEEFTMLAFAQMGSGCTRENLESIITDFLNHLGIDFKIVGD
SCMVFGDTLDVMHGDLELSSALVGPIPLDREWGIDKPTIGAGFGLERLLKVK
HDFKNIKRAARSESYYNGISTNL*.
Where * is the “ochre” stop codon, TAA.

Note S4. The amino acid sequence of sfGFP-Q204TAG-His_6:
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGN
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQK
NGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTXSALSKDP
NEKRDHMVLLEFVTAAGITHGMDELYKEHHHHHH*.
Where * is the “ochre” stop codon, TAA; X = “amber” codon: TAG.

Note S5: The amino acid sequence of EGFP-N149TAG-HA:
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGK
LPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHXVYIMADK
QKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK
DPNEKRDHMVLLEFVTAAGITLGMDELYKMYPYDVPDYA*.
Where * is the “ochre” stop codon, TAA; X = “amber” codon: TAG.

Note S6: The amino acid sequence of sfGFP-N149TAG-His_6:

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKL

PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGN

YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHXVYITADKQK

NGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDP

NEKRDHMVLLEFVTAAGITHGMDELYKEHHHHHH*

Where * is the "ochre" stop codon, TAA; X = amber codon: TAG.

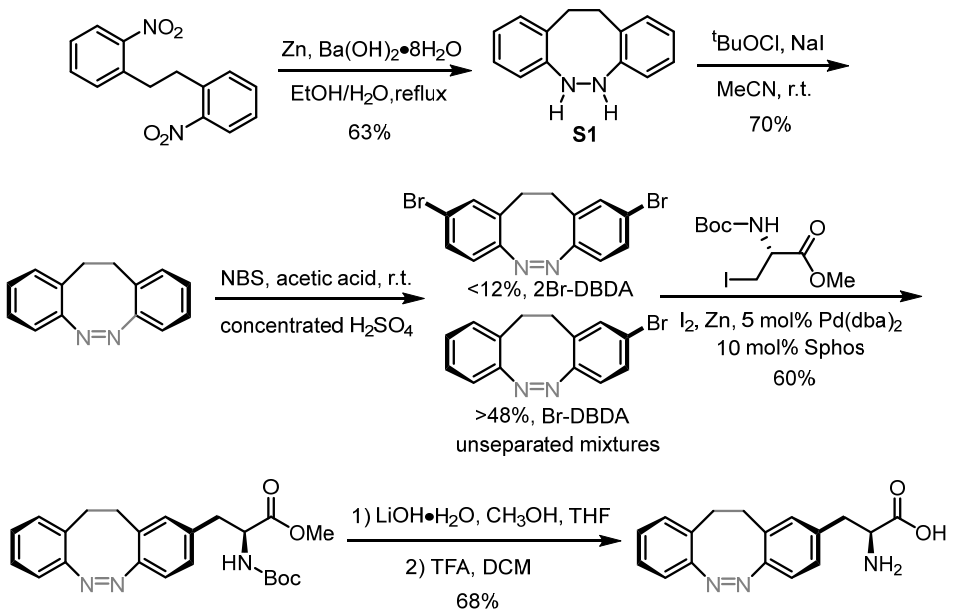
Supplemental Tables: Primers used for plasmid construction

Table S1: Primers used for cloning

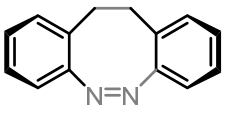
	Primer	Sequence (5'- 3')
*Mm*PylRS-A302T	pEVOL-A302T-F	TGCCTGAGACCCATACCTGCTCCAAACCTTTAC
	pEVOL-A302T-R	AGTTGTAAAGGTTTGGGGTAAGCATGGGTCTC
*Mm*PylRS-Y384F	pEVOL-Y384F-F	ATTCCTGCATGGTCTTCGGGGATACCCTTGAT
	pEVOL-Y384F-R	TATCCCCGAAGACCATGCAGGAATCGCCTAC
*Mm*PylRS-V401L	pEVOL-V401L-F	TGGAACTTTCCTCTGCACTGGTCGGACCCATACC
	pEVOL-V401L-R	TATGGGTCCGACCAGTGCAGAGGAAAGTTCCAGG
*Mm*PylRS-W417T	pEVOL-W417T-F	ATTGATAAACCCACCATAGGGGCAGGTTTCGGGC
	pEVOL-W417T-R	ACCTGCCCCTATGGTGGGTTTATCAATACCCCATTC
*Mm*PylRS-N346(small-intelligent mutation library)-C348(small-intelligent mutation library)	pEVOL-346NDT-348NDT-F	TACCATGCTGNDTTTCNDTCAGATGGGATCGGGATGCAC
	pEVOL-346NDT-348VMA-F	TACCATGCTGNDTTTCVMACAGATGGGATCGGGATGCAC
	pEVOL-346NDT-348TGG-F	TACCATGCTGNDTTTCTGGCAGATGGGATCGGGATGCAC
	pEVOL-346NDT-348ATG-F	TACCATGCTGNDTTTCATGCAGATGGGATCGGGATGCAC
	pEVOL-346VMA-348NDT-F	TACCATGCTGVMATTCNDTCAGATGGGATCGGGATGCAC
	pEVOL-346VMA-348VMA-F	TACCATGCTGVMATTCVMACAGATGGGATCGGGATGCAC
	pEVOL-346VMA-348TGG-F	TACCATGCTGVMATTCTGGCAGATGGGATCGGGATGCAC
	pEVOL-346VMA-348ATG-F	TACCATGCTGVMATTCATGCAGATGGGATCGGGATGCAC
	pEVOL-346TGG-348NDT-F	TACCATGCTGTGGTTCNDTCAGATGGGATCGGGATGCAC
	pEVOL-346TGG-348VMA-F	TACCATGCTGTGGTTCVMACAGATGGGATCGGGATGCAC
	pEVOL-346TGG-348TGG-F	TACCATGCTGTGGTTCTGGCAGATGGGATCGGGATGCAC
	pEVOL-346TGG-348ATG-F	TACCATGCTGTGGTTCATGCAGATGGGATCGGGATGCAC
	pEVOL-346ATG-348NDT-F	TACCATGCTGATGTTCNDTCAGATGGGATCGGGATGCAC
	pEVOL-346ATG-348VMA-F	TACCATGCTGATGTTCVMACAGATGGGATCGGGATGCAC
	pEVOL-346ATG-348TGG-F	TACCATGCTGATGTTCTGGCAGATGGGATCGGGATGCAC

	pEVOL-346ATG-348ATG-F	TACCATGCTGATGTTCATGCAGATGGGATCGGGATGCAC
	pEVOL-346NDT-348NDT-R	CAGCATGGTAAACTCTTCGAGGTGTTCTT
	N=A, T, C, G; D=A, T, G; V=A, C, G; M=A, C	
*Mm*PylRS-OE PCR	pEVOL-OVERLAPPING-R1	ACGCGTCGACTTACAGGTTGGTAGAAATCCCGTTATAGTAAG
	pEVOL-OVERLAPPING-F	CCGGACTAGTATGGATAAAAAACCACTAAACACTCTG

Synthetic route for *cis*-DBDAA

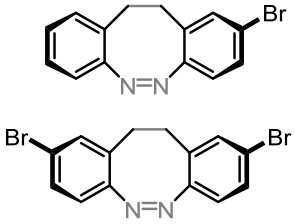


Scheme S1. Synthetic route for L-*cis*-DBDAA ncAA. r.t. = room temperature; $Pd(dba)_2$: bis(dibenzyl-ideneacetone)palladium(0); SPhos: 2-dicyclohexylphosphino-2',6'-dimethoxybiphenyl.

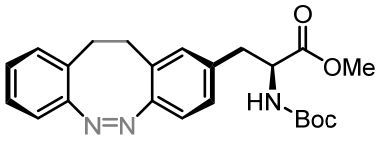


(*Z*)-11,12-dibenzo[*c,g*][1,2]diazocine (DBDA)： The zinc powder was activated with dilute hydrochloric acid to remove the oxide on the surface, then washed with distilled water and centrifuged to discard the supernatant for three times. A solution (100 mL) of barium hydroxide (10.1 g, 32.3 mmol) in water was added to the ethanol (300 mL) suspension of dinitrobiphenyl (5.0 g, 18.4 mmol) in a round bottom flask, and then added the activated zinc powder slowly (9.4 g, 143 mmol) after the flask was heated at 70 °C. The mixture

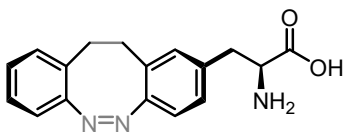
was heated to 95 °C and continuously refluxed for about 4 h. After the reaction was cooled to room temperature, introduction of CO_2 gas into the reaction mixture was to neutralize the unreacted barium hydroxide and precipitate as barium carbonate for separation. After filtration, the solvent was removed by evaporator, and then the residue was purified by flash column chromatography (eluent volume ratio: PET: EA = 5:1). Finally, white solid **S1** was precipitated by recrystallization from PET and DCM mixture, with yield of 63%. Dissolving **S1** (2.0 g, 9.42 mmol) and sodium iodide (5.70 g, 18.8 mmol) in 80 mL of anhydrous acetonitrile under protection of nitrogen atmosphere, and then adding *t*BuOCl (4.10 g, 18.8 mmol) slowly. The mixture was stirred at room temperature for 3 h. Sodium thiosulfate aqueous solution (1.0 M, 100 mL) was added to quench reaction, then extract the mixture with EA (50.0 mL × 3) and the organic phase was dried with anhydrous magnesium sulfate. The solvent was removed by evaporation, and purified by flash chromatography (eluent volume ratio: PET: EA = 10:1), and vacuum dried to obtain a yellow solid with yield of 70% (1.40 g). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.17-7.09 (m, 2H), 7.04-6.94 (m, 4H), 6.82 (d, J = 7.8 Hz, 2H), 3.04 - 2.70 (m, 4H). MS (ESI) calcd. for $C_{14}H_{13}N_2^+$ 209.11 [M+H$^+$], found 209.19.



(*Z*)-2-bromo-11,12-dibenzo[*c,g*][1,2]diazocine: DBDA (500 mg, 2.38 mmol) and NBS (322 mg, 1.81 mmol) were added in 5 mL of acetic acid with a small amount of concentrated sulfuric acid, and the mixture was stirred at room temperature for 10 min. Then, the reaction mixture was immediately quenched with 10% sodium thiosulfate aqueous solution, and then was extracted with DCM (50.0 mL × 3). The organic phase was dried with anhydrous magnesium sulfate. After purification by column chromatography (eluent volume ratio: PET: EA = 10:1), a yellow oil mixture was obtained by vacuum drying. Through HPLC-MS analysis for the unseparated products, the mixture contained target product, DBDA-Br, >48% yield, dibromo product, DBDA-2Br, <12%.[2] The mixture was used in further steps without additional purification.



***N*-Boc-DBDA-Ala-OMe:** Dry a round bottom flask, a stirring bar and a Schlenk tube before using. First, zinc dust (818 mg, 12.93 mmol) was added to the round bottom flask with ultra-dry DMF solvent (3.0 mL), and the mixture was stirred at low speed (300 rpm), followed by adding a catalytic amount of iodine (172 mg, 0.65 mmol) to activate the zinc dust. After the solution became colorless, *N*-Boc-iodoalanine methyl ester (1.42 g, 4.31 mmol) was added immediately, followed by adding an additionally catalytic amount of iodine (172 mg, 0.65 mmol). After the solution became colorless again, the mixture was continuously stirred at room temperature for 10 minutes to finish preparation of the organozinc reagent. $Pd(dba)_2$ (49.6 mg, 0.09 mmol), Sphos (35.3 mg, 0.09 mmol), **DBDA-Br** (500 mg, 1.72 mmol) and organozinc reagent was added to the Schlenk tube sequentially with additional ultra-dry DMF solvent (2.0 mL), and stirred at 70 ℃ for 12 hours under nitrogen atmosphere. After cooling to room temperature, the reaction mixture was quenched by water (50.0 mL) and extracted with EA (50.0 mL × 3). The organic phase was dried with anhydrous magnesium sulfate. The solvent was removed by evaporation, and purified by flash chromatography (eluent volume ratio: PET: EA = 2:1), then vacuum dried to obtain a yellow oil with a yield of 60% (423 mg). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.12 (td, *J* = 7.4 1.7 Hz, 1H), 7.04 - 6.94 (m, 2H), 6.92 - 6.84 (m, 1H), 6.81 (dd, *J* = 7.8, 1.2 Hz, 1H), 6.73 (d, *J* = 18.8 Hz, 2H), 4.96 - 4.83 (m, 1H), 4.46 (q, *J* = 6.8 Hz, 1H), 3.52 (s, 3H), 3.01-2.83 (m, 4H), 2.73 (d, *J* = 9.2 Hz, 1H), 1.41 (s, 9H). ^{13}C NMR (101 MHz, Chloroform-*d*) δ 171.12, 154.48, 153.92, 153.37, 133.92, 129.51, 128.52, 127.27, 126.99, 126.45, 126.01, 125.68, 118.07, 117.74, 78.99, 53.31, 51.05, 36.97, 30.81, 30.53, 28.68, 27.28, 17.42. HRMS (ESI) calcd. for $C_{23}H_{28}N_3O_4^+$ 410.2080 [M+H$^+$], found 410.2059.



***cis*-DBDAA:** Dissolving ***N*-Boc-DBDA-Ala-OMe** (600 mg, 1.46 mmol) in a mixed solvent system of methanol: tetrahydrofuran: water = 1:1:1, then 8 eq. $LiOH{\cdot}H_2O$ (468 mg, 11.7 mmol) was added to the mixture and stirred at room temperature for 6 h. The hydrolysis process was tracked and detected by TLC analysis until completed consumption of the starting material. Then, diluted hydrochloric acid (1.0 M) was added to neutralize the pH, then the mixture was extracted with EA (20.0 mL × 3). The organic phase was dried with anhydrous magnesium sulfate, and then evaporated to remove the solvent for consecutive deprotection of the amine. The resulting residue was dissolved in 1.0 mL of DCM with 1.0 mL of trifluoroacetic acid, and stirred at room temperature for 10 minutes. And then the resulting reaction mixture was blown with nitrogen flow to get rid of the TFA. The crude product was then dissolved with 1.0 mL DMSO and purified by reverse preparative on a C18 chromatography with 5% water and 95% acetonitrile as the mobile phase. The collected fractions were evaporated to remove the solvent to obtain the final ***cis*-DBDAA** product as a yellow solid (293 mg, yield 68%). ^{1}H NMR (600 MHz, Deuterium Oxide) δ 7.14 - 7.10 (m, 1H), 7.05 - 7.02 (m, 2H), 6.96 (td, J = 6.0 Hz, 1H), 6.88 (d, J = 4.0 Hz, 1H), 6.81 (d, J = 7.8 Hz, 1H), 6.78 - 6.74 (m, 1H), 3.26 (dd, J = 7.8, 5.4 Hz, 1H), 2.84 - 2.70 (m, 5H), 2.54 (dd, J = 13.6, 7.8 Hz, 1H).^{13}C NMR (101 MHz, Deuterium Oxide) δ 181.94, 153.82, 152.29, 138.28, 130.72, 129.81, 128.90, 127.99, 127.65, 126.85, 119.00, 118.60, 57.21, 40.26, 38.72, 31.00, 30.63. HRMS (ESI) calcd. for $C_{17}H_{17}N_3O_2^+$ 294.1248 [M-H^+], found 294.1257.

Examination of the photo-switching properties of DBDAA in aqueous solution

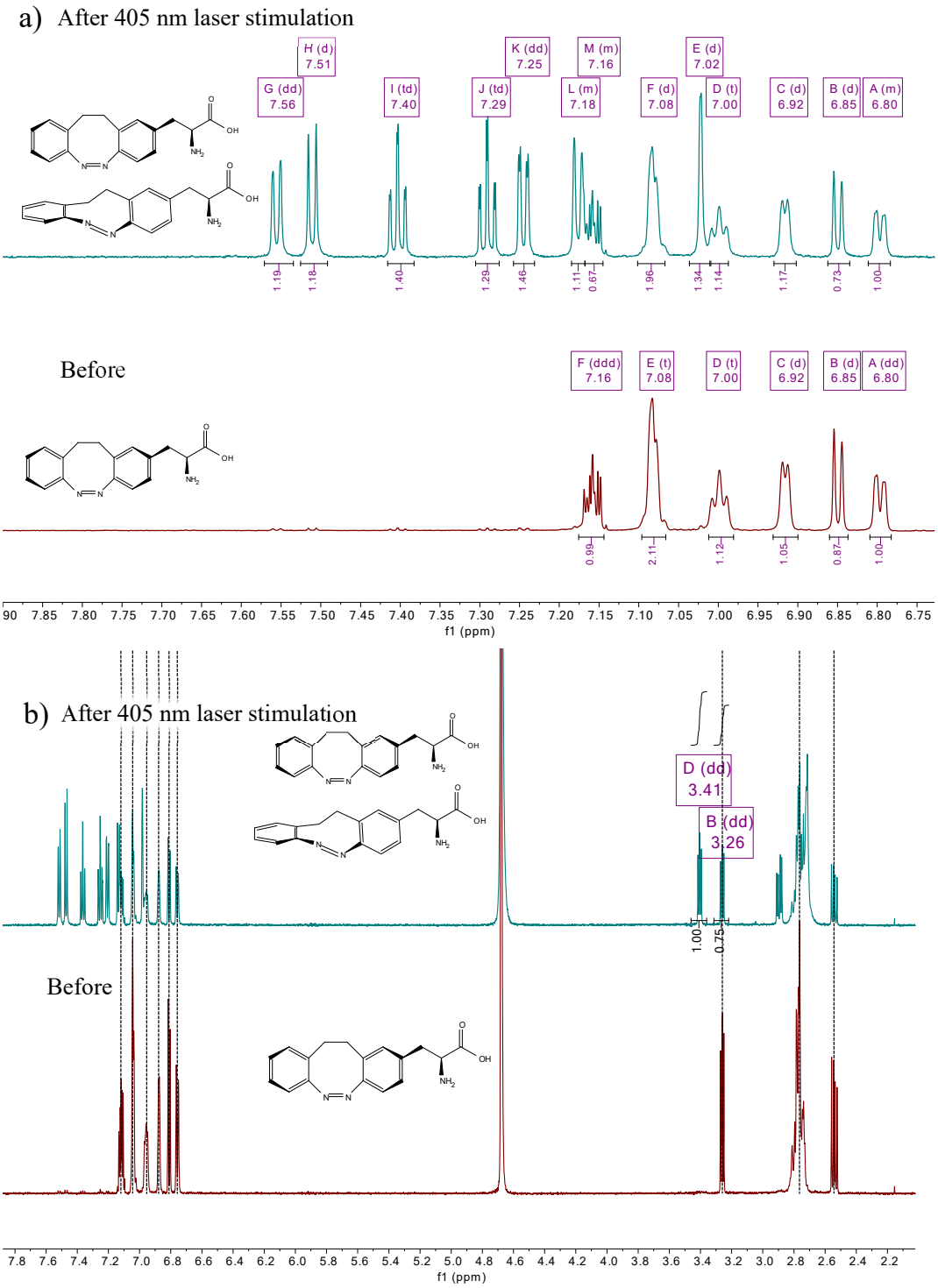


Figure S1. The photo-stationary state of DBDAA. a) 1.0 mM in D_2O with 2% NaOH, ^1H 800 MHz, 298 K, before and after 20W 405 nm LED irradiation for 2 min at 298 K.

power density ≈ 13.4 mW·cm^{-2}. PSS_{405} = *Z*:*E* = 1:1.34. b) 3.0 mM in D_2O with 2% NaOH, ^1H 600 MHz, 298 K, before and after 405 nm laser irradiation for 2 min at 298 K. 405 nm laser power density = 500 mW·cm^{-2}. PSS_{405} = *Z*:*E* = 3:4 = 1.33.

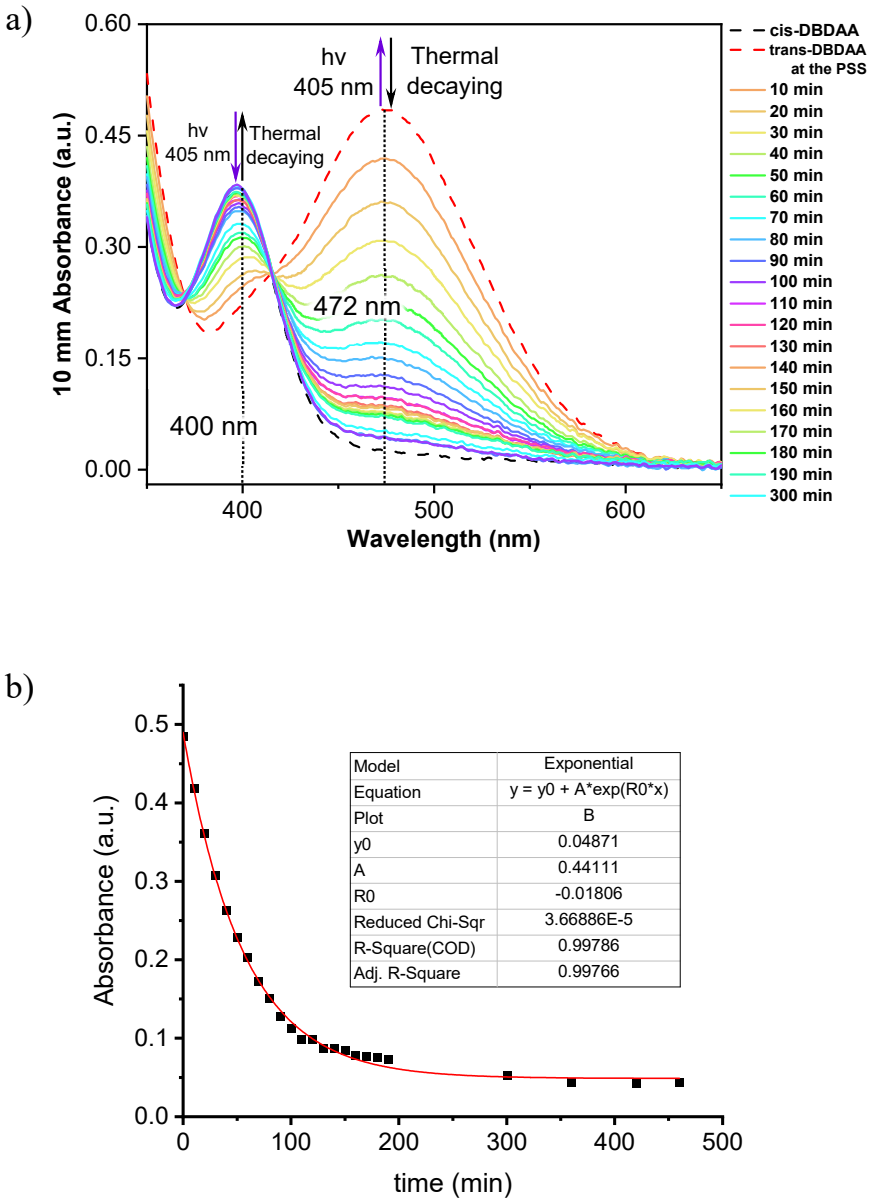


Figure S2. Absorption spectra (a) top panel) and time-dependent absorbance at 472 nm (b) lower panel) for 100 µM DBDAA in H_2O after 405 nm photo-stimulation for 2.0 min.

Figure note for Fig. S2: As it can be seen in the top panel of Fig. S2, a characteristic absorption peak around 400 nm (λ_{max} = 398 nm) of *cis*-DBDAA (the black dotted line) was observed, which could be assigned as its n → π* transition excitation. After continuous 405 nm photo-stimulation for 2 min, the photo-switching of *cis*-DBDAA to *trans*-DBDAA reached a photo-stationary state (PSS) in which the ratio of R*t* (*trans*-DBDAA/total DBDAA) is a constant. At the PSS, the dynamic equilibrium system is a mixture of *cis*-DBDAA and *trans*-DBDAA (R*t*= *t*DBDAA/total DBDAA = 4/7 = 57.1% at the PSS_{405nm}) determined based on the ^{1}H NMR integration of each isomer in Figure S1 (assuming the PSSs are the same under both conditions). At the PSS, a new distinctive absorption band around 472 nm appeared (the red dotted line), which could be assigned as the n → π* transition of *trans*-DBDAA. At the same time, the absorption peak at 400 nm decreased significantly, implying a high conversion of *cis*- to *trans*-isomer induced by the 405 nm LED.

By tracking the time-dependent absorption after the 405 nm photo-stimulation in dark, the absorbance at 472 nm descended gradually while the absorbance at 400 nm reverted close to the original value (the colour-coded solid lines in Fig. S2a), which suggested a thermal isomerization from *trans*- to *cis*-DBDAA. The decaying rate was then obtained (lower panel, $t_{1/2}$ = 38.5 min) by plotting the characteristic absorbance of *trans*-DBDAA at 472 nm versus time and by fitting the curve with an exponential decay function. The photo-isomerization from *cis*- to *trans*-isomer induced by the 405 nm stimulation. On the other hand, the isomerization from *trans*- to *cis*-isomer can be promoted by 532 nm stimulation, which leads to complete back-switching with the R*t* ≈ 0 at PSS_{532nm}.

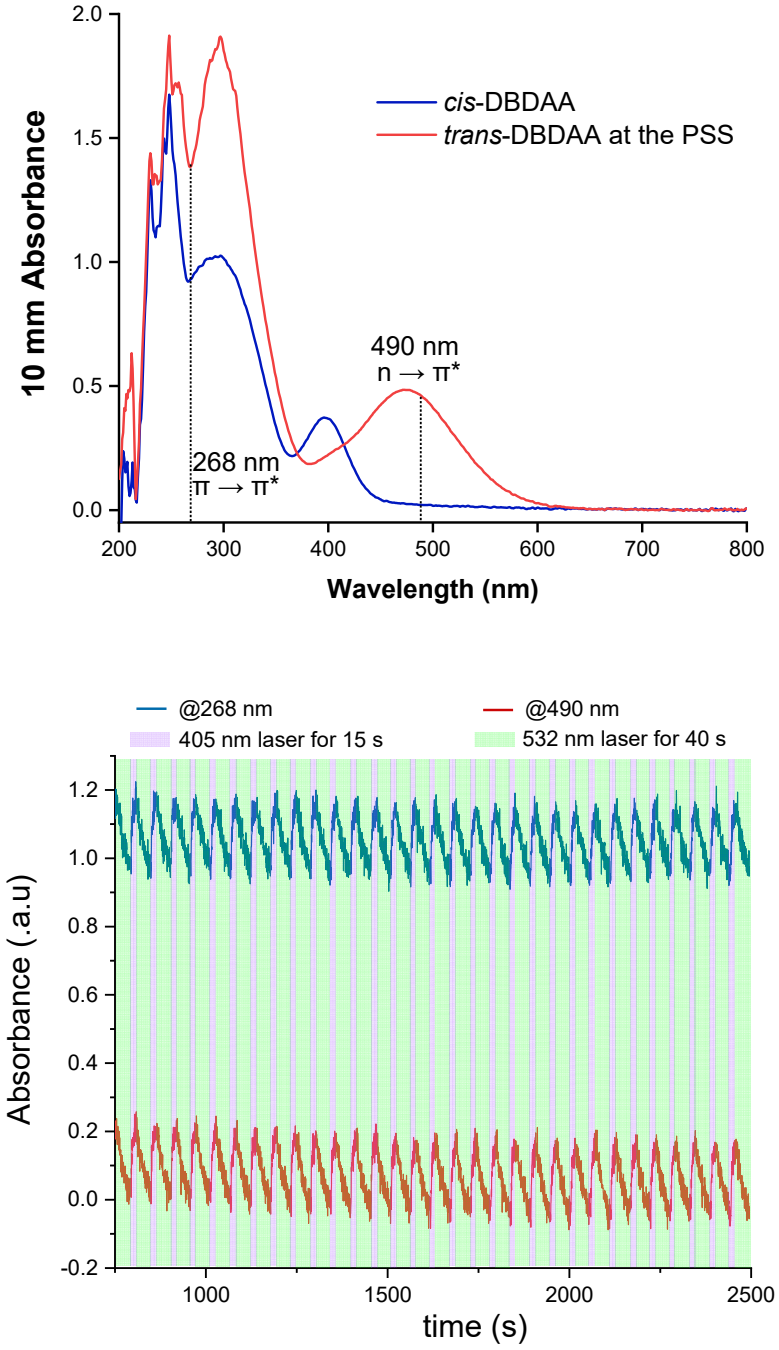


Figure S3. Top panel: the characteristic π → π* (268 nm) and n → π* (490 nm) transition of *cis*-DBDAA and *trans*-DBDAA at the PSS. Lower panel: real-time tracking of the absorption evolution at 268 nm (blue line) and 490 nm (red line) for 100 μM DBDAA in H_2O. No significant decay of the absorbance change was observed after

32 rounds of photo-switching cycles. 405 nm laser (15 s in each cycle, 500 mW·cm^{-2}) and 532 nm laser (40 s in each cycle, 100 mW·cm^{-2}) are used as excitation light sources.

X-ray crystal diffraction for the structure elucidation of *cis*-DBTD, *cis*-DBDA and *trans*-DBDA

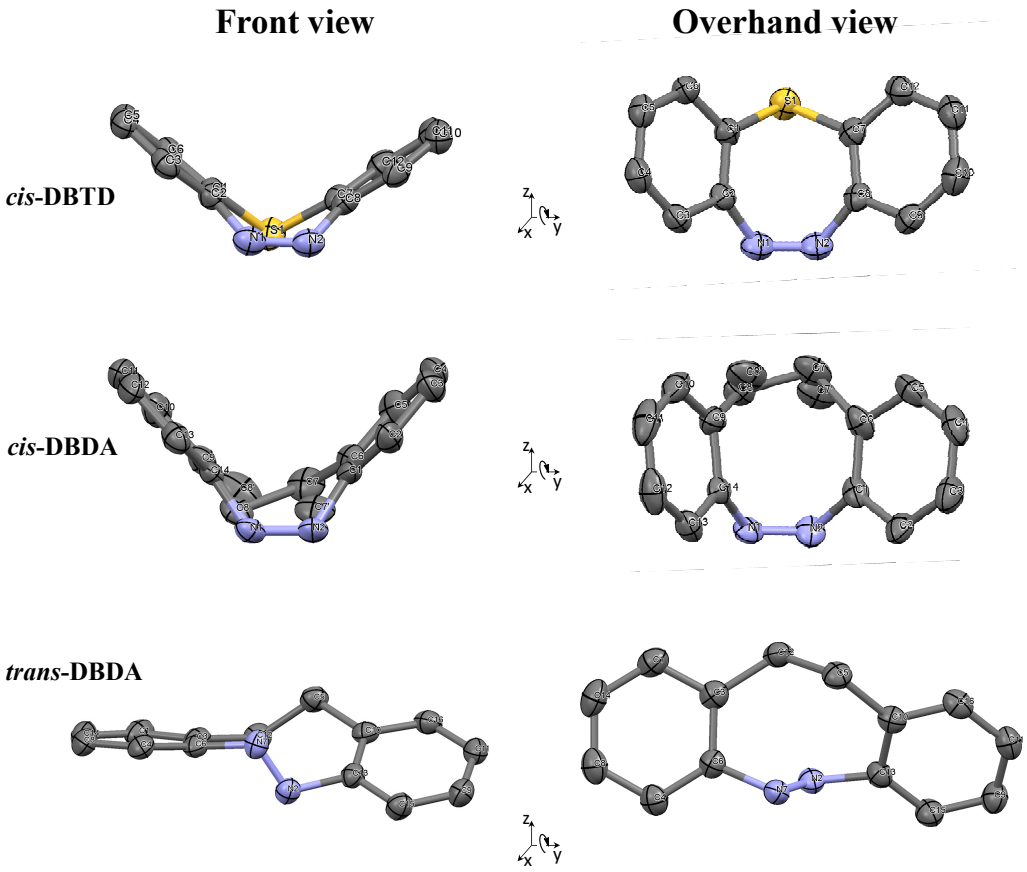


Figure S4. X-ray crystal structure of *cis*-DBTD (CCDC: 1905279)[3], *cis*-DBDA (CCDC: 757546)[4] and *trans*-DBDA (CCDC: 1992126).[5] The similar "bowl-shape" morphology of *cis*-DBTD and *cis*-DBDA provided us the rationale to choose the mutation sites and to construct the DNA library based on previously trported *Mm*DBTDARS DNA library[1] for evolving the *Mm*PylRS to recognize *cis*-DBDAA specifically.

The DNA library construction on the *Mm*PylRS and the selection procedures for the directed evolution

For the first step in the directed evolution of PylRS, selection of the bind sites in the enzymatic pocket for library construction is based on the reported *p*BpaRS mutant (for incorporation of a photo-cross-linker ncAA, *p*-benzoyl-L-phenylalanine, *p*Bpa[6]) and the DBTDARS mutant (for incorporation of a seven-membered cyclic azobenzene ncAA, dibenzo[*b*,*f*][1,4,5]thiadiazepine–alanine, DBTDA[1]). The chemical structure of *p*Bpa and *cis*-DBTDA (bowl-shaped and rigid structure) has a great similarity to *cis*-DBDAA (also bowl-shaped and rigid structure) in which the two benzene rings are connected by a linker with a relatively large size and the two benzene rings are folded with each other (Fig. S4). In the directed evolution of *Mm*PylRS for *p*Bpa and *cis*-DBTDA, it is considered that the N346 and C348 residues on the bottom region of wild-type PylRS (Fig. S5) produce steric clashes to the bowl-shaped ncAA substrate, thus affecting the recognition. Therefore, the codons at both the 346 and 348 site of *Mm*PylRS were chosen to construct a "small-intelligent" DNA library to avoid redundant codons and stop codons.

In the site-saturation mutagenesis, the NNK (N=A, T, C, G and K=G, T, a more restricted degenerate codons) codon library is often used to encode all 20 natural amino acids and 1 stop codon (TAG). However, the distribution of each amino acid in the NNK library is unequal due to the codon redundancy. By analyzing the occupancy probability of each amino acid in the NNK library, Jiang group further minimized the NNK library into a "novel" small-intelligent library[7] without inherent amino acid biases, stop codons, or rare codons in *E. coli* system. In detail, the site-saturation mutagenesis was generated by using a primer mixture containing four types of small randomized codon library or fixed codons on designate sites, including NDT, VMA (N=A, T, C, G; D=A, T, G; V=A, C, G; M=A, C), ATG, and TGG. In the primer mixture, the four types of small randomized codons were also combined with an appropriate stoichiometry (NDT: VMA: ATG: TGG=12: 6: 1: 1) to give a complete small-intelligent library. As a

result, 20 natural amino acids were correlated with 20 degenerate codons, which could eliminate redundant codons and achieve even distribution of each amino acid, thus avoiding the loss of library richness during selection. In our case, the residues, N346 and C348, were chosen to randomize as two small-intelligent libraries simultaneously in the construction. Firstly, the primers in the Table S1 (Primers used for plasmids construction, on page 6-7 in the ESI) were synthesized commercially and mixed with the appropriate stoichiometry. Then, we were able to perform an overlap extension PCR to construct a full plasmid library with both the N346 and C348 sites as small-intelligent libraries.

As for those fixed mutation sites, the distal benzene ring in *cis*-DBDAA was expected could have a lone pair (LP) – π interaction with the oxygen atom at W417T site to assist the binding, which is akin to the DBTDARS-DBTDA pair.[1] Therefore, the mutation on W417T was retained from previously discovery. On the other hand, the N346 site plays an important role in the aminoacylation of the $tRNA^{Pyl}$ toward the cognate amino acid.[8] Therefore, the fixed mutation on A302T site was to rebuild a strong hydrogen-bonding with the amino acid moiety of *cis*-DBDAA to compensate for the loss of hydrogen-bonding donor on the N346 mutations. Furthermore, the fixed mutation on Y384F is capable of accelerating the process of aminoacylation of the $tRNA^{Pyl}$ according to previous report[9] (Figure S5). Lastly, the V401 site is located in the open front of the active cavity of the *Mm*PylRS (Figure S5). If the size of the mutated residue at V401 site is properly tuned while keeping as a hydrophobic residue, it might be helpful to lock the *cis*-DBDAA substrate inside the binding pocket. As a result, the V401 site is site-directedly mutated as either V or L, which were then mixed as a two-amino-acid randomized codon.

Collectively, the site-directed mutagenesis was performed on A302T, Y384F, W417T sites. And V401 was randomly mutated as V or L residue in combination with a small-intelligent mutation library randomized on both 346 and 348 sites.

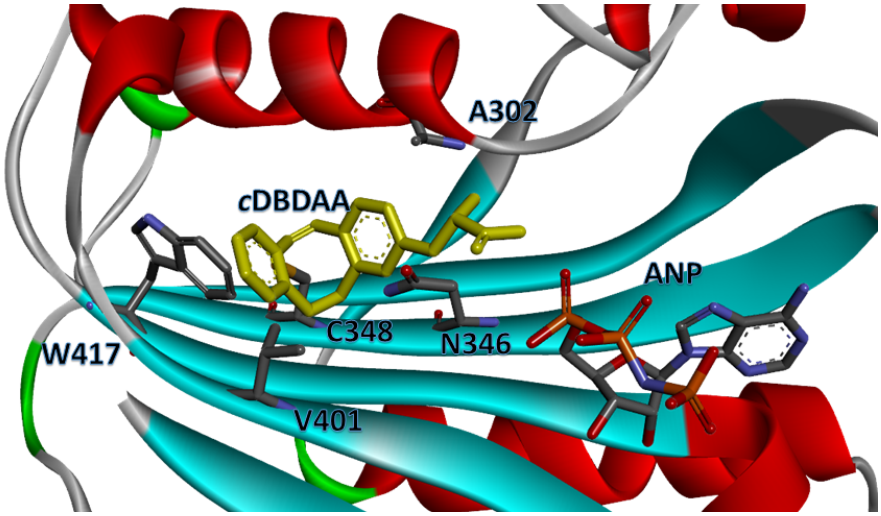


Figure S5. Preliminary prediction for the complex between *Mm*PylRS and *c*DBDAA (yellow small molecule) to understand the potential interactions or steric clashes. The key residues for the fixed mutation or DNA library construction and ANP (phosphoaminophosphonic acid-adenylate ester) were displayed.

In the direct evolution of *Mm*PylRS synthetase, the positive selection is based on the chloramphenicol-resistance gene (CmR), an antibiotic-based selection. The positive selection plasmid (named as *p*Evol-*Mm*PylRS-PylT-CmR-TcR) is constructed by ourselves based on the *p*Evol-*Mm*PylRS-PylT plasmid, which contains two antibiotics resistance gene, the chloramphenicol-resistance gene (CmR) and the tetracycline-resistance gene (TcR), and the *Mm*PylRS gene with the DNA library on the specified sites. The TcR gene was included for exclusive survival of the transformed *E. coli* strain (Top10 competent cells) in the culture medium against environment contamination, which is native and without any TAG mutation. On the other hand, the chloramphenicol-resistance (CmR) gene bears an amber stop codon on Q98TAG was subcloned into the *p*Evol-*Mm*PylRS-PylT plasmid to achieve the antibiotic-based selection purpose. The *p*Evol-*Mm*PylRS-PylT-CmR-TcR plasmid library were transformed into self-made Top10 competent cells that harbors the plasmid, *p*BAD-sfGFP-Q204TAG with an ampicillin resistance gene (AmpR) and a fluorescent reporter protein gene. Colonies were grown on the GMML agar in Petri dishes supplemented with tetracycline (34 μg•mL^{-1}), ampicillin (100 μg•mL^{-1}), chloramphenicol (40 μg•mL^{-1}), L-arabinose (0.2%) and 2.0 mM *c*DBDAA for 72 h. In the presence of 40 μg•mL^{-1}

chloramphenicol as a screening pressure, survival of Top10 cells relies on the full-length expression of the chloramphenicol acetyltransferase gene (CmR) by suppressing the amber stop codon. As a result, *Mm*PylRS mutants capable of incorporating either natural amino acid or *cis*-DBDAA should then readthrough the amber stop codon to express the full-length chloramphenicol acetyltransferase, making the host *E. coli* cells resistance to the growth inhibition by chloramphenicol. The *p*BAD-sfGFP-Q204TAG was used as a secondary selection factor that the survived cells are able to co-express the full-length sfGFP as a fluorescence indication. During the selection process, growth of the Top10 cells which were unable to suppress the amber stop codon will be eliminated. After the first positive selection, all colonies were collected, and the plasmids within the colonies were recovered for sequencing.

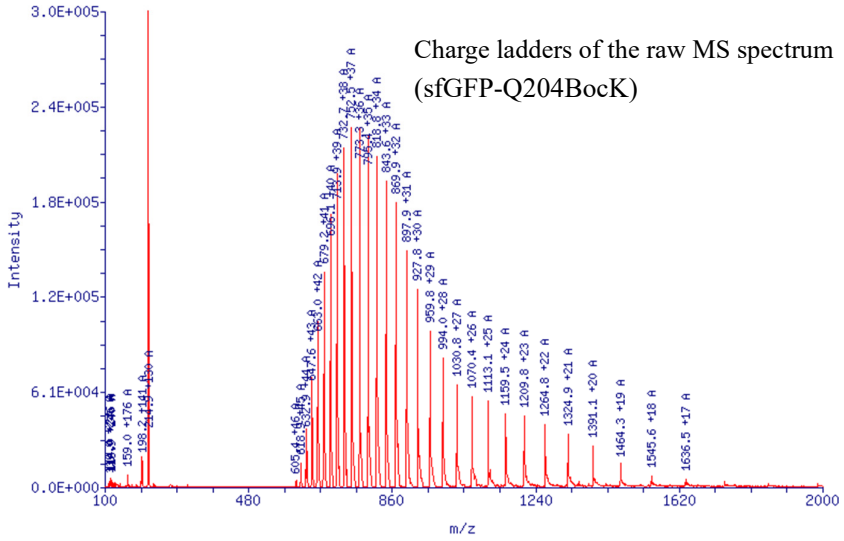
Next, the collected plasmid library was used for the first negative selection. The negative selection depends on the full-length expression of a toxic RNase gene (barnase) to eliminate the cells containing *Mm*PylRS mutants which mis-incorporate natural amino acids to suppress the amber stop codon. Accordingly, the recycled plasmid library was transformed into another self-made Top10 competent cells harboring *p*YSneghc (constructed by ourselves) plasmid which contains the gene of the toxic barnase (barnase-Q3TAG-G66TAG, bactericidal to *E. coli*). Then, colonies were grown on LB agar in Petri dishes supplemented with tetracycline (34 $\mu g \cdot mL^{-1}$), ampicillin (100 $\mu g \cdot mL^{-1}$), *L*-arabinose (0.2%). Mis-incorporation of natural amino acids resulted in full-length expression of the toxic barnase RNase, leading to the survival and propagation of cells containing PylRS mutants that exclusively recognize *cis*-DBDAA. Then, the recovered and sequenced plasmid library was used for next round of selection. The positive and negative selection processes were cycled and iteratively repeated with the concentration of chloramphenicol or *L*-arabinose gradually increased until a few consensuses DNA sequence from the library was obtained. The higher concentration of chloramphenicol or *L*-arabinose applied a higher selection pressure for higher ncAA fidelity of the *Mm*PylRS mutant. After sequencing, the identified plasmid was subjected to protein expression verification at a relatively large scale (50 mL).

Protein expression and purification

For protein over-expression, the plasmid pET-sfGFP-Q204TAG-His_6 harboring a sfGFP gene with a hexa-histidine-tag on *C*-terminal for purification and pEvol-*Mmc*DBDAARS-PylT plasmid for genetically encoding *cis*-DBDAA were co-transformed into *E. coli* BL21 (DE3) competent cells. Transformed bacterial cells were grown on LB agar medium in Petri dishes supplemented with tetracycline (34 $\mu g \cdot mL^{-1}$) and ampicillin (100 $\mu g \cdot mL^{-1}$) overnight. Bacterial colonies were picked randomly and inoculated in 5.0 mL LB medium supplemented with tetracycline (34 $\mu g \cdot mL^{-1}$) and ampicillin (100 $\mu g \cdot mL^{-1}$) for subsequent DNA sequencing. The correctly sequenced strains were stored as a glycerol stock of *E. coli* BL21 (DE3) in -80 °C freezer. The glycerol stock of *E. coli* BL21 (DE3) was then inoculated in LB medium with tetracycline (34 $\mu g \cdot mL^{-1}$) and ampicillin (100 $\mu g \cdot mL^{-1}$) and cultured overnight at 37 °C, 280 rpm. 1.0 mL of the pre-culture was transferred into 50 mL fresh LB medium with tetracycline (34 $\mu g \cdot mL^{-1}$) and ampicillin (100 $\mu g \cdot mL^{-1}$), then cultured at 37 °C, 280 rpm. Until OD600 reaching around 0.6, 2.0 mM *cis*-DBDAA, 1.0 mM IPTG and 0.2% arabinose were added to the culture system to induce recombinant protein over expression for 8-12 h at 37 ºC, 250 rpm. Then, the culture was centrifuged at 4 ºC for 20 min to collect bacterial pellet, and then the pellet was resuspended in a lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH = 8.0). After 30 times ultrasonic fragmentation (2 s each with 6 s interval), the cell lysates were centrifuged (30 min, 12,000 g, 4 ºC) to get rid of insoluble debris. Then His-tagged proteins in cell lysate were purified by Ni-NTA affinity column (HisPur™ Ni-NTA Resin, Thermo Scientific) according to manufacturer's protocol (2 h, 4 ºC). The protein-bound resins were washed with the lysis buffer and then a washing buffer (300 mM NaCl, 50 mM Na_2HPO_4, 50 mM imidazole, pH = 8.0) for three times or more to get rid of other bound proteins. The His_6-tagged proteins were exchanged by an elution buffer (300 mM NaCl, 50 mM Na_2HPO_4, 250 mM imidazole, pH = 8.0) and collected in fractions. Then, the eluted fractions were transferred into the Amicon Ultra-15 Centrifugal Filter (10 k MWCO, Millipore) for buffer exchanging to PBS buffer (pH = 7.4). Then the purified

proteins were analyzed by deconvoluted ESI-MS verification as well as SDS-PAGE to check purity, and further subjected to in-gel fluorescence imaging followed by Coomassie blue staining. The concentration of over-expressed proteins was determined by measuring the absorbance at 485 nm (Thermo Nanodrop 2000C), which was reported in previous literature.[10] sfGFP-N149TAG-His_6 was also over-expressed and purified through the same procedures.

LC-MS analysis of sfGFP-Q204*c*DBDAA



Charge ladders of the raw MS spectrum (sfGFP-Q204BocK)

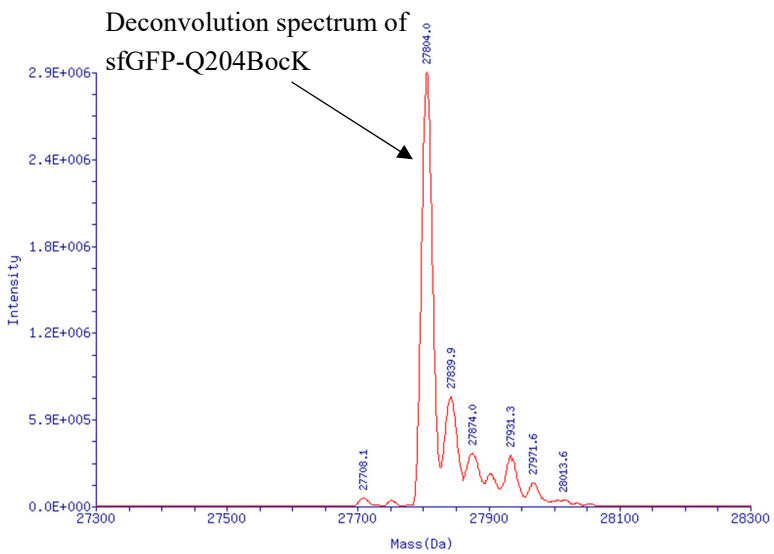
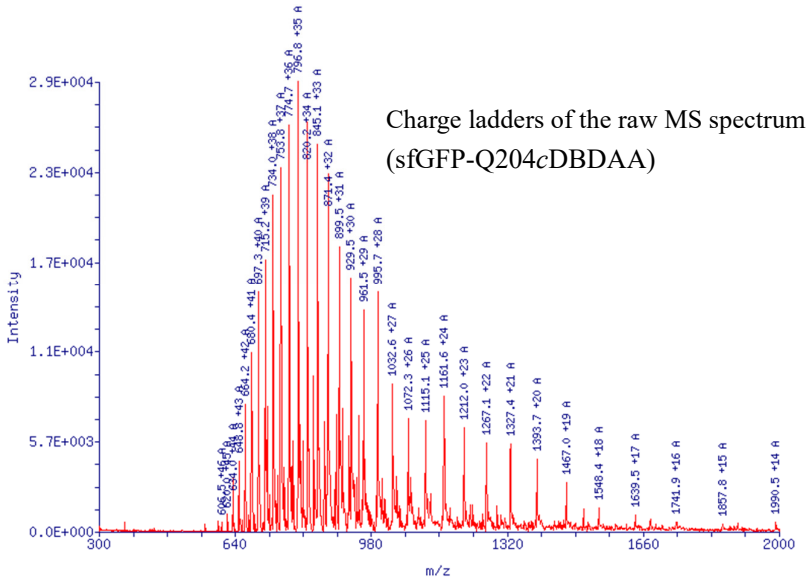


Figure S6. Deconvoluted LC-MS analysis for sfGFP-Q204BocK with the charge ladder and the deconvoluted mass spectrum shown. Found [M-Met+H$^+$]: 27804.0 Da



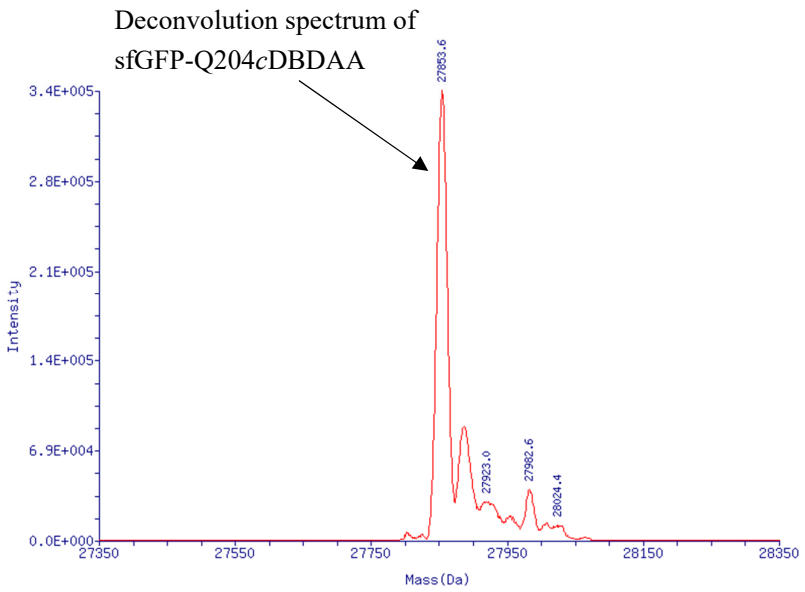


Figure S7. Deconvoluted LC-MS analysis for sfGFP-Q204*c*DBDAA with the charge ladder and the deconvoluted mass spectrum shown. Found [M-Met+H^+]: 27853.6 Da. Calcd. mass 27853.0 Da.

The incorporation efficiency of the *Mmc*DBDAARS and the opti-*Mmc*DBDAARS

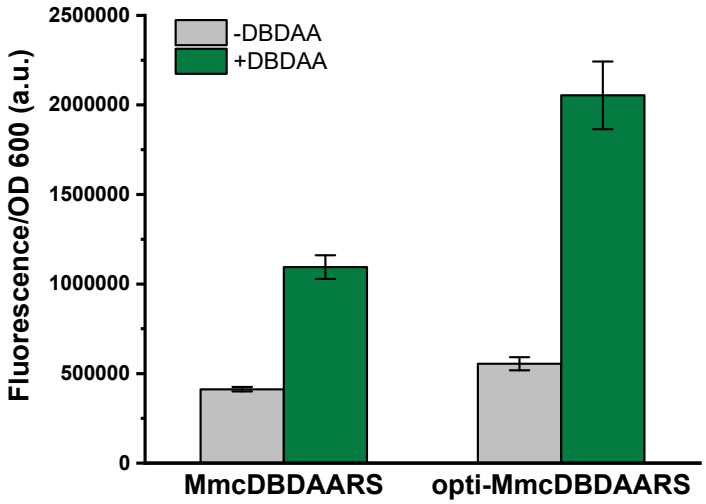


Figure S8. Fluorescence/OD600 screening toward the *Mmc*DBDAARS and the opti-*Mmc*DBDAARS for their incorporation efficiency of *cis*-DBDAA (2.0 mM). HORIBA Fluoromax-4 Spectrofluorometer was used to evaluate the fluorescence intensity of sfGFP-Q204DBDAA inside the living *E. coli* cells and Thermo Nanodrop 2000C for determining OD600 to undertand the cell density in the samples. Error bars represent three independent expressions.

The moleculer docking of *cis*-DBDAA to the evolved *Mmc*DBDAARS

The crystal structure of *O*-methyl-L-tyrosine-*Mm*OmeRS (PDB: 3QTC) was acquired from the Protein Data Bank (RCSB). And then replacing with the mutated amino acid (A302T, N346A, C348G, Y384F, V401L and W417T) in the corresponding residues on the *Mm*OmeRS template to create the *Mmc*DBDAARS. Before docking, water molecules, solvent moleclues and the original ncAA ligand were removed from the crystal structure. The optimal configuration and conformation of *cis*-DBDAA was obtained from the crystal data and was optimized by DFT calculation via the Gaussian 09 software (calculated at b3lyp-D3/6-311g** level of thoery), and further optimized to a minimized ligand energy by applying a CHARMM force field. The effective binding sites were defined within a 5 Å spherical space covering all of the active residues to furnish a *Mm*cDBDARS-*c*DBDAA complex. Then, we used the CDocker algorithm via the Discovery Studio to search for a series of plausible docking results with the interaction scoring and the interaction energy ranked.

	Index	Name	Visible	Tagged	Visibility Locked	Resolution	Forcefield	-CDOCKER_ENERGY	-CDOCKER_INTERACTION_ENERGY
1	1	3QTC	☑ Yes	☐ No	☑ Yes	1.75	CHARMm		
2	2	DBDAA	☐ No	☐ No	☐ No			28.7185	41.3633
3	3	DBDAA	☐ No	☐ No	☐ No			28.3893	41.859
4	4	DBDAA	☐ No	☐ No	☐ No			27.8376	40.7048
5	5	DBDAA	☐ No	☐ No	☐ No			25.2663	40.74
6	6	DBDAA	☐ No	☐ No	☐ No			25.2017	38.8099
7	7	DBDAA	☐ No	☐ No	☐ No			25.0654	40.6392
8	8	DBDAA	☐ No	☐ No	☐ No			24.7362	40.4167
9	9	DBDAA	☐ No	☐ No	☐ No			24.357	39.1571
10	10	DBDAA	☐ No	☐ No	☐ No			24.1049	38.8906
11	11	DBDAA	☐ No	☐ No	☐ No			24.0356	37.9519

Figure S9. The docking results with the scoring, CDocker energy and interaction energies of *cis*-DBDAA binding to the *Mmc*DBDAARS are shown.

Figure note for figure S9: The results showed that the *cis*-DBDAA moiety was accommodated in a hydrophobic cavity surrounded by L305, L309, V401L and I405 residues with an optimal binding energy (-41.9 kcal·mol^{-1}, Fig. S10-11, ESI†). The carboxylate of *cis*-DBDAA was orientating towards the hydroxyl of A302T (Fig. S10a-b) in a similar way as *Mm*OMeRS[8] and *Mm*PSCaaRS,[11] forming strong hydrogen-bonds with both A302T and -NHCO- on L301. Besides, this binding posture also promotes an additional sulfur-oxygen interaction with the M344. Thus, the triple interactions together compensate the evolved N346A mutation for important aminoacylation (Fig. S10a). Meanwhile, small-sized residual mutations, e.g., N346A and C348G, provided sufficient space to weaken steric clash between the "bowl" shaped *cis*-DBDAA and "bottom" of the binding pocket. Additionally, an alkyl-π and a n-π interactions on proximal and distal benzenes of *cis*-DBDAA were found, respectively, via V401L and W417T mutations, which constrains the tilted DBDA moiety in proper posture and distinguishes from other amino acids (Fig. S10b). These unique interactions together are indispensable for successful anchoring of *cis*-DBDAA in *Mmc*DBDAARS that inspires evolution for encoding *trans*-DBDAA after photo-isomerization.

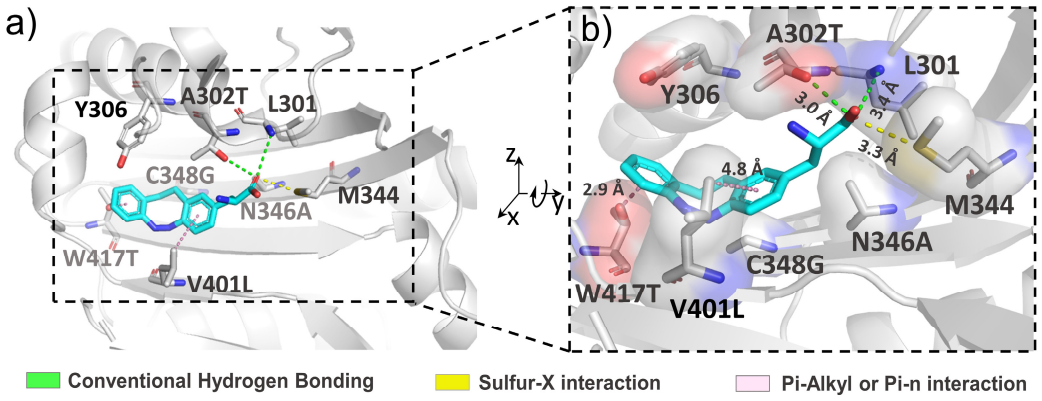


Figure S10. The simulative binding pattern of *cis*-DBDAA into the active pocket of *Mmc*DBDAARS. a) 3D display of the docking posture of *cis*-DBDAA in the optimal binding energy from an overhand view. Key residues and corresponding interactions were denoted. b) The docking posture and interactions shown in a magnified front view.

Hydrogen-bonding (green): L301 and A302T; sulfur-oxygen interaction (yellow): M344; alkyl-π & n-π interactions (lavender): V401L and W417T.

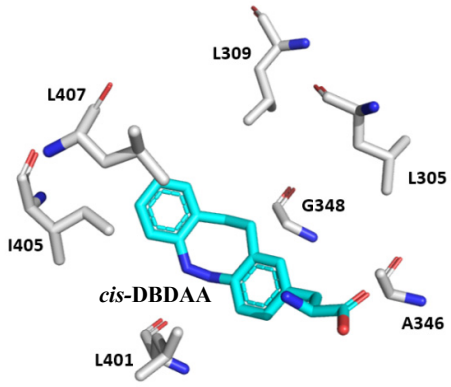


Figure S11. The hydrophobic residues (L305, L309, N346A, C348G, V401L, I405 and L407) surrounding in active binding pocket of the evolved *Mmc*DBDAARS, and the posture of *cis*-DBDAA in the binding pocket represented. The residues (gray) and *cis*-DBDAA (cyan) are shown and labelled in stick displaying mode.

Applying the evolved *Mmc*DBDAARS on GCE system in mammalian cells

To extend the evolved *Mmc*DBDAARS GCE system into mammalian cells, an enhanced green fluorescent protein (EGFP) was chosen for fluorescence detection with high sensitivity. Therefore, a plasmid *p*EM-14E,[12] bearing a codon optimized *Mmc*DBDAARS-$tRNA^{Pyl}_{CUA}$ pair and the EGFP-N149TAG gene with an HA-tag on *C*-terminal for immunoassay, was transfected into human embryonic kidney 293T (HEK293T) cells (Fig. S12a). Transfections without *p*EM-14E plasmid was used as one of the negative controls (Fig. S12b-c). As expected, only cells, both transfected with *p*EM-14E and incubated with *cis*-DBDAA (2.0 mM) as cognate ncAA, showed strong green fluorescent signal, implying a successful GCE (Fig. S12b). And the nonsense suppression efficiency is as 7.4% of WT reporter (Fig. S13-14). Western blot analysis

further verified that a band of EGFP-N149*c*DBDAA-HA was immuno-recognized by an anti-HA antibody, indicating a full-length GCE expression in the presence of *cis*-DBDAA (Fig. S12c). Treatment of 2.0 mM *cis*-DBDAA to living HEK293T cells for 30 h did not cause observable cytotoxicity, tested by CCK8 assay (Fig. S15)

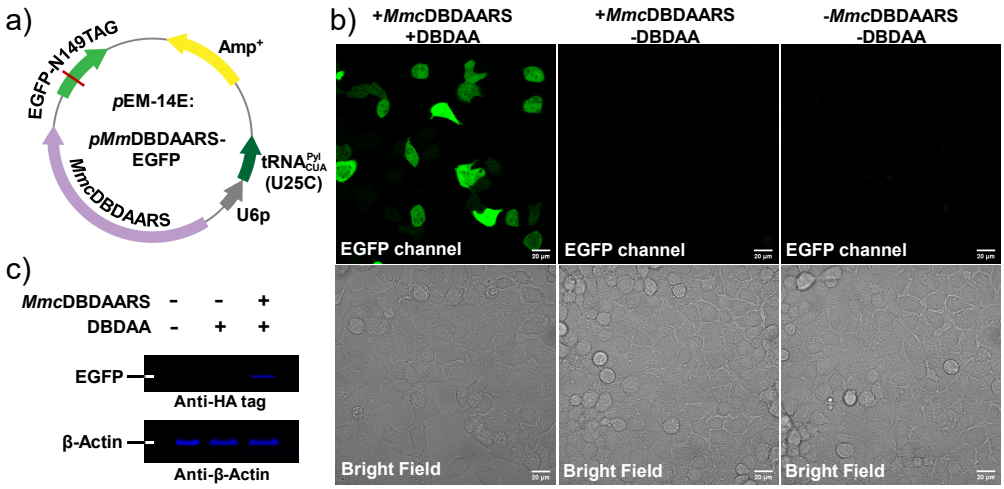


Figure S12. The GCE protein expression by applying *Mmc*DBDAARS-tRNA^{Pyl} pair to encode *cis*-DBDAA into EGFP in living HEK293T cells. a) The map of plasmid *p*EM-14E used for the transfection in HEK293T cells. b) Fluorescence confocal microscopic imaging after transfection for 30 h. Condition: 2 mM *cis*-DBDAA. c) Fluorescence imaging of the western blot analysis to verify the full-length expression.

Transfection procedures for HEK293T cell and Western-Blot analysis

HEK293T cells were seeded and cultured to 60%-70 % confluency per well in 35-mm glass-bottom tissue-culture dishes and grown in an antibiotic free DMEM medium supplemented with 10 % fetal bovine serum under the atmosphere containing 5 % CO_2 at 37 °C for 12 h. Then, the plasmid *p*EM-14E (*pMm*DBDAARS-tRNA^{Pyl}-EGFP149TAG-HA), which was purified by an E.Z.N.A Fastfiler Endo-free Plasmid Maxiprep kit, was transiently transfected into the cells by using Lipofectamine 3000 transfection reagent (Thermo scientific) in an opti-MEM medium according to

manufacturer's protocol. After transfection for 5-6 h, the medium was replaced with fresh DMEM medium either with 2.0 mM *cis*-DBDAA or without any ncAA. After 30 h of incubation, the cells were imaged with the Zeiss LSM 780 Microscope (40× objective) by monitoring in the EGFP fluorescence channel.

Then, Western-Blot analysis targeting the HA tag was used to confirm expression of the full-length EGFP-N149*c*DBDAA-HA. After washing by chilled PBS (4 ºC, 2.0 mL × 3), the transfected and over-expressed HEK293T cells were collected by a sterile disposable scraper, and were lysed on ice (10-20 min) through the RIPA lysis buffer. The cell lysates were centrifuged (10 min, 12000 g, 4 ºC) to get rid of insoluble debris. The concentration of supernatant used for the Western-Blot analysis was then determined via BCA assay. The protein samples were then heated with 5X loading buffer at 95 ℃ for 10 min, and was subjected to gel electrophoresis on a 15 % SDS-PAGE (150 V for 90 min). Then, the resolved protein samples in the PAGE gel were transferred to the polyvinylidene difluoride (PVDF) membrane via semi-dry electroblotting (25 V for 30 min). The PVDF membrane was blocked in a blocking buffer for 1 h. The blocked PVDF membrane was washed with TBS buffer (3.0 mL) for three times or more. Anti-HA-tag mouse monoclonal antibody was used to bind and detect the HA-tag in the *C*-terminal domain of EGFP-N149*c*DBDAA. The PVDF membrane was incubated with the anti-HA-tag mouse monoclonal antibody (1:500) overnight at 4 ℃, followed by staining with a biotin-labeled Goat Anti-mouse IgG(H+L) secondary antibody (1:500) overnight at 4 ℃, which was further rendered as fluorescence bands by a NeutrAvidin Oregon Green™ 488 conjugate. Then the fluorescent Western-Blot analysis was carried out with a CHAMPCHEMI multicolor fluorescence imaging system to read out the Oregon Green fluorescence signals on the PVDF membrane.

The nonsense codon suppression efficiency of *c*DBDAARS for encoding *cis*-DBDAA in living HEK293T cell

The plasmid *p*EM-14E-*c*DBDAARS (*pMm*DBDAARS-tRNA^{Pyl}-EGFP149TAG) and pEM-14E (*pMm*PylRS-tRNA^{Pyl}-wt-EGFP) were used for incorporation of *cis*-DBDAA into EGFP with a fluorescence reporter feature and as a benchmark for wild-type EGFP expression level in HEK293T cells, respectively. Therefore, the EGFP fluorescence imaging can be utilized to quantify the incorporation efficiency of the *Mm*DBDAARS based on the EGFP intensity. Accordingly, the plasmids, *p*EM-14E-*c*DBDAARS and *p*EM-14E, were transiently transfected into the HEK293T cells, respectively. Then, the *p*EM-14E-*c*DBDAARS transfected HEK293T cells were cultured in the presence or absence of 2.0 mM *cis*-DBDAA for 30 h to obtain two groups of fluorescence cell imaging results (the first and second rows, Fig. S13). The *p*EM-14E transfected HEK293T cells was cultured under the same conditions without treatment of *cis*-DBDAA (the third row, Fig. S13). Native HEK293T cells under the same culture conditions was also included as a negative control (the fourth row, Fig. S13). In each group, the expression experiments were repeated for at least five times to derive the standard deviation in the fluorescence intensity. As expected, the GCE incorporation of *cis*-DBDAA by the *c*DBDAARS gave a much weaker full-length EGFP level (much weaker fluorescence intensity) inside HEK293T cells intensity in comparison with the wild-type EGFP. Then, quantitative analysis of EGFP fluorescence intensity inside cells was conducted for each image via the ImageJ software. Quantitative results suggested that the nonsense suppression efficiency of *c*DBDAARS is estimated to be only 7.4% of the wild-type reporter protein (EGFP) level in living HEK293T cells. (Fig. S14).

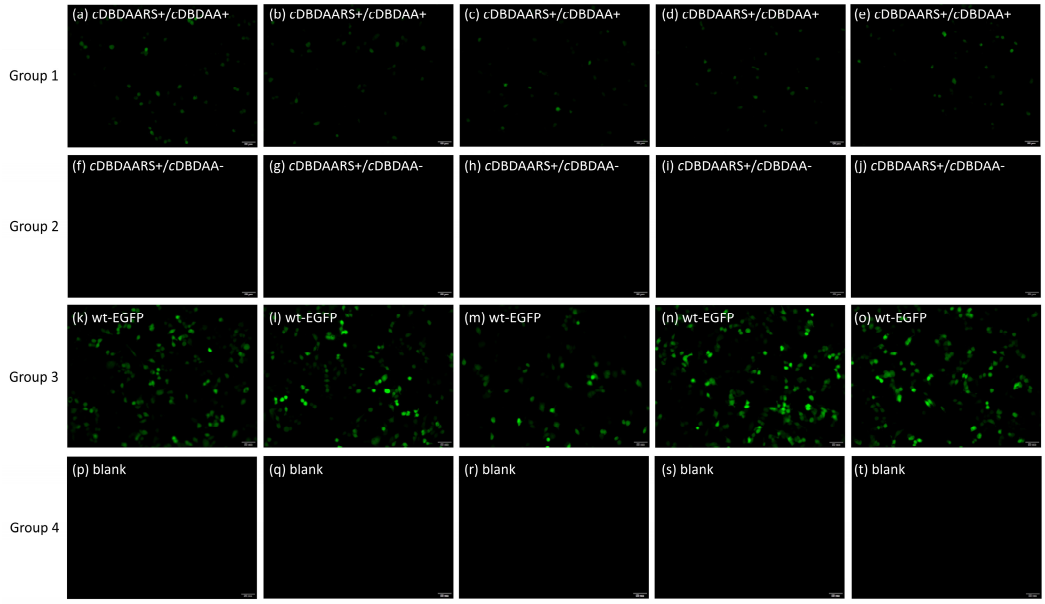


Figure S13. Fluorescence images of living HEK293T cells to quantify the full-length EGFP expression level based on the intensity. The plasmid *p*EM-14E-*c*DBDAARS (*p*MmDBDAARS-tRNA^{Pyl}-EGFP149TAG) or *p*EM-14E (*p*MmPylRS- tRNA^{Pyl}-EGFP) was transiently transfected into the HEK293 cells, respectively, to investigate the GCE incorporation efficiency. (Group 1, a-e) The cells were transfected with *p*EM-14E-*c*DBDAARS and incubated with 2.0 mM *cis*-DBDAA in opti-MEM medium. (Group 2, f-j) The cells were transfected with *p*EM-14E-*c*DBDAARS and cultured in opti-MEM without *c*DBDAA, providing as a negative control. (Group 3, k-o) The cells were transfected with *p*EM-14E (*p*MmPylRS-tRNA^{Pyl}-EGFP) without adding *cis*-DBDAA in opti-MEM medium. (Group 4, p-t) The native HEK293T cells were cultured under the same conditions as another negative control. After the transfection procedure followed by 30 hours incubation, the living cells were imaged in the EGFP channel on an epi-fluorescence microscope (Olympus IX83) to obtain the relative expression level. In each group, the expression experiments were repeated at least for five times to derive the standard deviation in the fluorescence intensity. Scale bar = 50 μm.

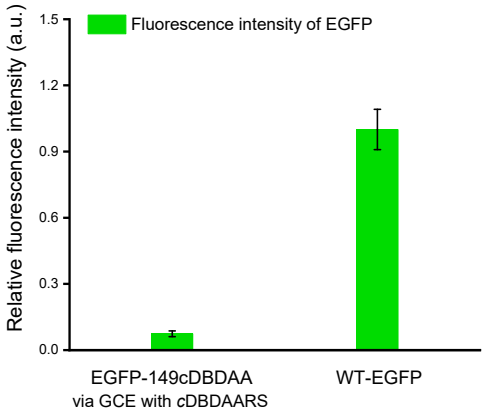


Figure S14. Quantitative analysis of fluorescence intensity in each group of cell imaging graphs, EGFP-149*c*DBDAA vs. WT-EGFP. The GCE suppression efficiency of *c*DBDAARS for *cis*-DBDAA is estimated to be 7.4% of the wild type reporter protein level (EGFP) in living HEK293T cells. The quantitative analysis was conducted via the ImageJ software. Error bars represent the standard deviation from five independent imaging experiments.

HEK293T cell viability under treatment of *cis*-DBDAA

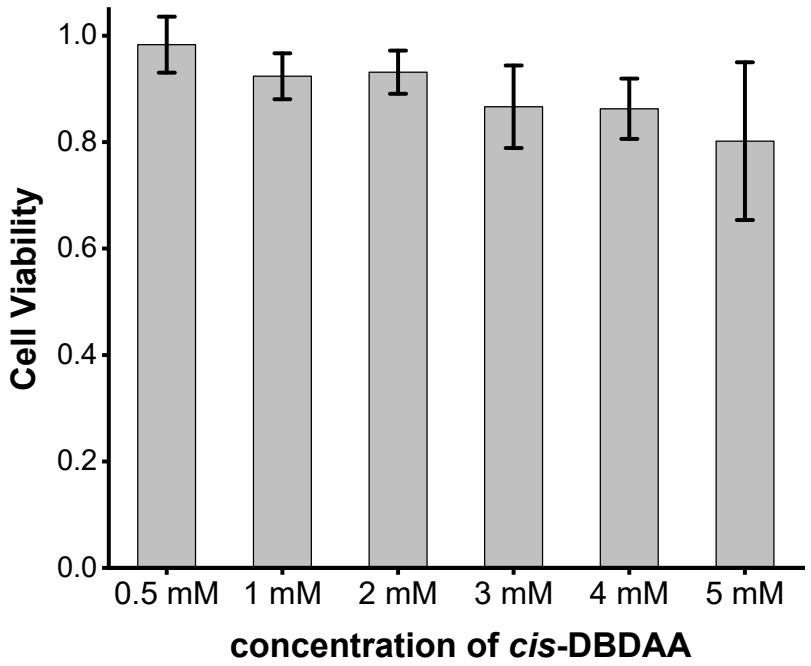


Figure S15. HEK293T cell viability under treatment of various concentration of *cis*-DBDAA for 30 h, which was determined by CCK-8 assay (Beyotime Biotechnology). HEK293T cells were seeded in 96 well plates with density of 30000-50000 per well, then cultured with *cis*-DBDAA dissolved in DMEM supplemented with 10% fetal bovine serum and 1% Pen-Strep for 30 h. The concentrations of *cis*-DBDAA were 0.5 mM, 1 mM, 2 mM, 3 mM, 4 mM and 5 mM, respectively. After incubation, the culture medium was replaced with fresh DMEM medium supplied with 10% CCK-8 reagent. The HEK293T cell viability was determined by measuring the absorbance at 450 nm (iMark^{TM} Microplate Reader, Bio-Rad). Error bars represented four independent experiments.

Determination of the levels of *cis*- and *trans*-DBDAA residue on proteins at the PSS

The characteristic absorbance changes of the DBDAA residues on sfGFP can be used to quantify the levels of both isomers at a specified stimulation wavelength. Accordingly, the real-time absorbance evolution of total 20.1 μM sfGFP-Q204DBDAA at 390 nm was tracked (Fig. 2b; Fig. S16,). The absorption of sfGFP-Q204DBDAA at 390 nm (the n-π* excitation of *cis*-DBDAA) before irradiation was 0.321, and the value becomes 0.283 after 405 nm stimulation for 25 s. The decrease in absorbance at 390 nm is due to the photo-isomerization of *cis*-DBDAA to *trans*-DBDAA when approaching to a PSS. The absorbance of wild-type sfGFP-Q204 at 390 nm was also measured to be 0.243, suggesting that about 61.3% (via the methods and equation: Eq 1. as followed) sfGFP-Q204*c*DBDAA isomerizes into sfGFP-Q204*t*DBDAA at that time point. Accordingly, the level of sfGFP-Q204*c*DBDAA remained at that time point was estimated to be 38.7%.

At 390 nm, R_{tDBDAA} equals to ratio of *t*DBDAA on sfGFP; ε_{Pwt} is the extinction coefficient of wildtype sfGFP, 12089 $cm^{-1}M^{-1}$ at 390 nm; ε_{PPSS} is the extinction coefficient of sfGFP-DBDAA at the PSS derived from the measured absorbance, 14080 $cm^{-1}M^{-1}$; ε_{cDBDAA} = 3881 $cm^{-1}M^{-1}$ is the extinction coefficient of *c*DBDAA on sfGFP; ε_{tDBDAA} = 803 $cm^{-1}M^{-1}$ is the extinction coefficient of *t*DBDAA on sfGFP (estimated from the ^{1}H NMR ratio of free *t*DBDAA at PSS); then

Eq 1.: $R_{tDBDAA} = (\varepsilon_{PPSS}-\varepsilon_{Pwt}-\varepsilon_{cDBDAA})/(\varepsilon_{tDBDAA}-\varepsilon_{cDBDAA})$

$= (14080-12089-3881)/(803-3881) = 0.613 = 61.3\%$

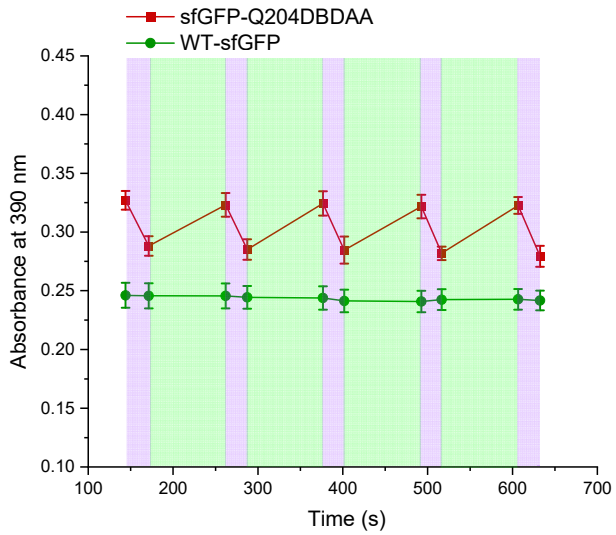


Figure S16. The real-time absorbance evolution of sfGFP-Q204DBDAA at 390 nm. 405 nm (500 mW·cm^{-2}, 25 s) and 532 nm lasers (100 mW·cm^{-2}, 90 s) were used as excitation light sources in an alternate sequence. The absorbance results were averaged to avoid systematic errors due to instrumental noise.

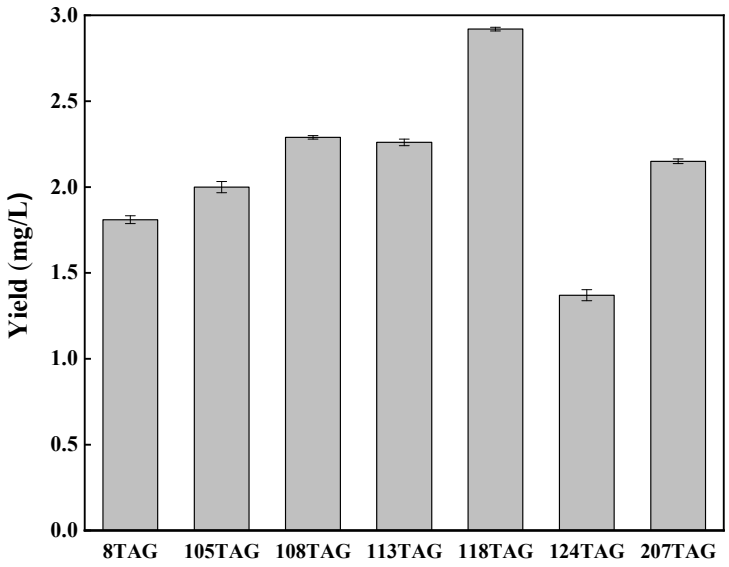


Figure S17. The expression yield of *sj*GST mutants site-specifically incorporated with *cis*-DBDAA was 1.3-2.9 mg·L^{-1}, which was measured by the BCA assay.

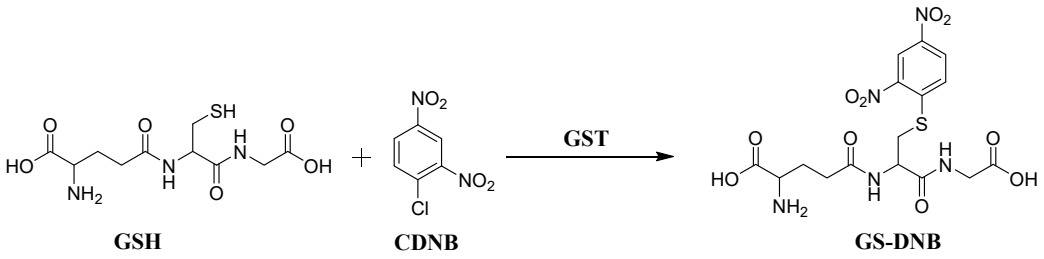


Figure S18. The catalytic ligation mechanism of *sj*GST. The characteristic absorption of GS-DNB is at 340 nm, which could be used to characterize the enzymatic activity of *sj*GST mutants. The enzymatic activity of *sj*GST mutants was measured by Glutathione S-transferase (GST) Activity Assay Kit, which was purchased from Solarbio.

Host-guest recognition between CB[7] and trans-DBDAA observed by ^{1}H NMR analysis

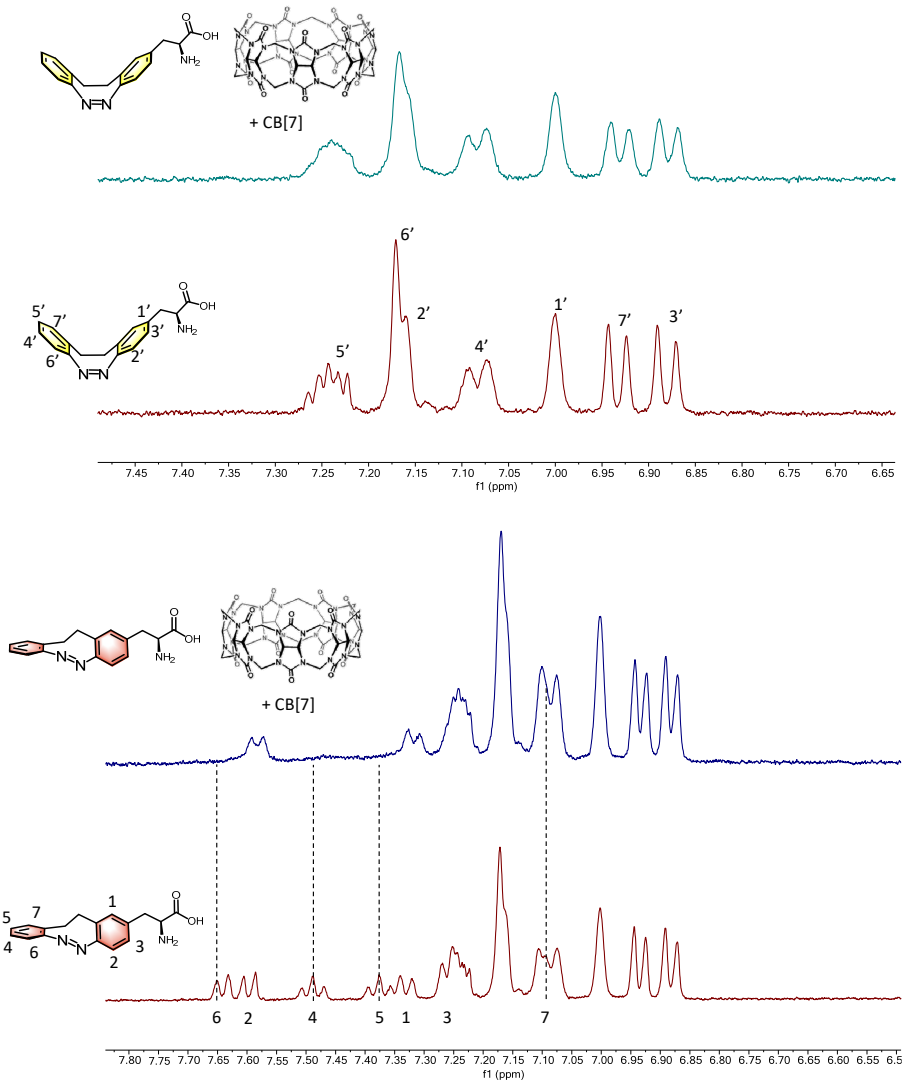


Figure S19. ^{1}H NMR monitored for the host-guest interaction between *cis*- or *trans*-DBDAA and CB[7] at 298 K (condition: 2.0 mM in D_2O, 400 MHz). Top panel: in dark; lower panel: after 405 nm stimulation for 1 min.

The expression of sfGFP-Q204DBDAA by utilizing the opti-*c*DBDAARS-tRNA^{Pyl} pair in the presence or absence of the 405 nm stimulation

sfGFP-Q204DBDAA was over-expressed by utilizing the opti-*c*DBDAARS-tRNA^{Pyl} pair in the presence or absence of the 405 nm stimulation (conditions: 2.0 mM *c*DBDAA), and the resulting proteins were purified for SDS-PAGE analysis. Under the continuous 405 nm stimulation, the expression system contained a mixture of *cis*-DBDAA and *trans*-DBDAA (PSS *cis: trans* =3:4, at the PSS) and the concentration of *cis*-DBDAA in the culture medium decreased (≈ 0.86 mM). As it can be seen in the results (Fig. S21) that opti-*c*DBDAARS is still able to express the full-length sfGFPQ204DBDAA under 405 nm exposure but with a much lower yield (≈ 0.24 mg•L^{-1}), which was about 48% of the yield expressed in dark (≈ 0.50 mg•L^{-1}). The GCE expression yield was significantly lowered with the *c*DBDAA converted to the PSS in which the concentration of *c*DBDAA should be 42.9% of the total DBDAA. The concentration ratio of *cis*-DBDAA in the culture medium at the PSS is presumably equal to the ratio determined by in-situ photo-activated proton NMR spectroscopy (Fig. S1).

Actually, a definitive conclusion cannot be provided because *cis*-DBDAA cannot be completely converted to *trans*-DBDAA but remains as a ratio at the PSS under 405 nm irradiation. But it is certain that even if *c*DBDAARS is able to incorporate *trans*-DBDAA, the expression level must be much lower than *cis*-DBDAA. There are two possible reasons for this phenomenon. First, *trans*-DBDAA indeed can be incorporated with a lower expression level. Second, the decrease of *cis*-DBDAA concentration in the system leads to the decrease in expression efficiency. Assuming the protein expression level for each isomer of the ncAA is proportional to the concentration of each isomer, the GCE expression level for *trans*-DBDAA via *c*DBDAARS is then calculated to be as low as 0.045 mg•L^{-1} by considering that the incorporation of both *cis*- and *trans*-

isomers at the PSS attributes to the total expression level. However, this value is too low to be of any significance.

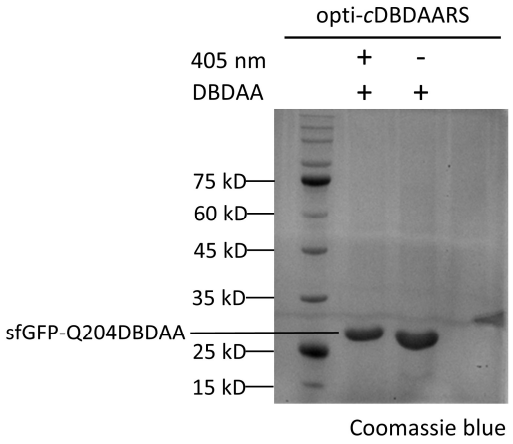


Figure S20. SDS-PAGE imaging for the purified sfGFP-Q204DBDAA expressed in the presence or absence of 405 nm stimulation by opti-*c*DBDAARS synthetase. (Condition: 2.0 mM *cis*-DBDAA)

Effect of continuous photo-stimulation on the growth of *Escherichia coli*

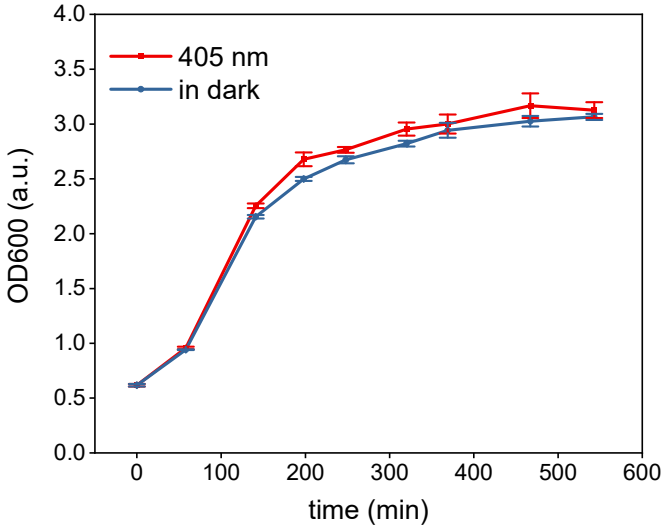


Figure S21. The growth curve of *E. Coli* in the presence or absence of 405 nm LED stimulation. The glycerol stock of *E. Coli* BL21 (DE3) was inoculated in LB medium

with tetracycline (34 μg·mL^{-1}) and ampicillin (100 μg·mL^{-1}) and cultured at 37 °C, 250 rpm. It can be seen that continuous 405 nm stimulation has no obvious effect on the growth of bacteria. Exposure to 405 nm (red line); kept in dark (blue line). The density of bacteria was measured via OD600 by Thermo Nanodrop 2000C.

LC-MS analysis of sfGFP-N149TAG expressed under 405 nm photo-stimulation

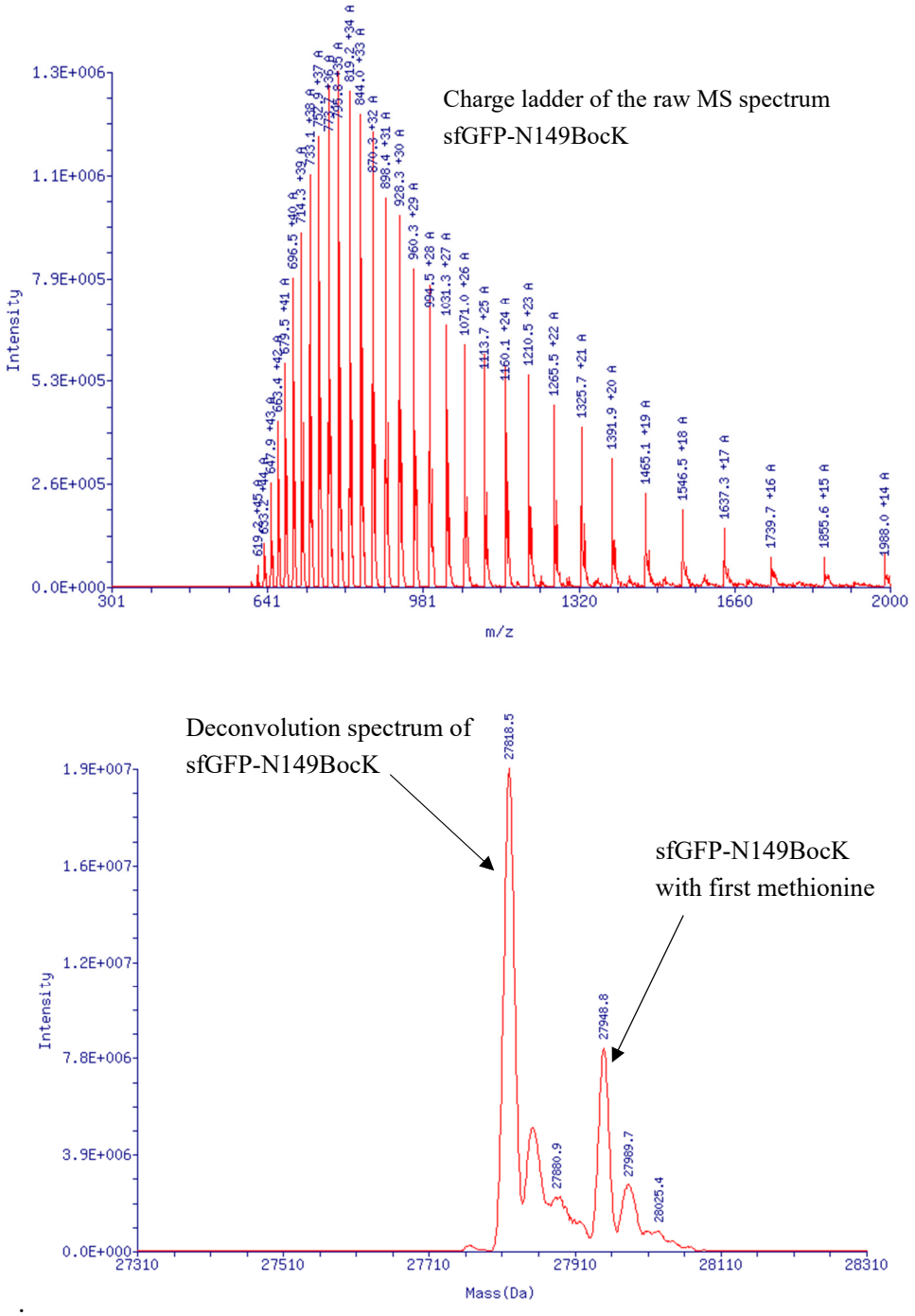


Figure S22. Deconvoluted LC-MS analysis of sfGFP-N149BocK with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H^+]: 27818.5 Da.

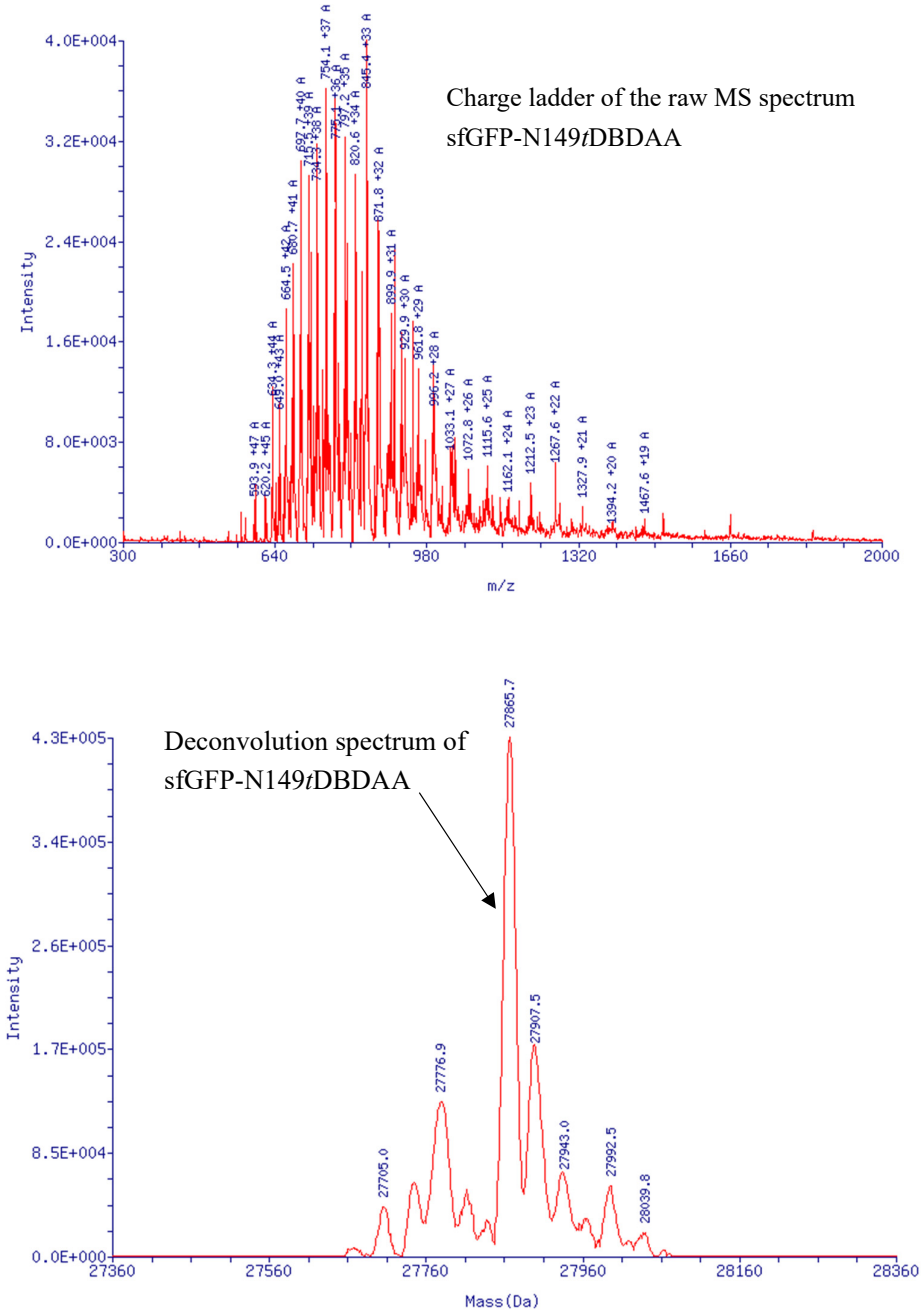


Figure S23. Deconvoluted LC-MS analysis of sfGFP-N149*t*DBDAA with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H^+]: 27865.7 Da. Calcd. mass 27867.5 Da. The protein was expressed with the 2 mM *cis*-DBDAA and continuous exposure to 405 nm LED.

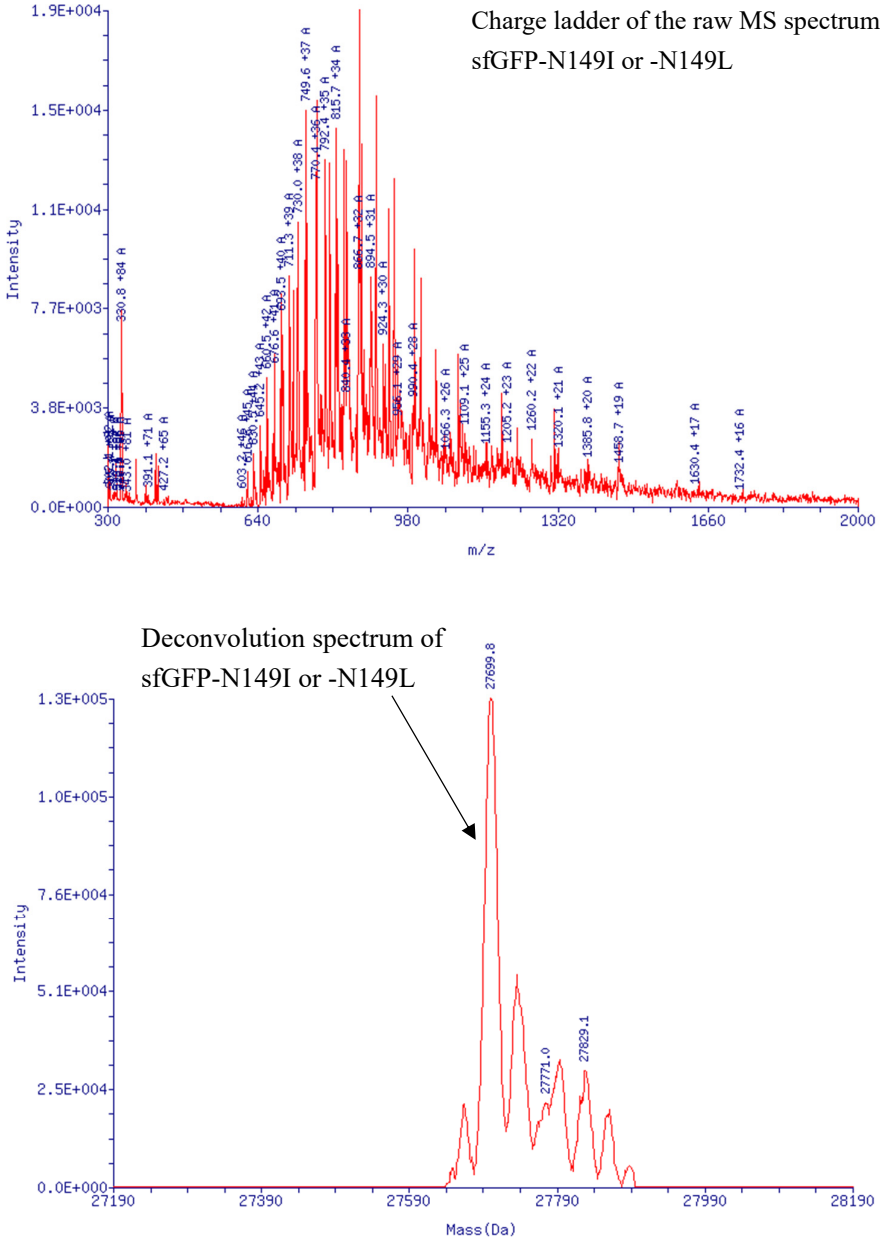


Figure S24. Deconvoluted LC-MS analysis of sfGFP-N149I or -N149L (comes from the expression conditions with 2 mM *cis*-DBDAA but without 405 nm LED exposure) with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H^+]: 27699.8 Da. The signal peaks suggested a slight suppression of amber stop codon which was translated as natural amino acids, either Ile or Leu.

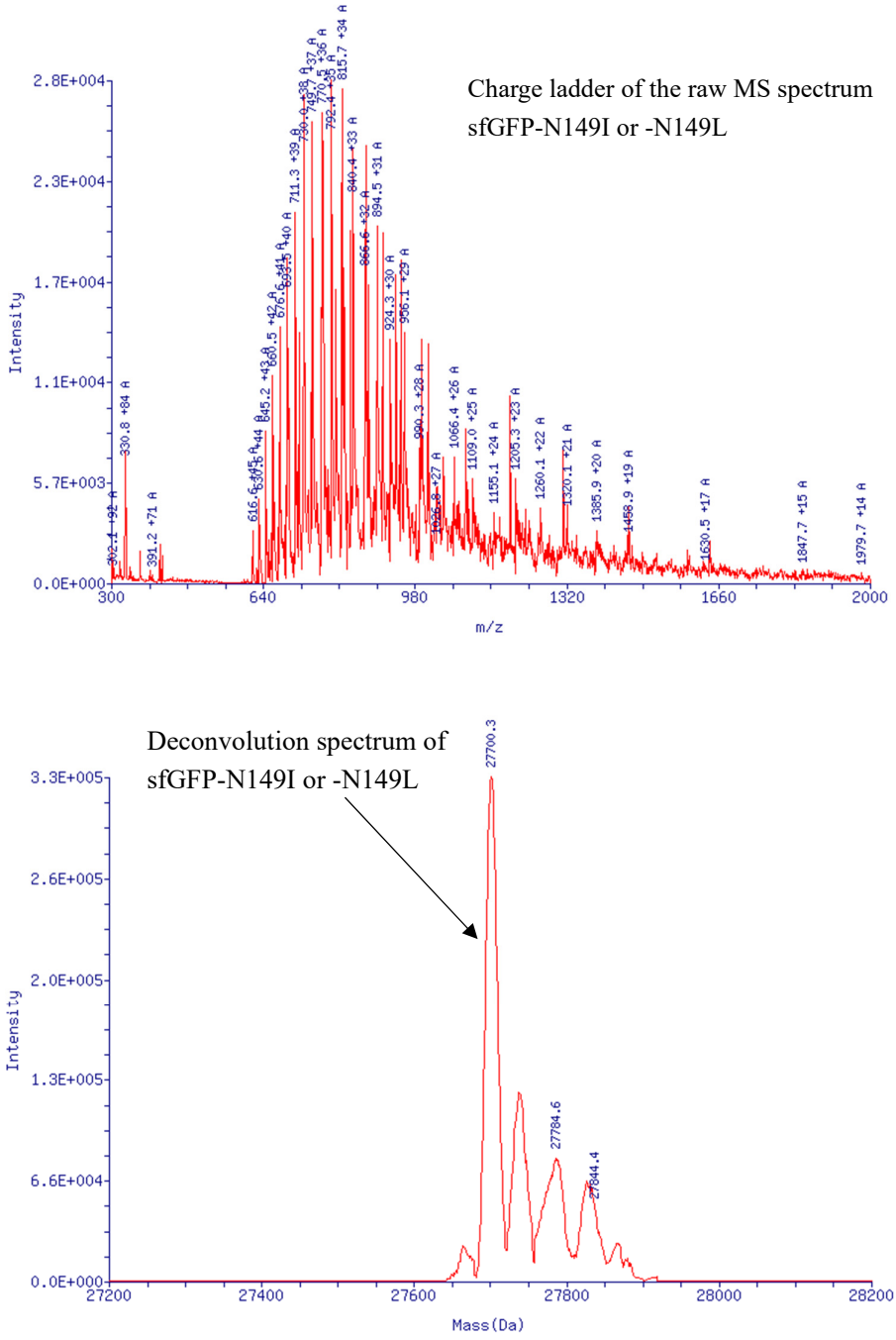


Figure S25. Deconvoluted LC-MS analysis of sfGFP-N149I or -N149L (comes from the expression conditions with only continuous exposure to 405 nm LED but without supply of *cis*-DBDAA) with the charge ladder and the deconvoluted mass spectrum. Found [M-Met+H^+]: 27700.3 Da. The signal peaks suggested a slight suppression of amber stop codon which was translated as natural amino acids, either Ile or Leu.

The photo-labelling of sfGFP with a Cy3 fluorophore via the sydnone-DBDA photo-click strategy

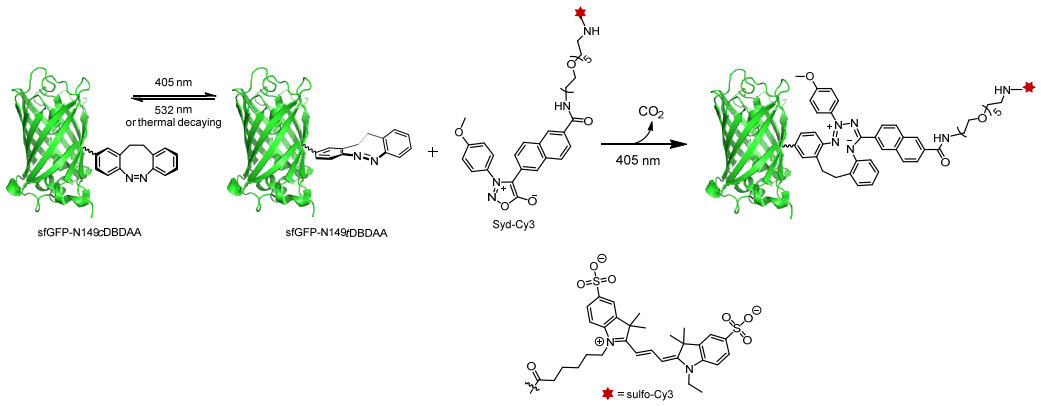


Figure S26. Photo-click labeling reaction between sfGFP-N149DBDAA and Syd-Cy3 to verify the presence of the DBDAA residue on the sfGFP. Only the recombinant sfGFP-N149 containing a DBDAA residue can be photo-labelled by the Syd-Cy3 probe to show a fluorescence band of Cy3 in the subsequent in-gel fluorescence SDS-PAGE analysis. The photo-click reaction conditions: 3.0 μM proteins and 30 μM Syd-Cy3 in PBS (pH = 7.4), 405 nm LED array (13.4 $mW{\cdot}cm^{-2}$) for 2 min, followed by electrophoresis in SDS-PAGE and in-gel fluorescent analysis.

Figure note for Figure 26: The continuous 405 nm illumination is to maintain the constant concentration of *t*DBDAA (at the PSS) in the bacterial expression system. Then, the configuration sensitivity of the *Mm*PylRS synthase (the *t*DBDAARS) for recognizing the ncAA substrate can be utilized to distinguish *t*DBDAA from *c*DBDAA. What needs to be pointed out here is that it usually takes at least 5 hours to complete the protein purification procedure (from lysis to desalting) after the induced over-expression in the *E. coli* cells. The short half-life (< 40 min) of *trans*-DBDAA leads to the configurational change from *trans*- at the PSS to complete *cis*-isomer during the purification procedure because our purification procedure cannot be carried out under

continuous 405 nm light irradiation. In fact, all the harvested sfGFP-N149*t*DBDAA has been isomerized to the form of sfGFP-N149*c*DBDAA via the fast thermal decaying.

When taking advantage of the sydnone-DBDA photo-click reaction to verify the presence of DBDAA residue on sfGFP-N149 position, the DBDAA residue had been switched back to *cis*-configuration even though the GCE expression was enabled in the presence of *trans*-DBDAA produced in-situ by the 405 nm photo-stimulation. And the 405 nm photo-stimulation plays two significant roles in this photo-click labeling procedure, first: making the *cis*-DBDAA residue photo-isomerize to *trans*-DBDAA into a PSS, and second: inducing the photolysis of Syd-Cy3 to produce the reactive nitrile imine intermediates (short-lived), then the cycloaddition between *trans*-DBDAA and nitrile imine can be realized to build a covalent conjugate.

The molecular docking analysis for Ile or Leu to the *Mmt*DBDAARS

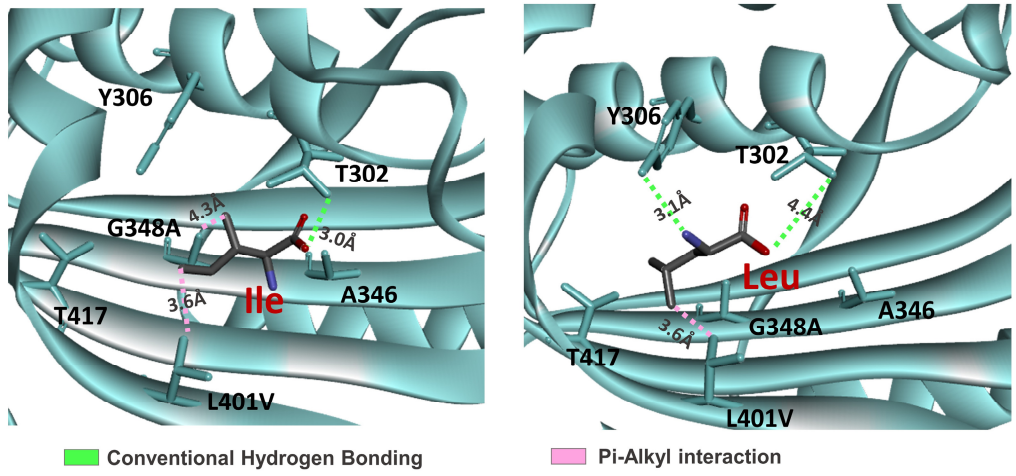


Figure S27. The background amber codon suppression possibly via binding of Ile or Leu into the active site of the evolved *Mmt*DBDAARS. The binding mode was analyzed by molecular docking. Ile or Leu was docked into the active cavity of *Mmt*DBDAARS with binding energies, -25.24 kcal·mol^{-1} and -31.18 kcal·mol^{-1}, respectively. The binding postures were displayed from front views. The relatively low docking interaction energy may be the reason for the poor expression of designated proteins harbouring natural amino acids at the TAG codon site. The key amino acid residues

(cyan) in the active pocket of *Mmt*DBDAARS that interacted with either Ile or Leu (gray) are denoted and displayed in stick mode.

Reference

1. Q. Xiong, T. Zheng, X. Shen, B. Li, J. Fu, X. Zhao, C. Wang and Z. Yu, *Chem. Sci.*, 2022, **13**, 3571-3581.

2. X. Wu, J. Deng, G. Guo, Y. Zheng, Q. Xiong, T. Zheng, X. Zhao and Z. Yu, *Chem. Eur. J.,* 2021, **27**, 11957-11965.

3. J. Gao, Q. Xiong, X. Wu, J. Deng, X. Zhang, X. Zhao, P. Deng and Z. Yu, *Commun. Chem.*, 2020, **3**, 29.

4. R. Siewertsen, H. Neumann, B. Buchheim-Stehn, R. Herges, C. Näther, F. Renth and F. Temps, *J. Am. Chem. Soc.,* 2009, **131**, 15594-15595.

5. J. Deng, X. Wu, G. Guo, X. Zhao and Z. Yu, *Org. Biomol. Chem.*, 2020, **18**, 5602-5607.

6. V. K. Lacey, G. V. Louie, J. P. Noel and L. Wang, *ChemBioChem*, 2013, **14**, 2100-2105.

7. L. Tang, H. Gao, X. Zhu, X. Wang, M. Zhou and R. Jiang, *BioTechniques*, 2012, **52**, 149-158.

8. J. K. Takimoto, N. Dellas, J. P. Noel and L. Wang, *ACS Chem. Biol*., 2011, **6**, 733-743.

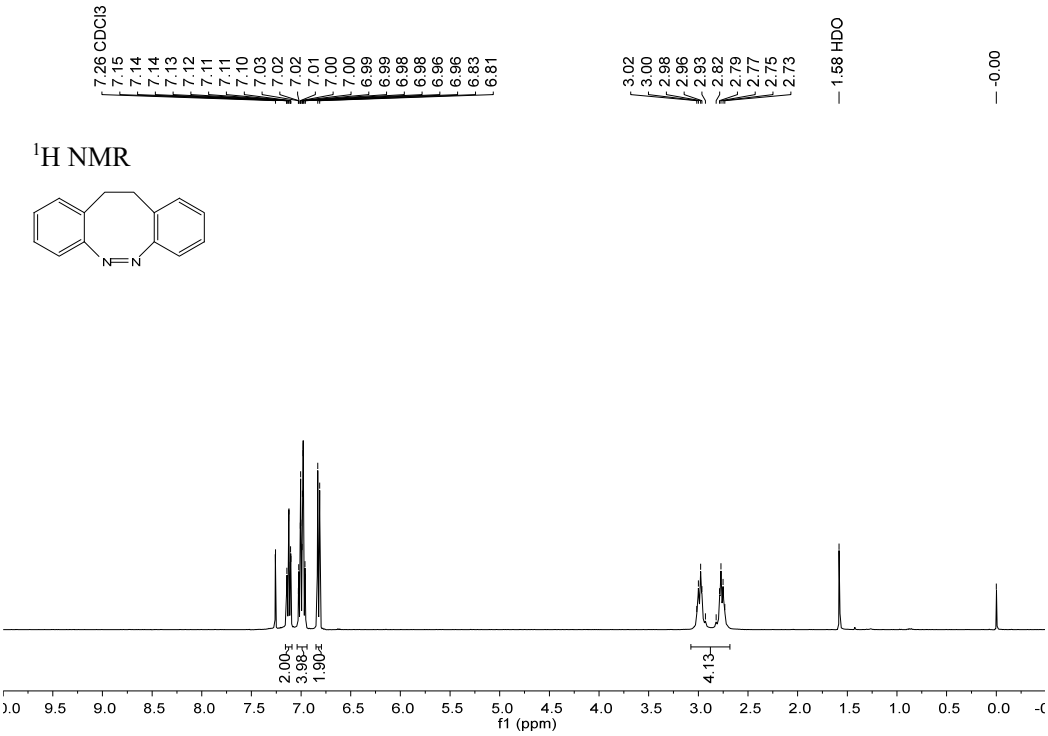
9. T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto and S. Yokoyama, *Chem. Biol.*, 2008, **15**, 1187-1197.

10. A. Khmelinskii, P. J. Keller, A. Bartosik, M. Meurer, J. D. Barry, B. R. Mardin, A. Kaufmann, S. Trautmann, M. Wachsmuth, G. Pereira, W. Huber, E. Schiebel, M. Knop, *Nat. Biotechnol.,* 2012, **30**, 708-714.

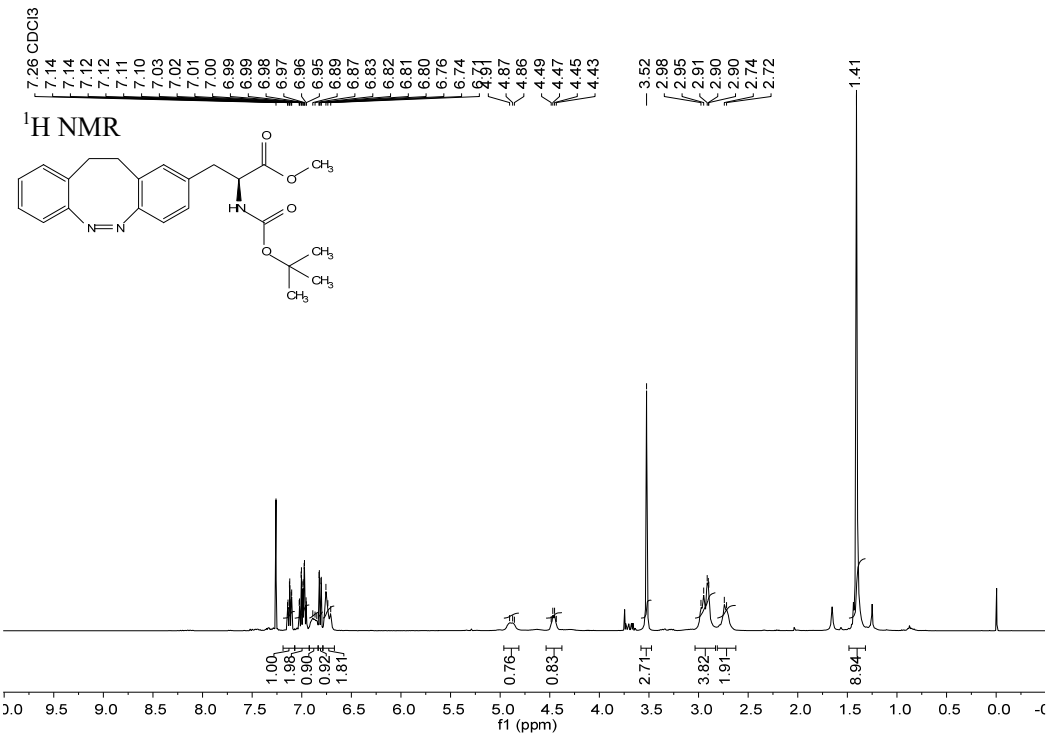
11. C. Hoppmann, V. K. Lacey, G. V. Louie, J. Wei, J. P. Noel and L. Wang, *Angew. Chem. Int. Ed.*, 2014, **53**, 3932-3936.

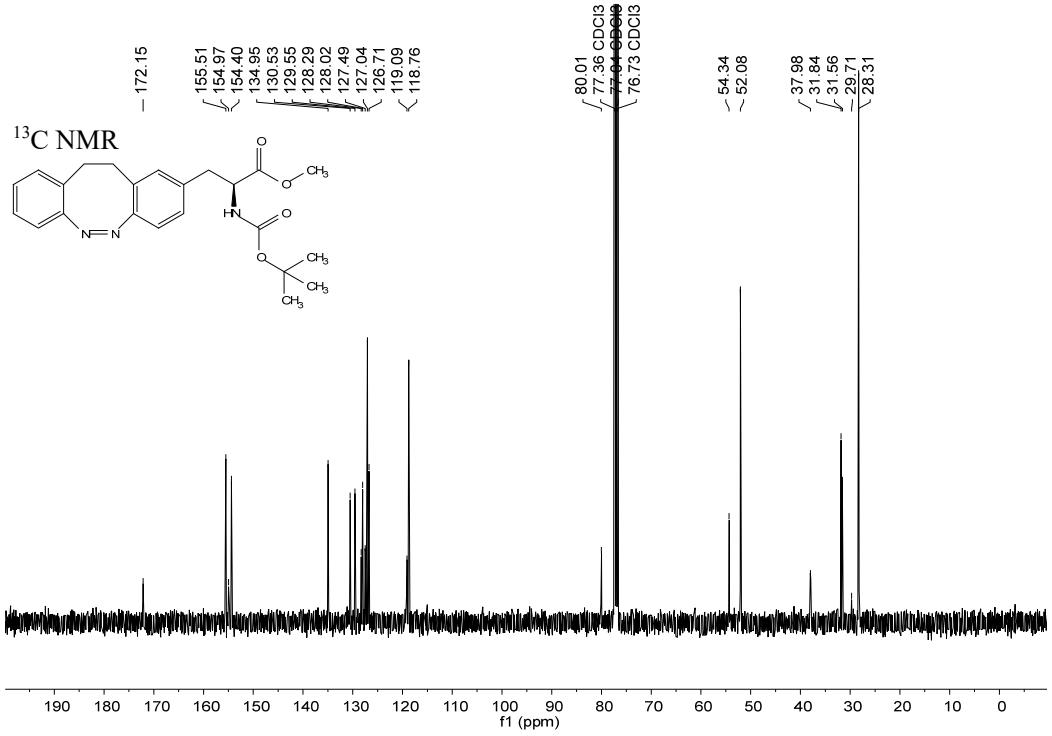
12. Y. H. Tsai, S. Essig, J. R. James, K. Lang and J. W. Chin, *Nat. Chem.*, 2015, **7**, 554-561.

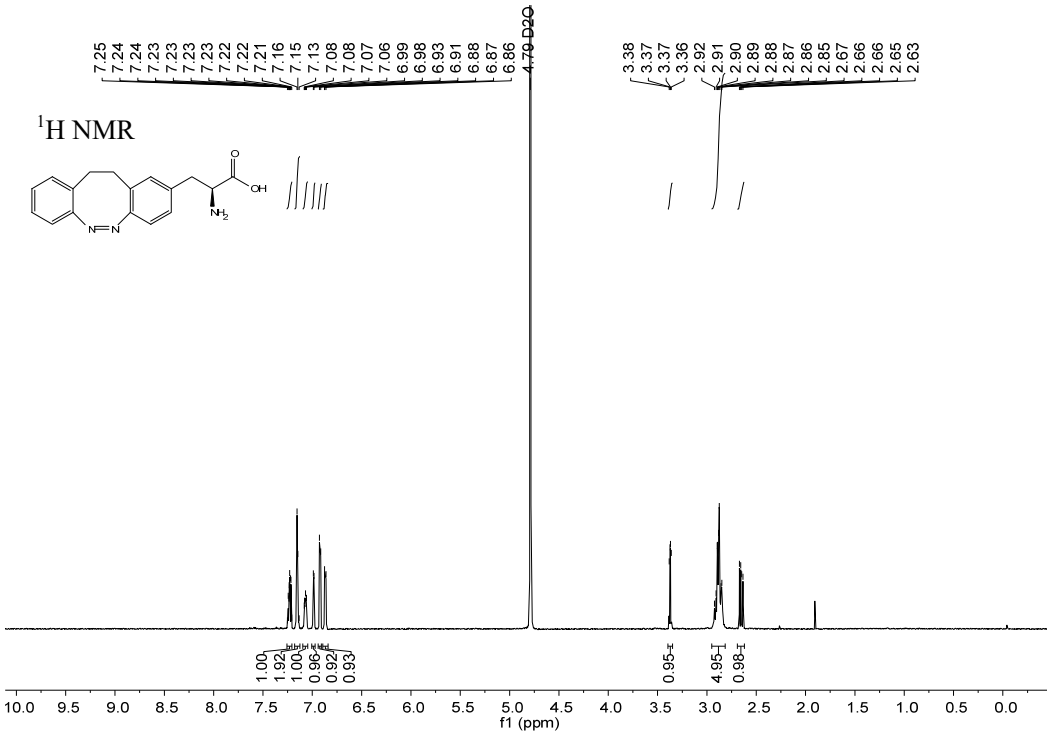
^{1}H NMR and ^{13}C NMR Spectra

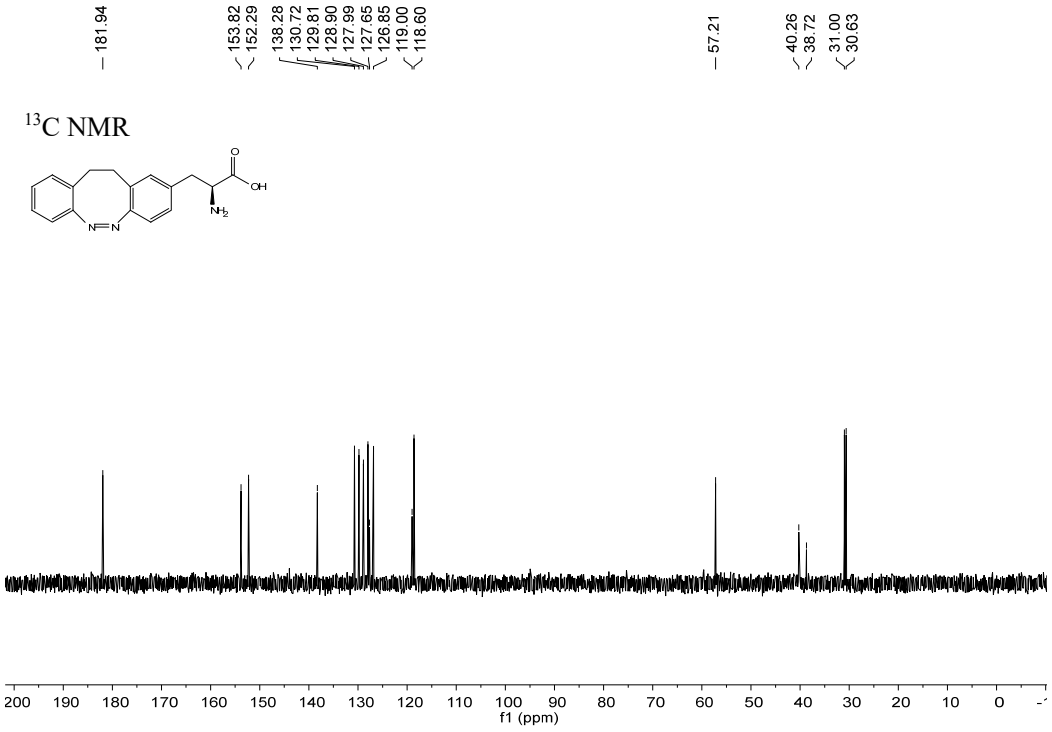


^{1}H NMR




172.15
155.51
154.97
154.40
134.95
130.53
129.55
128.29
128.02
127.49
127.04
126.71
119.09
118.76
80.01
77.36 CDCl3
77.04 CDCl3
76.73 CDCl3
54.34
52.08
37.98
31.84
31.56
29.71
28.31
13C NMR
f1 (ppm)


7.25
7.24
7.24
7.23
7.23
7.23
7.23
7.22
7.22
7.21
7.16
7.15
7.13
7.08
7.08
7.07
7.06
6.99
6.98
6.93
6.91
6.88
6.87
6.86
4.79 D2O
3.38
3.37
3.37
3.36
2.92
2.91
2.90
2.89
2.88
2.87
2.86
2.85
2.67
2.66
2.66
2.65
2.63
1H NMR
1.00
1.92
1.00
0.96
0.92
0.93
0.95
4.95
0.98
f1 (ppm)


181.94
153.82
152.29
138.28
130.72
129.81
128.90
127.99
127.65
126.85
119.00
118.60
57.21
40.26
38.72
31.00
30.63
13C NMR
O
OH
NH2
N=N
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)