Chem, Volume *11*

Supplemental information

Biosynthesis and genetic encoding of activated nitriles for fast protein conjugation and tunable fluorogenic labeling

Elwy H. Abdelkader, Haocheng Qianzhu, Gottfried Otting, and Thomas Huber

Supplemental Information

Table of contents:

Supplemental methods:

Selection of functional G1PylRS enzymes recognizing CysCNP

A previously established library of G1PylRS mutants [S1-S4], encoded on the pBK-G1RS plasmid, was transformed into *E. coli* DH10B cells harbouring the pBAD-H6RFP reporter plasmid. After transformation, the culture was directly inoculated into 25 mL LB medium supplemented with 100 mg/L carbenicillin, 50 mg/L kanamycin, 0.4% L-arabinose and 0.5 mM CNP-thiol. This culture served as the sample for the first round of positive selection (**1P+**). Overnight expression at 37 °C led to readily detectable level of RFP expression.

Following the overnight incubation, the cells were harvested, resuspended in 5 mL of PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) and diluted 100-fold to a concentration suitable for fluorescence-activated cell sorting (FACS). FACS was performed using a FACSAria Fusion cell sorter (BD Biosciences, USA; **Figure S2**). Cells with high RFP levels were collected from the **1P+** sample, representing 1.6% of the total population (indicated by violet shades in **Figure S2**), and subjected to a subsequent round of negative selection (**2N–**) in the absence of CNP-thiol.

Cells exhibiting low RFP expression (45.3%) were collected from the **2N–** sample and aliquoted to inoculate media under positive (**3P+**, with CNP-thiol) and negative (**3P–**, without CNP-thiol) conditions. The top RFP-expressing cells from the **3P+** sample (2.4%) were collected and underwent another negative selection round (**4N–**, without CNP-thiol). From the **4N–** sample, cells showing low RFP expression (33.1%) were collected and aliquoted again to inoculate media under positive (**5P+**, with CNP-thiol) and negative (**5P–**, without CNP-thiol) conditions. The cultures demonstrated a clear response to the presence of CNP-thiol and the top 3.2% of RFP-fluorescent cells from the **5P+** sample were collected.

Approximately 2,000 of these cells were recovered by plating on LB agar plates containing 100 mg/L carbenicillin and 50 mg/L kanamycin. Isolated colonies were analyzed in 96-well plates. Sixty enzyme candidates were inoculated into media under both positive (with 0.5 mM CNP-thiol) and negative (without CNP-thiol) conditions. The red fluorescence intensity was measured as an indicator of RFP expression and normalized to the OD_{600} of the cell culture using a TECAN Infinite 200 Pro M Plex plate reader (Tecan, Switzerland; **Figure S3**). DNA sequence analysis identified two different candidates as G1PylRS mutants incorporating CysCNP. The amino acid mutations of these candidates are listed in **Table S1**.

In vivo protein expression and purification

E. coli B-95.ΔAΔ*fabR* cells [S5] were co-transformed with the pRSF-G1(CysCNP)RS plasmid (for CysCNP or CysCNPym incorporation) or pRSF-G1mCNPRS plasmid[2] (for mCNP incorporation) along with the pCDF plasmid carrying the gene of the protein of interest, which were synthesized by Twist Bioscience, USA (**Table S5**). Cells were initially cultured in LB medium supplemented with 25 mg/L kanamycin and 25 mg/L spectinomycin at 37 °C. Subsequently, a 0.5 mL aliquot of the overnight culture was used to inoculate 50 mL LB medium supplemented with 25 mg/L kanamycin, 25 mg/L spectinomycin and the cells were grown at 37 °C to an OD_{600} of 0.6–1. At this point, the culture was supplemented with either

0.5 mM CNP-thiol, 0.5 mM CNPym-thiol or 1 mM mCNP and the temperature was reduced accordingly. Protein expression was then induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

Following protein expression, cells were harvested by centrifugation at 4,000 *g* for 15 minutes at 4 °C. The cell pellet was resuspended in buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol, 10 mM imidazole) and lysed on ice by sonication using 50% power and 50% pulse length for 10 min (ultrasonic homogenizer Omni-Ruptor 4000, Omni International, USA). The cell lysate was centrifuged at 30,000 *g* for 1 h at 4 °C. The clarified supernatant was then loaded onto a 1 mL His GraviTrap column (Cytiva, USA). After washing with 20 column volumes of buffer B (same as buffer A but with 20 mM imidazole), the target protein was eluted using 5 column volumes of buffer C (same as buffer A but with 500 mM imidazole). Finally, buffer exchange to PBS (pH 7.4) was performed using an Amicon ultrafiltration centrifugal filter with a 10 kDa molecular weight cut-off (Merck Millipore, USA).

Effect of CNP-thiol concentration on the amber codon suppression efficiency

E. coli B-95.ΔAΔ*fabR* cells were co-transformed with pRSF-G1(CysCNP)RS and pCDF-RFP amber plasmids and the cells were recovered at 37 °C on LB agar plates containing 25 mg/L kanamycin and 25 mg/L spectinomycin. Afterwards, the cells were used to inoculate 10 mL LB medium supplemented with 25 mg/L kanamycin and 25 mg/L spectinomycin to prepare the pre-culture. Following growth at 37 °C for 16 h, the preculture was used to inoculate 500 mL LB media containing the same antibiotics concentration. The culture was grown at 37 °C until OD_{600} 0.6-1, where 30 mL fractions of the culture were aliquoted into clean 250 mL flasks and a serial concentrations (1–1000 μM) of CNP-thiol were added. Next, the cells were grown at 25 °C for 15 min and protein expression was induced by the addition of 1 mM IPTG. After 16 h at 25 °C, 100 μL from each flask was transferred into 96-well plates to measure the RFP fluorescence and OD_{600} (**Figure S6**).

Incorporation of CysCNP using high cell-density fermentation

E. coli B-95.ΔAΔ*fabR* cells were co-transformed with pRSF-G1(CysCNP)RS and pCDF-NT_Ubi amber plasmids and the cells were recovered at 37 °C in LB medium containing 50 mg/L kanamycin and 50 mg/L spectinomycin. Afterwards, the cells were used to inoculate 25 mL TB medium supplemented with 50 mg/L kanamycin and 50 mg/L spectinomycin (pre-culture). Following growth at 37 °C for 16 h, the pre-culture was used to inoculate 225 mL rich fermenter medium containing 50 mg/L kanamycin and 50 μg/L spectinomycin in a Labfors 5 bioreactor (Infors-HT, Switzerland). The cells were grown at 37 °C with air flow = 1.25 L/min, minimal pO_2 = 20%, stirring cascade: 500–1200 rpm, pH = 7.0. When the OD_{600} value of the culture reached 15–20, the temperature was reduced to 18 °C and 0.5 mM CNP-thiol (dissolved in 0.5 mL DMSO) was added. After 15 minutes, protein overexpression was induced by the addition of 1 mM IPTG. After 16 h, the fermentation was stopped and the cells were harvested. The final OD_{600} values were 43–51 [S6].

The harvested cells were lysed as described above and the clarified cell lysate was loaded onto a 5 mL HisTrap FF column connected to an ÄKTA pure 25 chromatography system (Cytiva, USA). The column was washed with 20 column volumes buffer A and the protein was eluted with 3 column volumes buffer C. Next, the buffer of the eluted protein was exchanged to PBS buffer using an Amicon ultrafiltration centrifugal tube with a molecular weight cut-off of 10 kDa. The yield of NT_Ubi CysCNP was 1.8 ± 0.1 g/L (mean ± SD, n=3). High purity and CysCNP incorporation fidelity were confirmed by SDS-PAGE analysis and intact protein mass spectrometry, respectively (**Figure S5**).

In vitro arcNAT click reaction with ABTs

ABT stock solution preparation:
For ABTs **1-5**: dissolve 0.1 mmol ABT in 50 µL DMSO, then add 950 µL 100% glycerol to obtain a 100 mM stock solution. For ABT **6**: dissolve 0.1 mmol ABT in 400 µL of 1 M aqueous KOH solution, then add 600 µL 100% glycerol to obtain a 100 mM stock solution.

For the reaction with ABTs, a 50 µM protein sample in degassed PBS buffer containing 30 mM TCEP was mixed with the appropriate ABT concentration from the 100 mM stock solution: 20 mM for proteins containing mCNP or CysCNP, or 2 mM for proteins containing CysCNPym. The reaction mixture was incubated with shaking. Afterwards, the buffer was exchanged to PBS using an Amicon ultrafiltration centrifugal filter. Relative quantum yields were determined according to the protocol of Würth et al. on a Cary 60 UV-Vis spectrophotometer and Cary Eclipse Fluorescence spectrophotometer (Agilent, USA) using either quinine sulfate or ribovlavin as reference standard [S7]. The CIE 1931 xy chromaticity diagrams were created with the Chromaticity Diagram Template in OriginPro 2021 (OriginLab Corporation, USA)

Live-cell labelling with the NAT click reaction

Cultures expressing CysCNPym containing proteins were grown as described previously (*In vivo* protein expression and purification section). After 5 h expression at 30 °C, 2 mL cultures were harvested by centrifugation at 4,000 *g* for 10 minutes at 25 °C. To remove excess CNPym-thiol and CysCNPym, the supernatant was discarded, and the cell pellet was resuspended in 1 mL PBS, followed by centrifugation at 4,000 *g* for 5 min at 25 °C. This step was repeated twice. At this point, cells could be stored at 4 °C. For labelling with 1,2-aminothiol-bearing biomolecules, cells were resuspended and incubated for 4 h at 25 °C in 200 µL of the tagging solution containing 20-200 µM C-RFP, 30 mM TCEP, degassed PBS buffer pH 7.4. For labelling with the fluorogenic ABT, cells were resuspended and incubated for 4 h at 25 °C in 200 µL of the tagging solution containing 2 mM ABT, 30 mM TCEP, degassed PBS buffer pH 7.4.

Live-cell imaging following the NAT click reaction

For confocal fluorescence microscopy, *E. coli* cells labelled with C-RFP were washed three times with PBS buffer by centrifugation at 4,000 *g* for 2 minutes to remove any excess C-RFP. 10 μL of the resuspended cells (diluted threefold compared to the overnight culture) were loaded onto 1% agarose pads and covered with coverslips [S8]. Cells labelled using the fluorogenic NAT click reaction required no washing and were directly prepared for microscopy.

Imaging was performed using a Zeiss LSM780 UV-NLO confocal microscope (Carl Zeiss Microscopy GmbH, Germany). Consistent laser power and detection gain were maintained for samples labelled with the same fluorophore. RFP was excited with a 561 nm laser (570–695 nm emission channels), while fluorophores **y2** and **y6** were excited with 405 nm (419–695 nm emission channels) and 458 nm (517–695 nm emission channels) lasers, respectively. The focus was adjusted based on the fluorescence channel and the transmission channel was used to capture the brightfield image simultaneously. Image processing was done using ZEN lite 3.10 software (Carl Zeiss Microscopy GmbH, Germany). Statistical analysis of the labelled cells used in the imaging was conducted on an FACSAria Fusion cell sorter (BD Biosciences, USA) using the appropriate laser and filter combinations for each fluorophore: RFP: 560 nm laser and a 610/20 nm bandpass filter; **y2**: 405 nm laser and a 525/50 filter; **y6**: 405 nm laser and a 610/20 nm bandpass filter.

Intact protein mass spectrometry analysis

Intact protein analysis was conducted using a Thermo Fisher Scientific UltiMate 3000 HPLC system connected to an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific, USA). The HPLC system is connected to a ZORBAX 300SB-C3 column (3.5 μm, 4.6 x 50 mm; Agilent Technologies, USA). Approximately 30 pmol of the protein sample was injected and separated using a 500 μL/min linear gradient of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B), with solvent B increasing from 5% to 80% over 7 min. Data was acquired in positive ion mode using an electrospray ionization (ESI) source. The Intact protein mass was determined by deconvolution using Xcalibur 3.0.63 software (Thermo Fisher Scientific, USA).

Kinetic analysis of CysCNP and CysCNPym reactivities in the NAT click reaction

To evaluate the reactivity of CysCNP and CysCNPym, the second-order rate constants (k_2) for their NAT click reaction with 1,2-aminothiols were determined. For CysCNP, 100 μM NT_Ubi CysCNP samples in 96-well plate were incubated with serial concentrations of L-cysteine (1–16 mM) in PBS buffer containing 20 mM TCEP at 37 °C. The exponential decay over time of the intrinsic fluorescence of CysCNP (λ_{ex} = 330 nm, λ_{em} = 400 nm) was determined under pseudo first-order conditions in order to determine the observed rate constants ($k`$; **Figure S8A)**. By plotting $k`$ against the concentration L-cysteine, the second-order rate constant (k_2) for the intermolecular NAT reaction between CysCNP and L-cysteine at 37 °C was calculated from the slope of the linear fit (**Figure S8E**).

For CysCNPym, 50 μM NT_Ubi CysCNPym ΣCΣK samples in 96-well plate were incubated with serial concentration of different ABTs (0.1–1.6 mM) in PBS buffer containing 30 mM

TCEP at 25 °C. The increase in fluorescence signal upon reaction (for ABT **1**: λ_{ex} = 365 nm, λ_{em} = 460 nm; ABT **2**: λ_{ex} = 380 nm, λ_{em} = 490 nm; ABT **6**: λ_{ex} = 450 nm, λ_{em} = 590 nm) was fitted to an exponential growth equation to determine the observed rate constants ($k`$) (**Figures S8B-8D)**. Subsequently, the second-order rate constants ($k_2$) for these reactions at 25 °C were obtained by plotting $k`$ versus the concentration of each ABT (**Figures S8F-8H**).

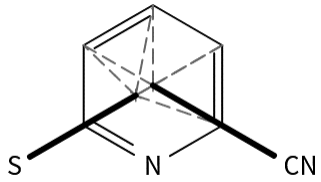


Figure S1. SciFinder query structure used to survey available nitrile-bearing pyridine-thiols. The query using variable attachment points returned 15 potential pyridine-thiols. Filtering by commercial availability returned 10 compounds. These compounds were manually assessed to identify pyridine-thiols with the nitrile group in the 2- or 4-position of the pyridine ring and for in-stock availability by the commercial supplier.

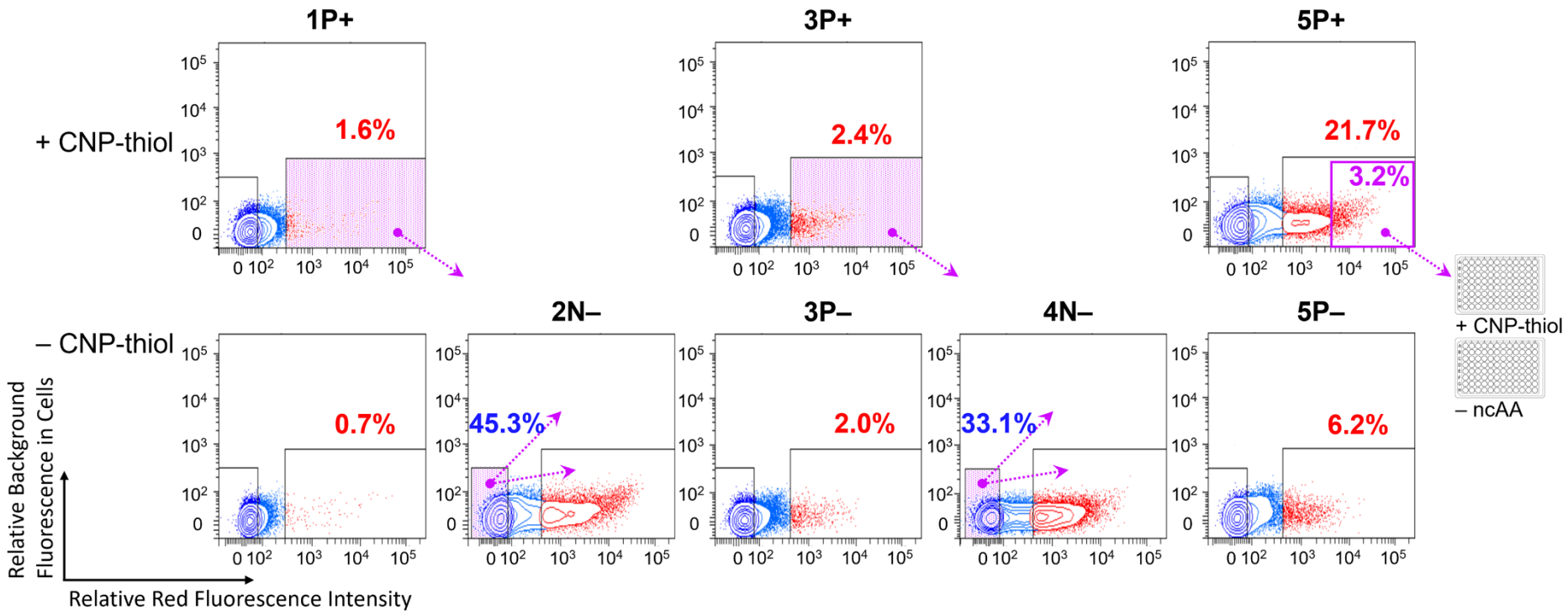


Figure S2. FACS screening of G1PylRS variants for activity and specificity in recognizing CysCNP. The horizontal axis of the scatter plots represents red fluorescence intensity (excitation at 560 nm), while the vertical axis indicates background fluorescence in cells excited at 488 nm. Positive (**P**) and negative (**N**) selection rounds are indicated with "+", denoting growth conditions with 0.5 mM CNP-thiol, and "-" for conditions without CNP-thiol. Violet-shaded regions highlight the collected cell populations. Arrows illustrate the subsequent selection strategy applied after amplification by culturing.

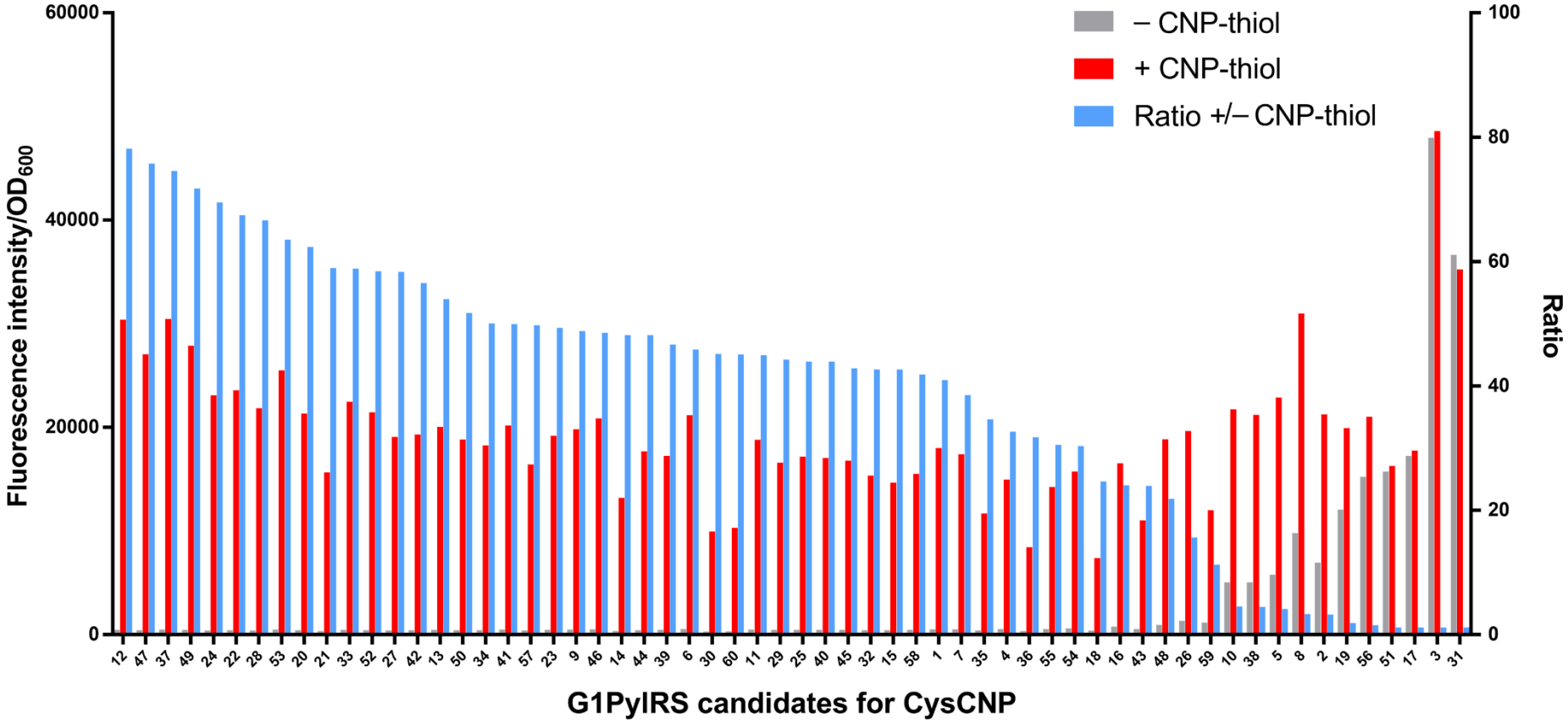


Figure S3. Activity and specificity screen of G1PylRS variants for CysCNP incorporation. Cells from the 3.2% fraction with the highest red fluorescence in the final selection round were cultured in 96-well plates with and without 0.5 mM CNP-thiol. Red fluorescence intensity indicative of the readthrough efficiency of the amber-interrupted reporter gene was then measured. The plot presents the colonies ranked in a descending order based on the ratio of red fluorescence in the + CNP-thiol wells compared to the – CNP-thiol wells. This ranking highlights the candidates with the highest activity and specificity for CysCNP incorporation.

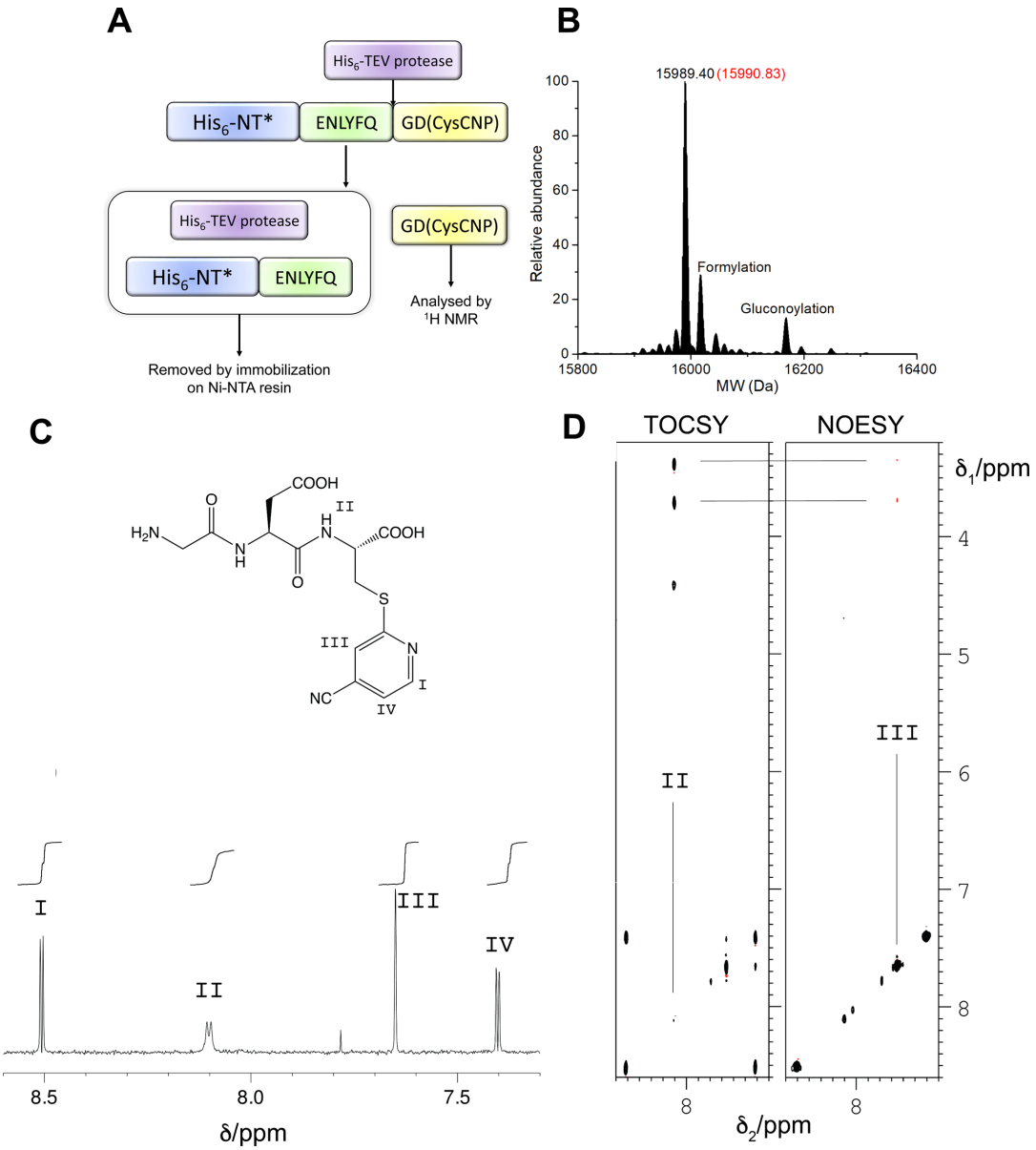


Figure S4. Confirmation of the biosynthesis and genetic encoding of CysCNP into proteins. **(A)** NT*_GDX construct used to produce the GD(CysCNP) tripeptide and the protocol of the isolation of the tripeptide. **(B)** Intact protein mass spectrometric analysis of CysCNP incorporation into the NT*_GDX construct. The observed and expected masses are indicated in black are red, respectively. **(C)** Chemical structure and 1D ^{1}H NMR spectrum of the tripeptide. Roman numbers refer to proton assignments in the NMR spectra. The 1D ^{1}H NMR spectrum was recorded for a 0.65 mM solution of the tripeptide in PBS buffer containing 10% D_2O. The spectrum was recorded on a Bruker 800 MHz NMR spectrometer at 25 ºC, using a double spin-echo for solvent suppression. The integral of the amide proton is attenuated by amide proton exchange. **(D)** 2D TOCSY and NOESY spectra of the tripeptide sample recorded under the same conditions as the 1D ^{1}H NMR spectrum in **c**. The TOCSY spectrum was recorded with a mixing time of 60 ms. The NOESY spectrum was recorded with a mixing time of 500 ms. Horizontal lines link the cross-peaks observed with the $C^{b}H_2$ group of the Cys-CNP residue. TOCSY cross-peaks link the aromatic protons I, III, and IV of the Cys-CNP residue.

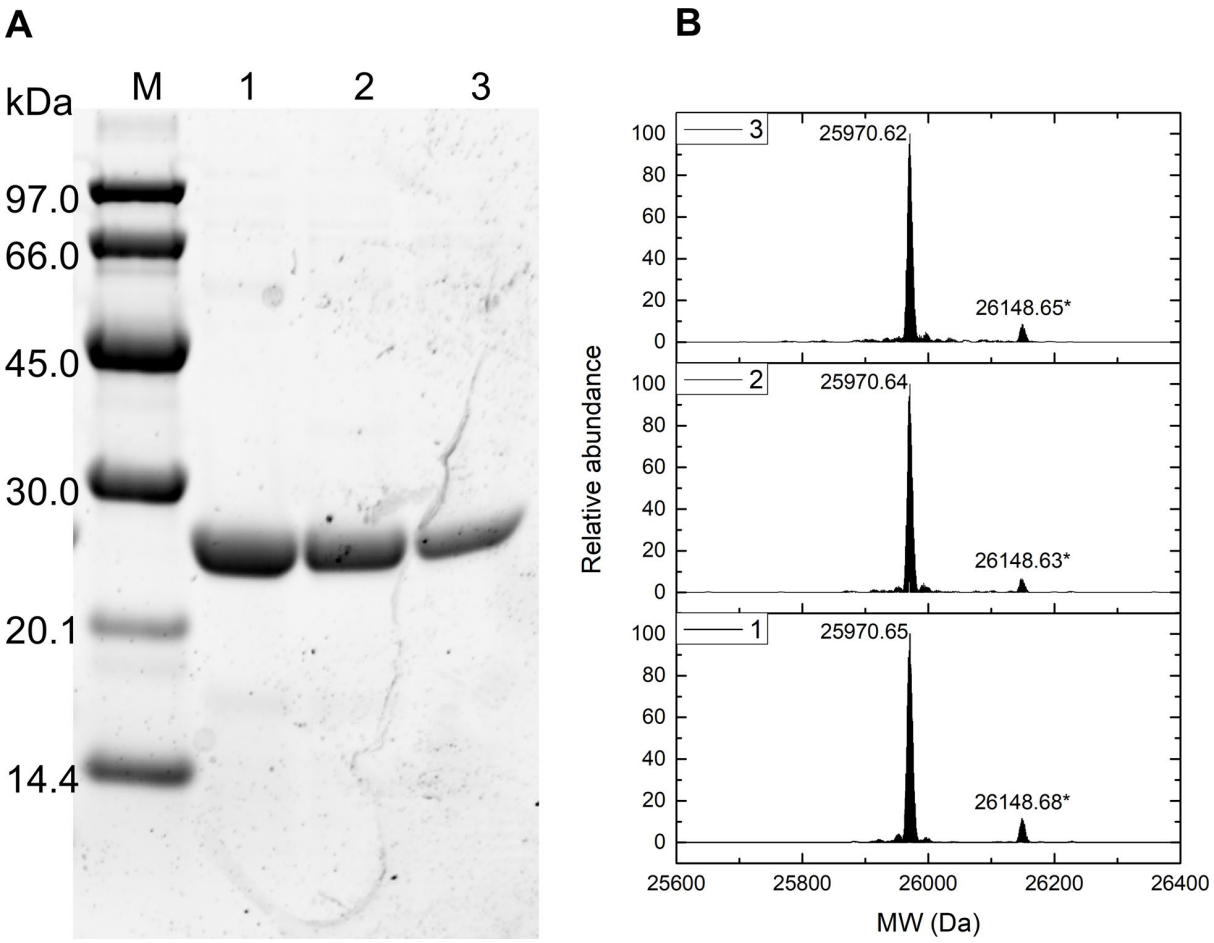


Figure S5. Analysis of NT_Ubi CysCNP produced using high cell-density fermentation. **(A)** Reducing SDS-PAGE analysis of the purified protein. M, protein molecular weight marker (the molecular weight of each band is indicated on the left). Proteins purified from each of three fermentation are shown in lanes 1–3. **(B)**, Intact protein mass spectrometry analysis of the proteins shown in lanes 1–3. The expected mass of NT_Ubi CysCNP is 25971.21 Da. The spectra show a minor peak at +178 Da (marked with an asterisk), which corresponds to N-terminal gluconoylation [S9].

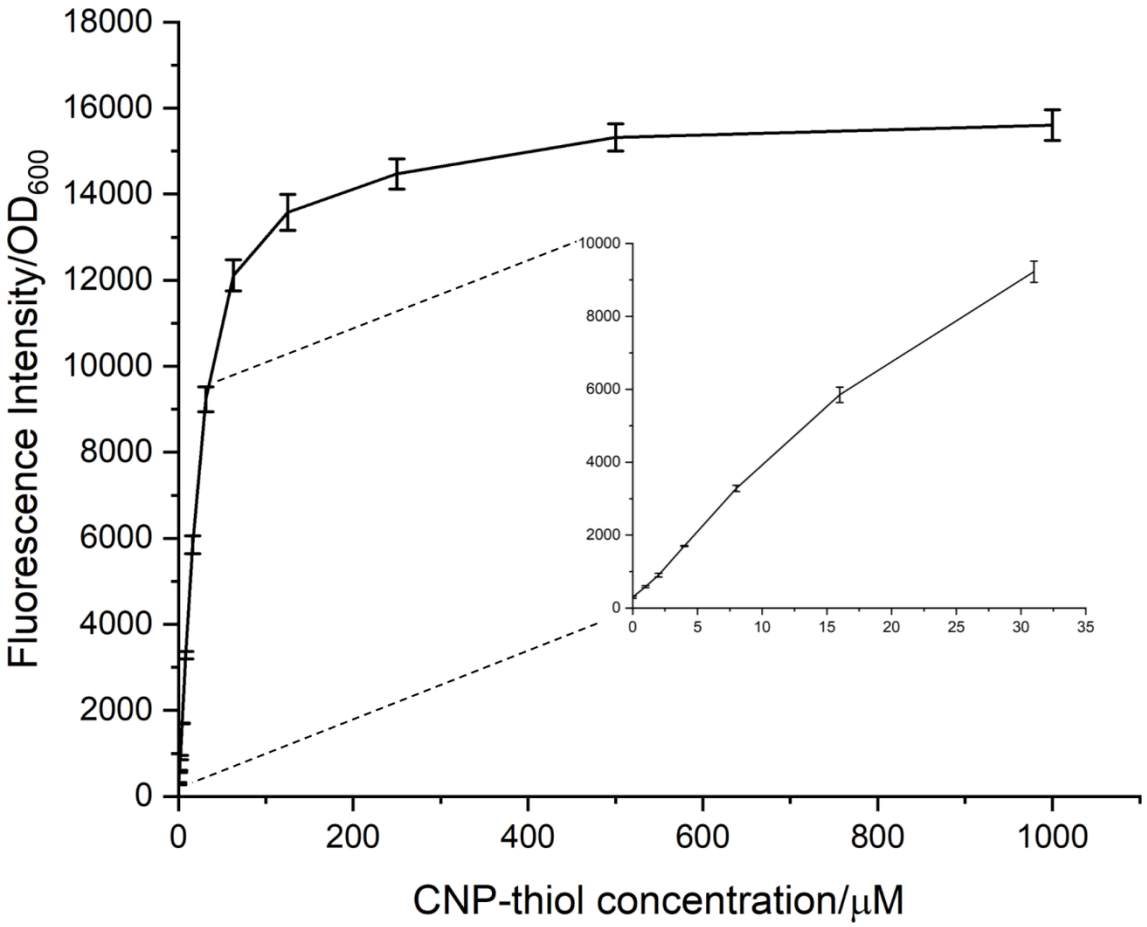


Figure S6. Effect of CNP-thiol concentration on the expression yield of RFP CysCNP. RFP fluorescence is normalized by the optical density (OD_{600}) and reported as the mean of three biological replicates ± standard deviation.

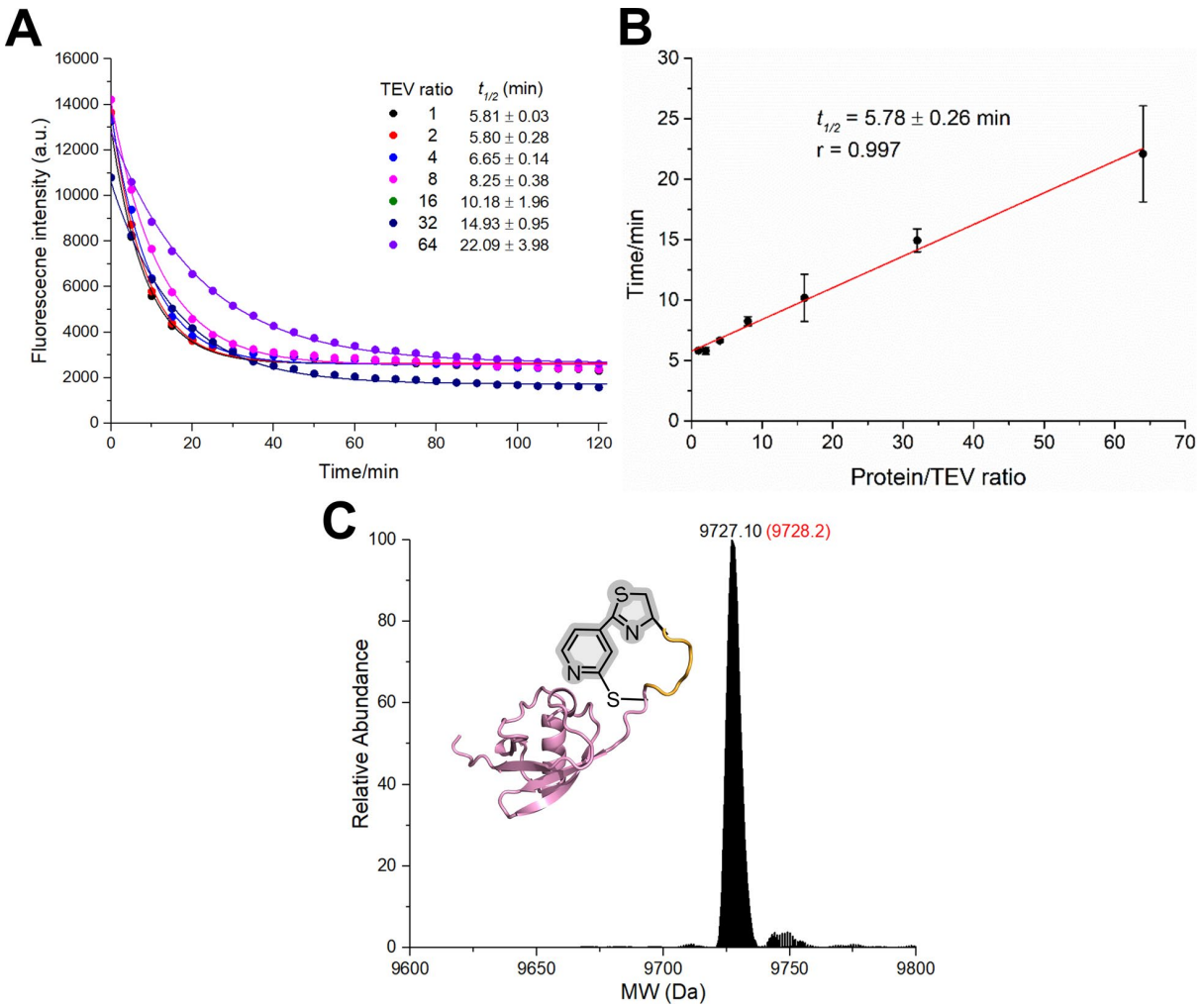


Figure S7. Kinetic analysis and determination of the half times $t_{1/2}$ of the intramolecular NAT macrocyclization reaction. **(A)** 50 µM NT_Ubi CysCNP samples were incubated with TEV protease at different serial protein/TEV protease ratio (1-64) in PBS buffer containing 30 mM TCEP at 37 °C, and the fluorescence was measured at λ_{ex} = 330 nm and λ_{em} = 400 nm every 1 min (data are plotted for the 5 min measurement points for simplicity). $t_{1/2}$ values for each protein/TEV protease ratio were determined by fitting the data to a single exponential equation. The $t_{1/2}$ are shown as the mean of 3 different measurements ± standard deviation. **(B)** plot of the half times $t_{1/2}$ versus protein/TEV protease ratio used to determine the intramolecular NAT reaction $t_{1/2}$, which is calculated from the zero ratio intercept. **(C)** Intact protein mass spectrometry analysis of the NAT macrocyclization reaction product. The mass spectrum confirms the formation of the shown macrocyclic Ubi product. The observed and expected masses are indicated in black are red, respectively.

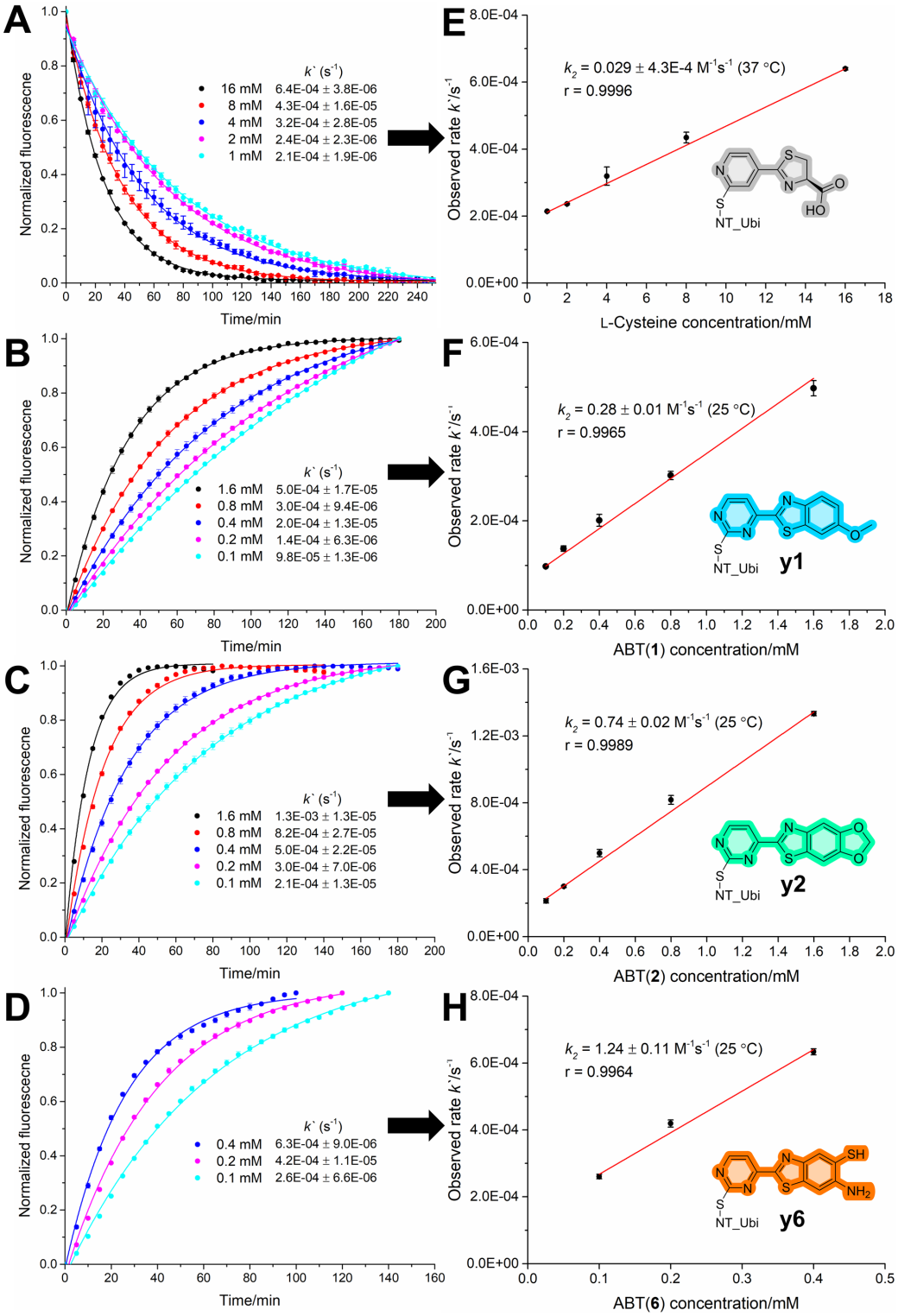
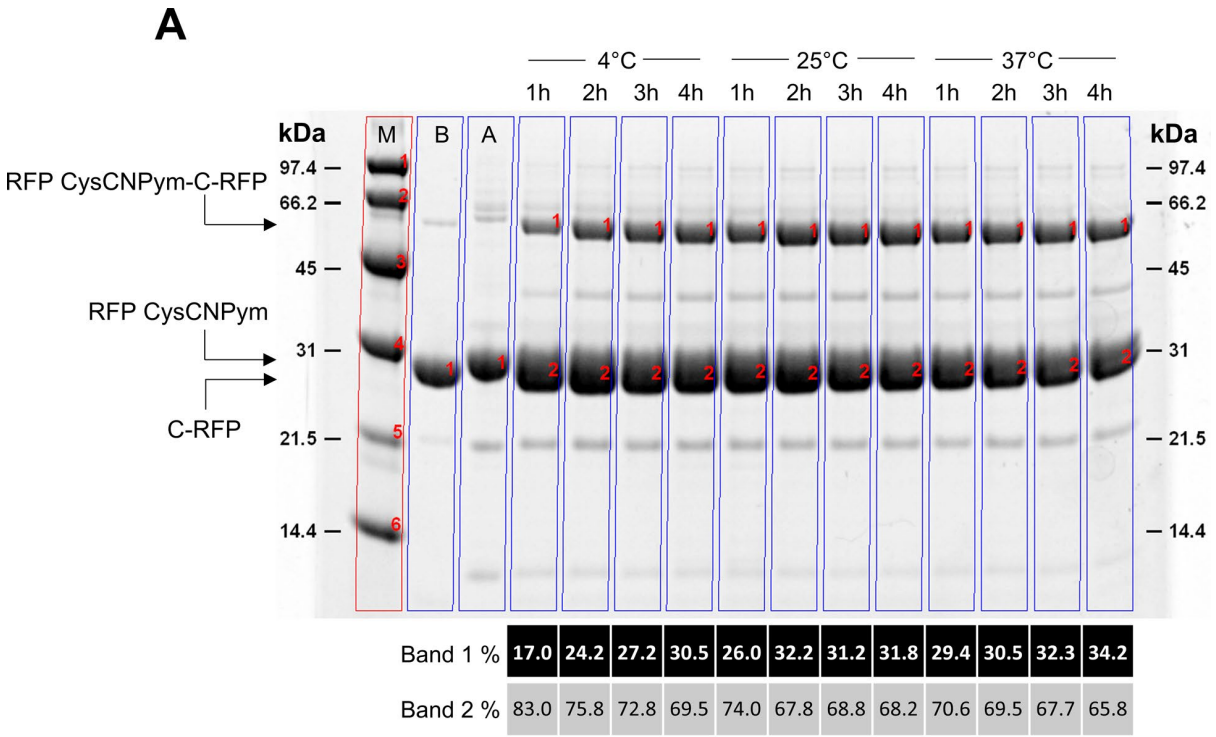


Figure S8. Kinetic analysis of the intermolecular NAT click reaction using fluorescence spectroscopy. **(A)–(D)** Rate constants (k`) between NT_Ubi CysCNP **(A)** or NT_Ubi CysCNPym ΣCΣK **(B–D)** and different 2-aminothiols (L-cysteine, ABT **1**, ABT **2**, ABT **6** for panels **(A)-(D)**, respectively) were determined by monitoring the change in fluorescence over time. The plots show the direct correlation between the observed rate constant (k`) and the concentration of the 2-aminothiol. **(E)–(H)**, plots of the observed rate constant $k$` versus the 1,2-aminothiol concentration used in the determination of the second-order rate constant k_2. Data are represented as the mean of three independent measurements ± standard error.



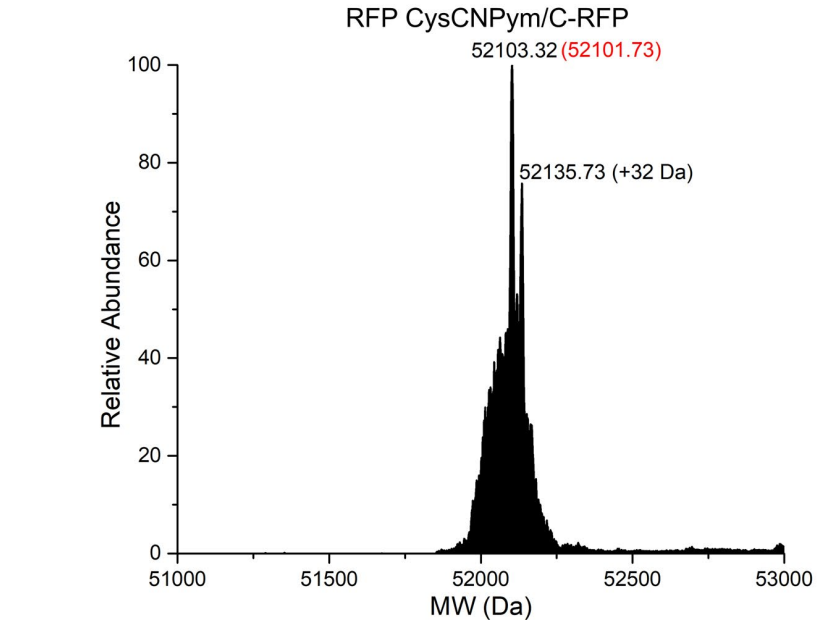


Figure S9. Formation of RFP CysCNPym/C-RFP conjugate via NAT click reaction. **(A)** SDS-PAGE analysis of the reaction between RFP CysCNPym (lane A, 26.8 kDa) and C-RFP (lane B, 25.4 kDa) at different temperatures (4 °C, 25 °C and 37 °C). The table below the gel presents the quantification of each band's volume relative to the total volume of bands 1 (RFP CysCNPym-C-RFP conjugate, 52.1 kDa) and 2 (unreacted RFP CysCNPym and C-RFP). The red numbers indicate the corresponding bands in each lane. Lane M shows the protein molecular weight markers. **(B)** Intact protein mass spectrometry analysis. The mass spectrum confirms the formation of the RFP CysCNPym/C-RFP conjugate (expected mass indicated in red). A minor peak observed at +32 Da corresponds to protein oxidation (+ 2 oxygen atoms).

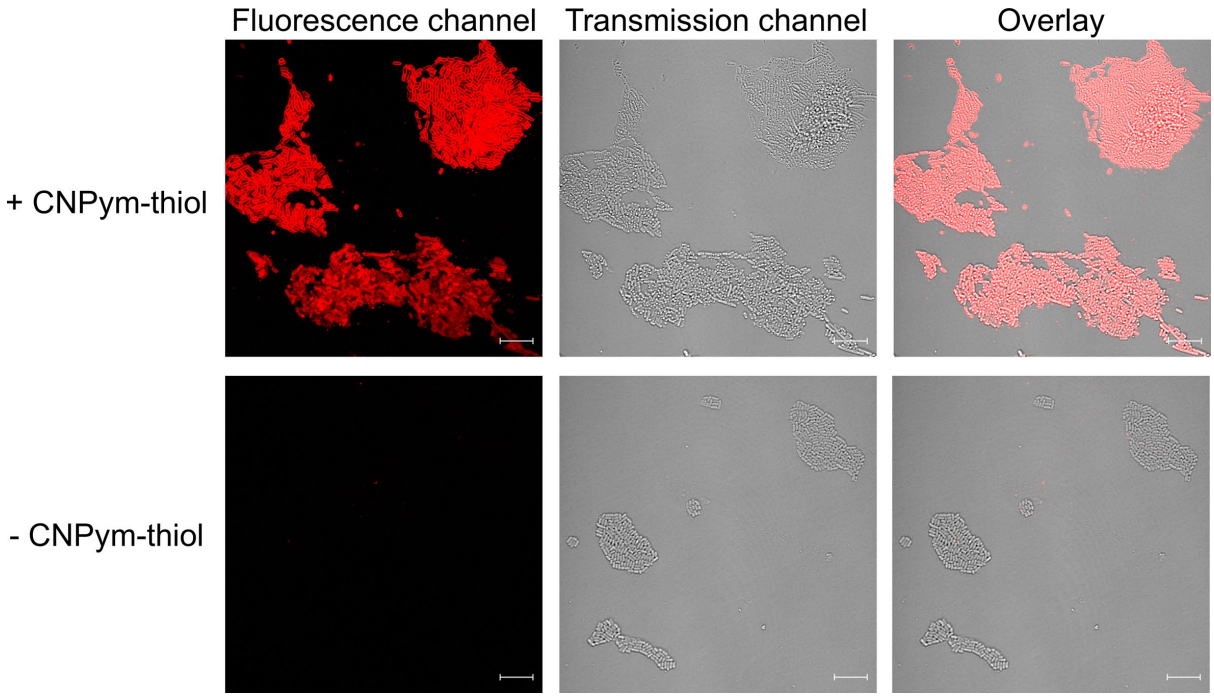


Figure S10. Uncropped confocal fluorescence microscopy images of live *E. coli* cells displaying OmpX CysCNPym after labelling with 200 μM C-RFP for 4 h at 25 °C. The top and bottom panels show cells grown and induced in culture media supplemented with or without 0.5 mM CNPym-thiol, respectively, followed by incubation with C-RFP (scale bar = 10 μm).

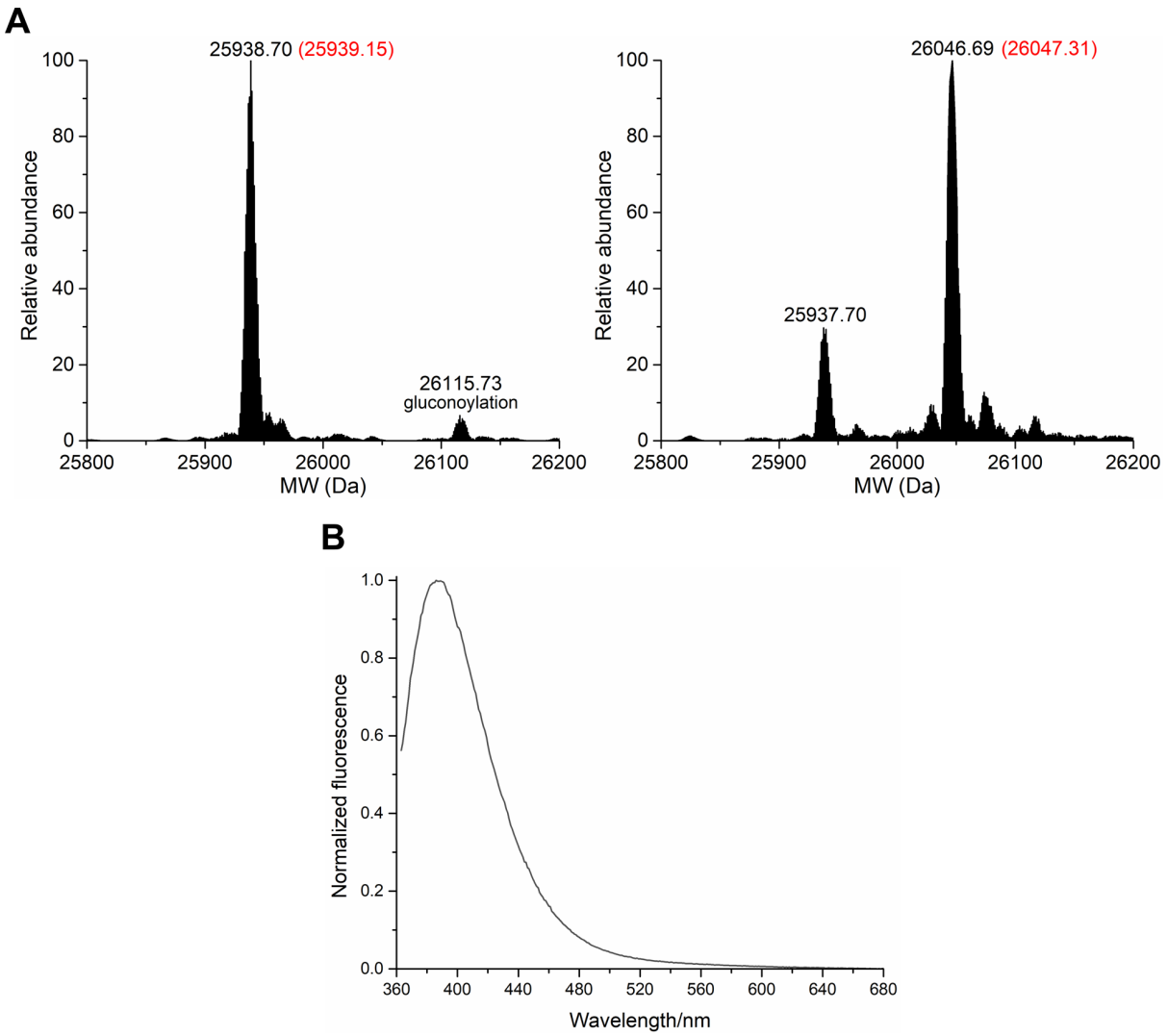


Figure S11. The NAT click reaction between NT_Ubi mCNP and ABT. **(A)** Intact protein mass spectrometry analysis of NT_Ubi mCNP (left panel) and its NAT reaction product with ABT (right panel). 100 μM NT_Ubi mCNP was incubated with 20 mM ABT in PBS buffer containing 30 mM TCEP for 16 h at 37 °C. The expected mass is indicated in red. **(B)** Fluorescence spectra of NT_Ubi mCNP-ABT product in PBS buffer after removing the unreacted ABT by ultrafiltration (λ_{Ex} = 320 nm).

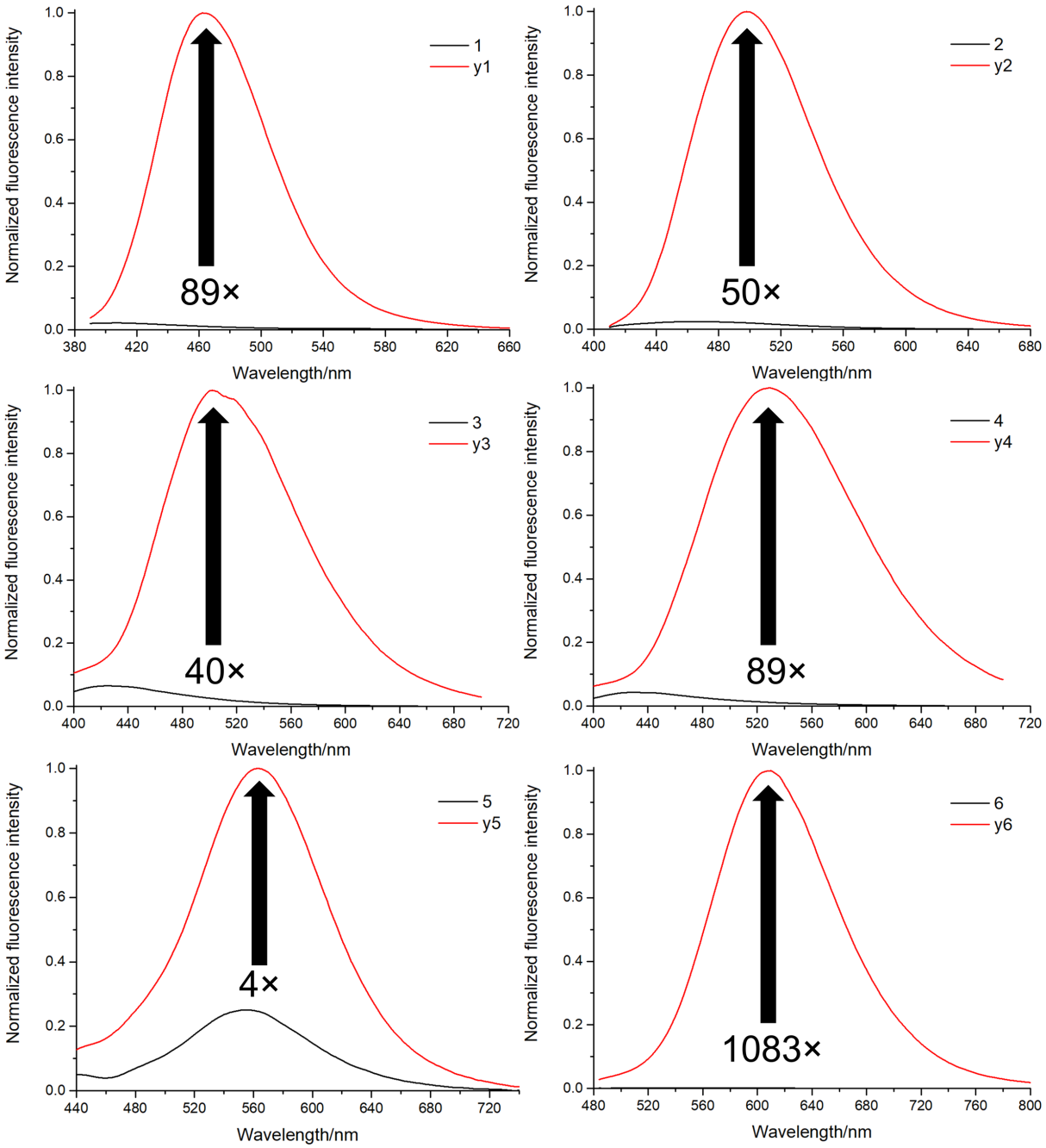


Figure S12. Quantification of the ON/OFF ratio in the arcNAT click reaction between NT_Ubi CysCNPym ΣCΣK and different ABTs. 1 mM ABTs **1–6** were incubated without (black spectra, OFF) or with (red spectra, ON) 100 µM NT_Ubi CysCNPym ΣCΣK in PBS buffer containing 30 mM TCEP at 37 °C for 4 h. Fluorescence emission spectra were then recorded with excitation at the respective excitation maxima ($^{max}\lambda_{Ex}$) listed in **Table S3**. The ON/OFF ratio, representing the ratio of fluorescence intensity at the maximum emission wavelengths ($^{max}\lambda_{Em}$) in the presence versus absence of protein, are indicated on the spectra.

Figure S13. Snapshots from Video S1: Progression of the fluorogenic arcNAT click reaction in real time. The snapshots were extracted at 30 s intervals for 180 s.

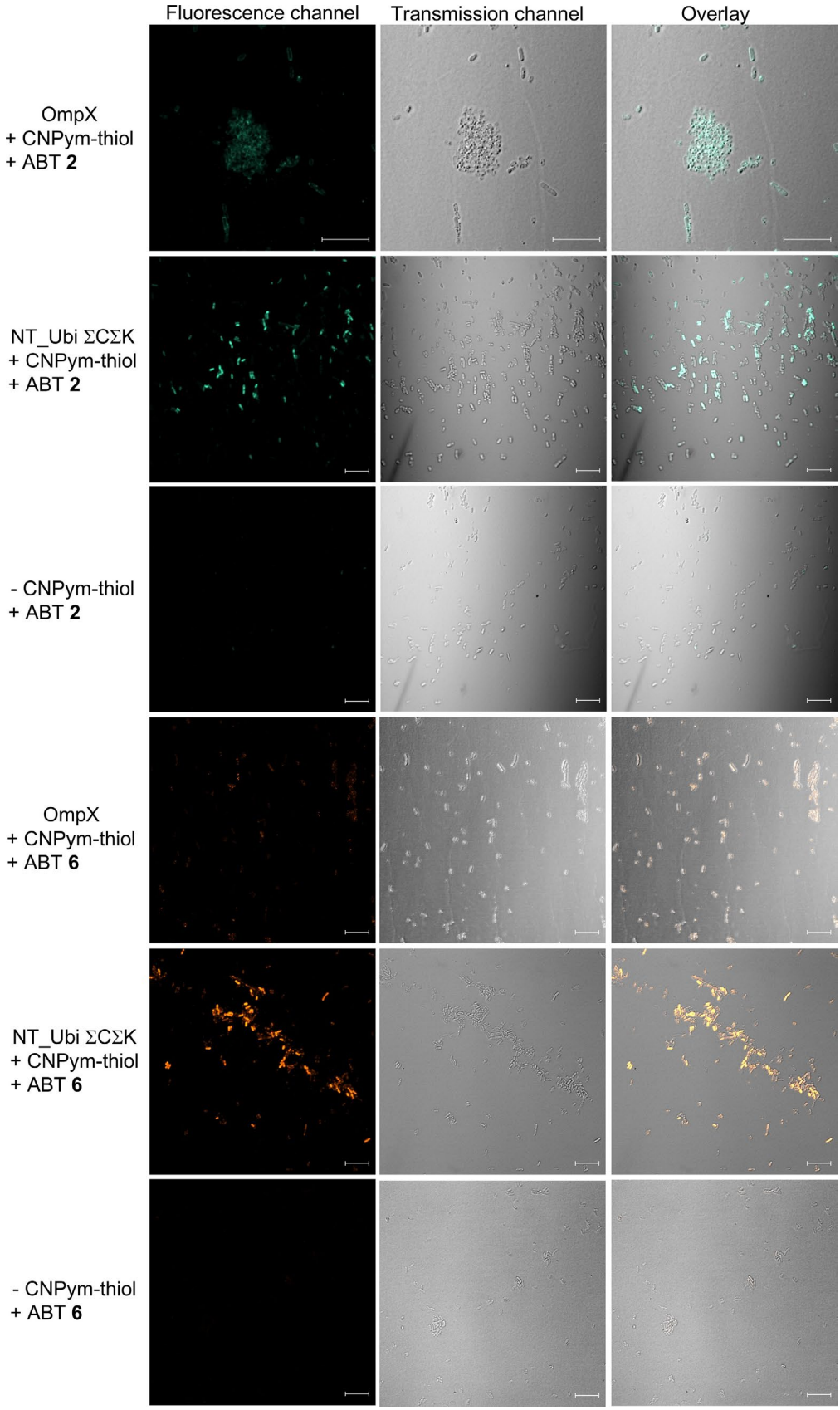
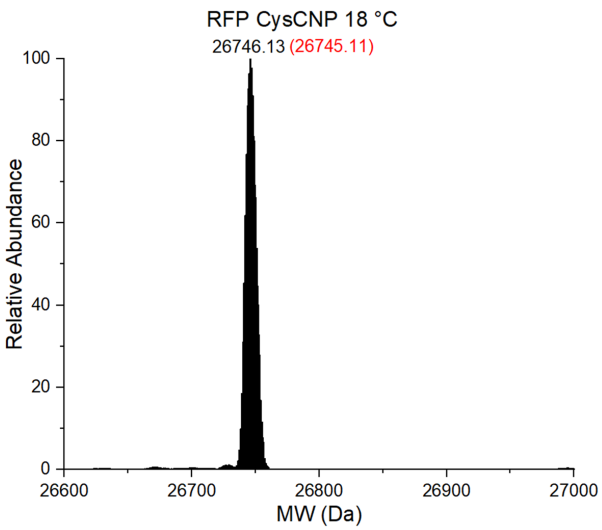
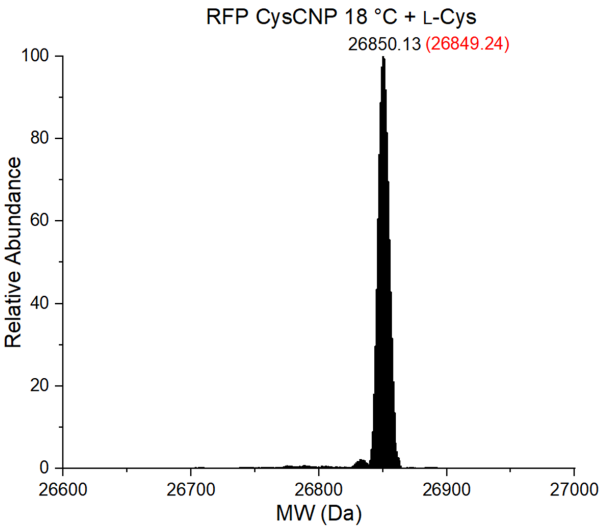
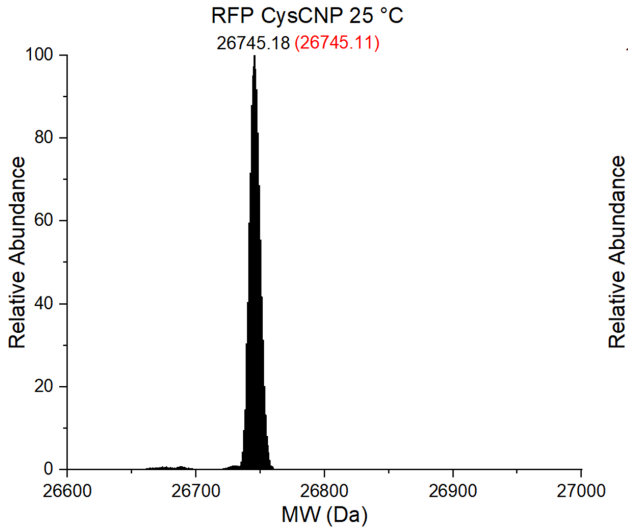
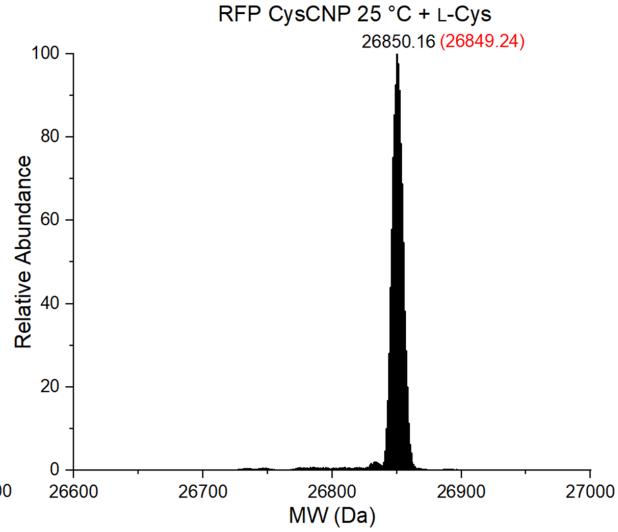


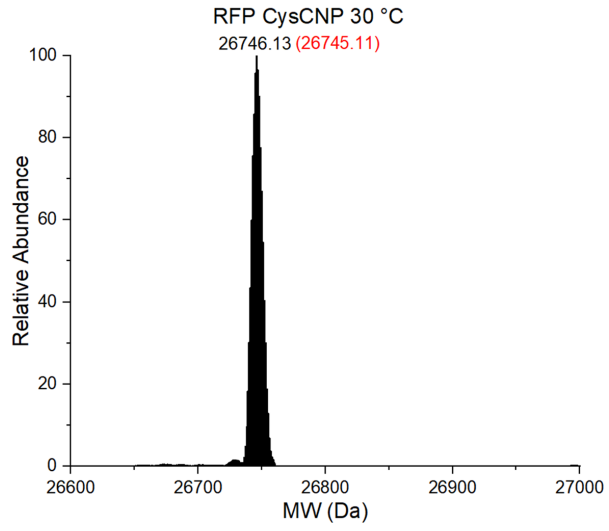
Figure S14. Uncropped confocal fluorescence microscopy images of the reaction of live *E. coli* cells displaying OmpX CysCNPym or expressing NT_Ubi CysCNPym ΣCΣK, or negative control cells (without supplementing CNPym-thiol) with ABT **2** or **6** (scale bar = 10 μm).

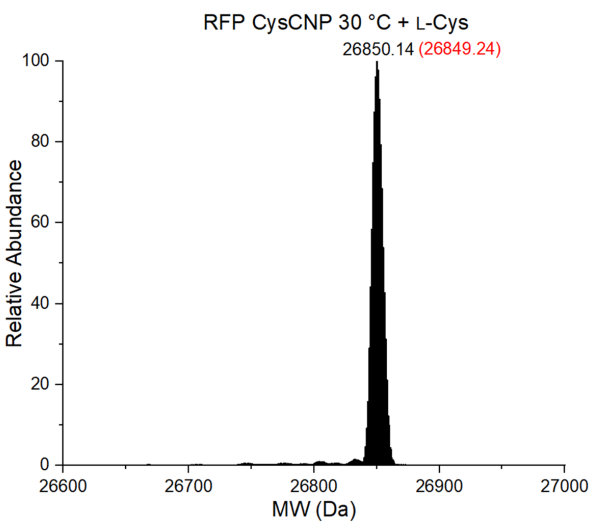

RFP CysCNP 18 °C
26746.13 (26745.11)
Relative Abundance
MW (Da)

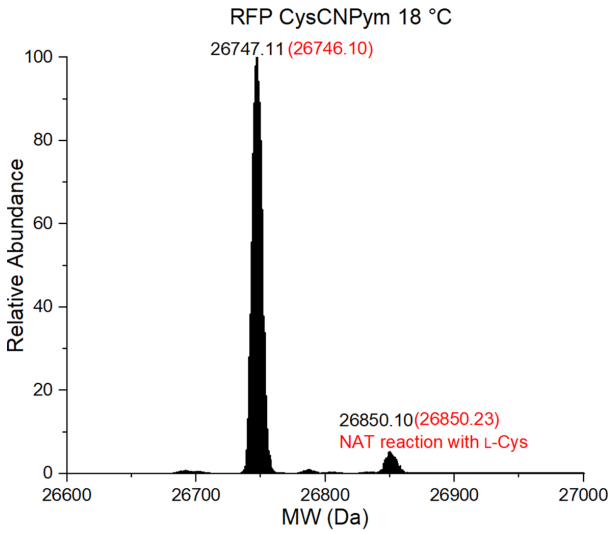

RFP CysCNP 18 °C + L-Cys
26850.13 (26849.24)
Relative Abundance
MW (Da)

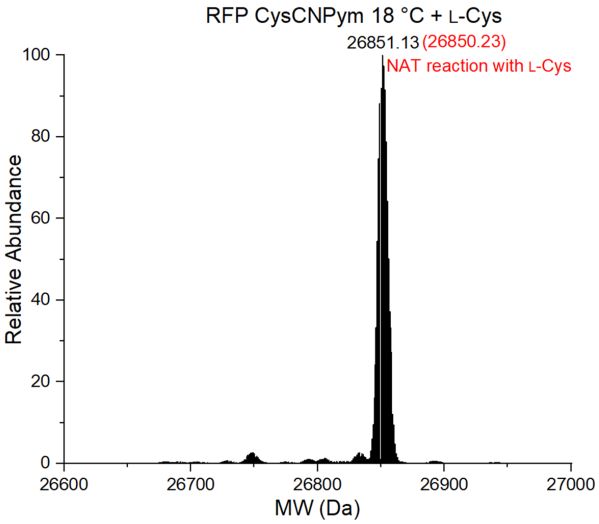

RFP CysCNP 25 °C
26745.18 (26745.11)
Relative Abundance
MW (Da)

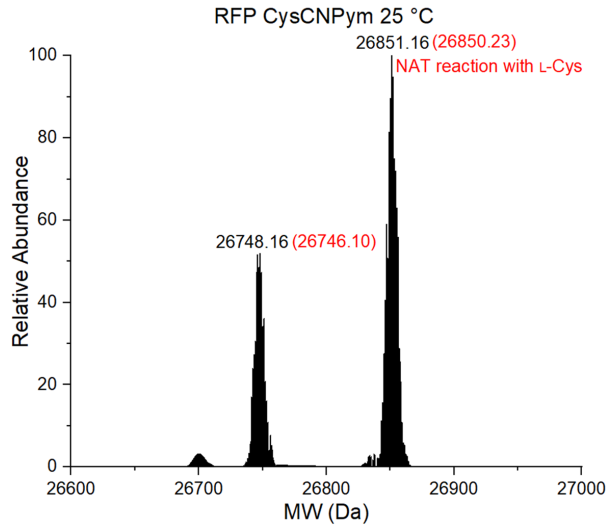

RFP CysCNP 25 °C + L-Cys
26850.16 (26849.24)
Relative Abundance
MW (Da)

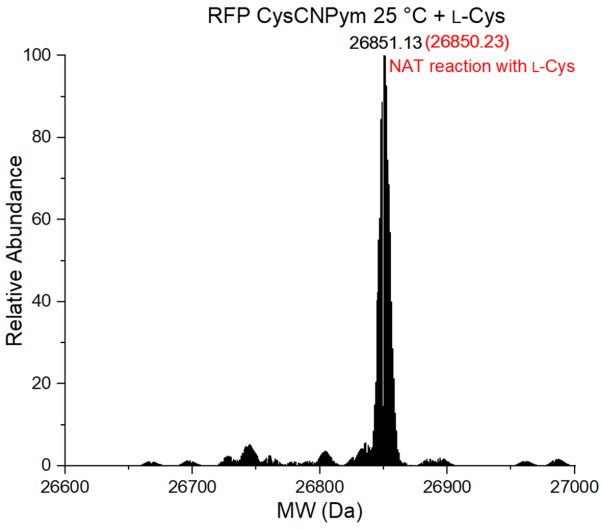

RFP CysCNP 30 °C
26746.13 (26745.11)
Relative Abundance
MW (Da)

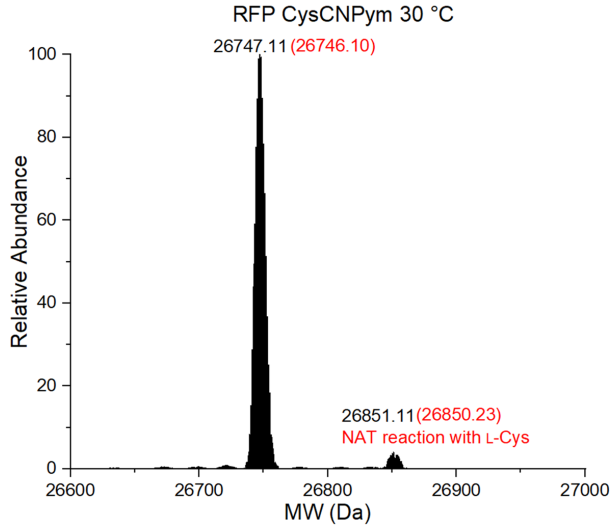

RFP CysCNP 30 °C + L-Cys
26850.14 (26849.24)
Relative Abundance
MW (Da)

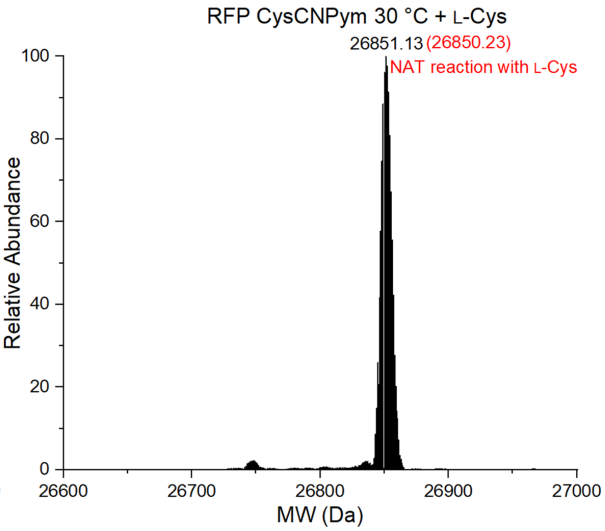

RFP CysCNPym 18 °C
26747.11 (26746.10)
26850.10 (26850.23)
NAT reaction with L-Cys
Relative Abundance
MW (Da)

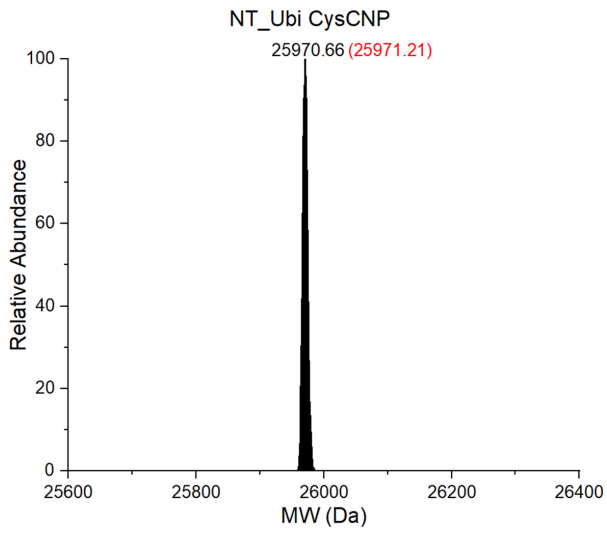

RFP CysCNPym 18 °C + L-Cys
26851.13 (26850.23)
NAT reaction with L-Cys
Relative Abundance
MW (Da)

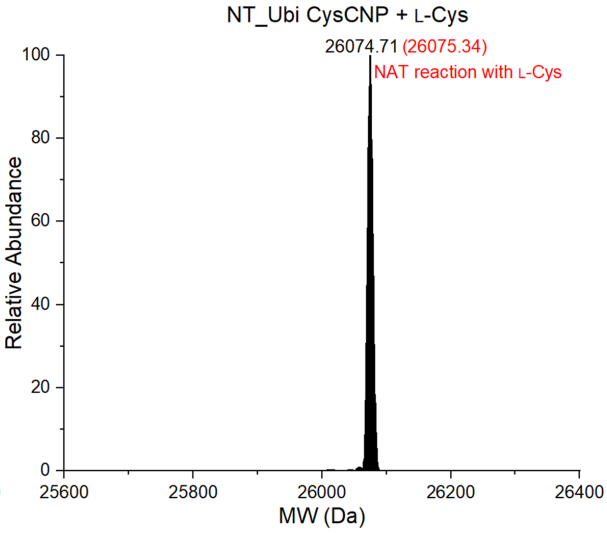

RFP CysCNPym 25 °C
26851.16 (26850.23)
NAT reaction with L-Cys
26748.16 (26746.10)
Relative Abundance
MW (Da)

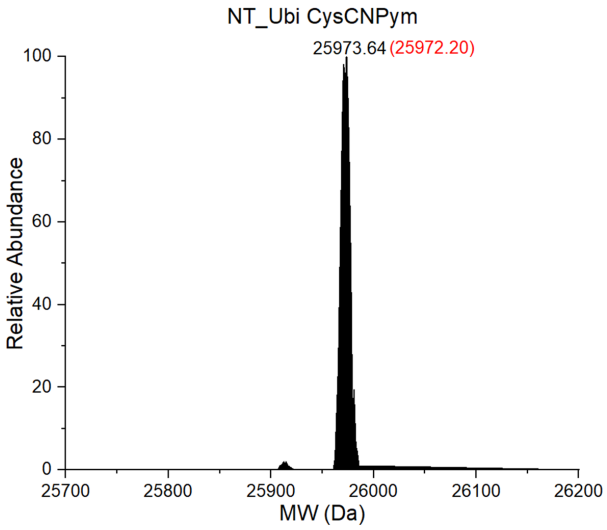

RFP CysCNPym 25 °C + L-Cys
26851.13 (26850.23)
NAT reaction with L-Cys
Relative Abundance
MW (Da)

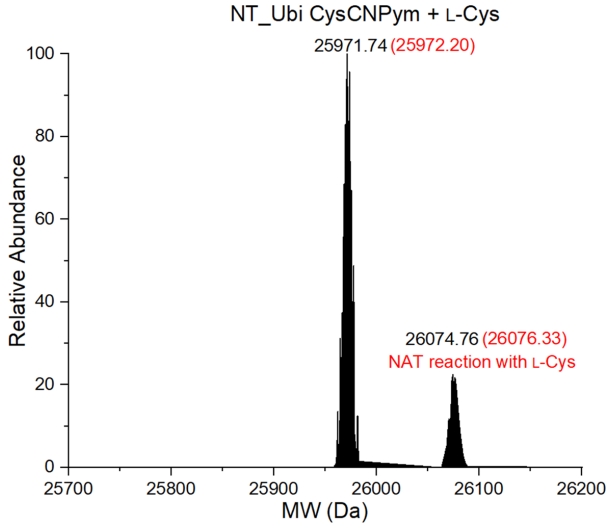

RFP CysCNPym 30 °C
26747.11 (26746.10)
26851.11 (26850.23)
NAT reaction with L-Cys
Relative Abundance
MW (Da)

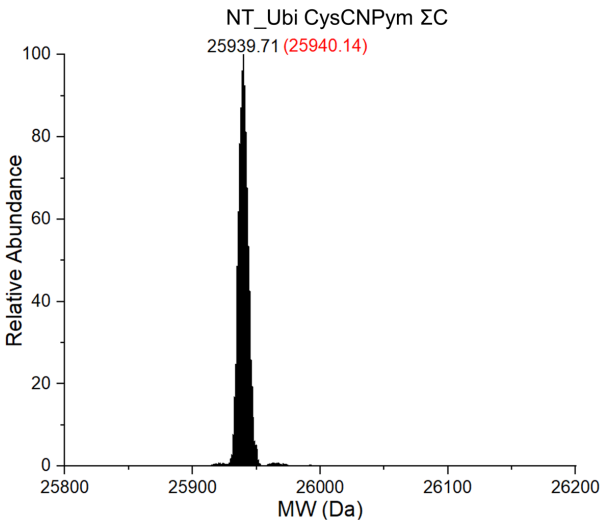

RFP CysCNPym 30 °C + L-Cys
26851.13 (26850.23)
NAT reaction with L-Cys
Relative Abundance
MW (Da)

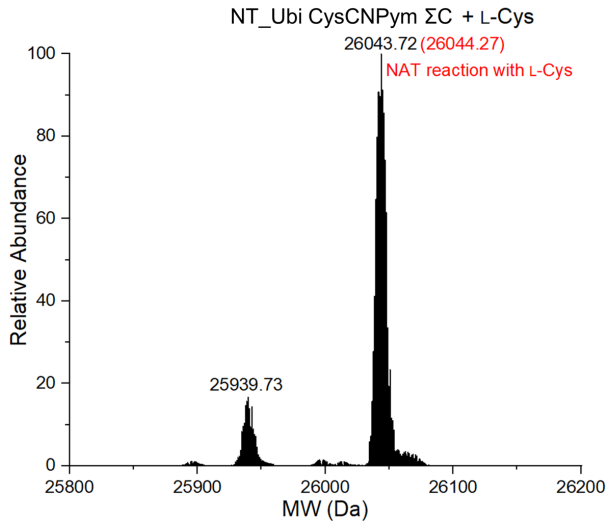

NT_Ubi CysCNP
25970.66 (25971.21)
Relative Abundance
MW (Da)


NT_Ubi CysCNP + L-Cys
26074.71 (26075.34)
NAT reaction with L-Cys
Relative Abundance
MW (Da)


NT_Ubi CysCNPym
25973.64 (25972.20)
Relative Abundance
MW (Da)


NT_Ubi CysCNPym + L-Cys
25971.74 (25972.20)
26074.76 (26076.33)
NAT reaction with L-Cys
Relative Abundance
MW (Da)


NT_Ubi CysCNPym ΣC
25939.71 (25940.14)
Relative Abundance
MW (Da)


NT_Ubi CysCNPym ΣC + L-Cys
26043.72 (26044.27)
NAT reaction with L-Cys
25939.73
Relative Abundance
MW (Da)

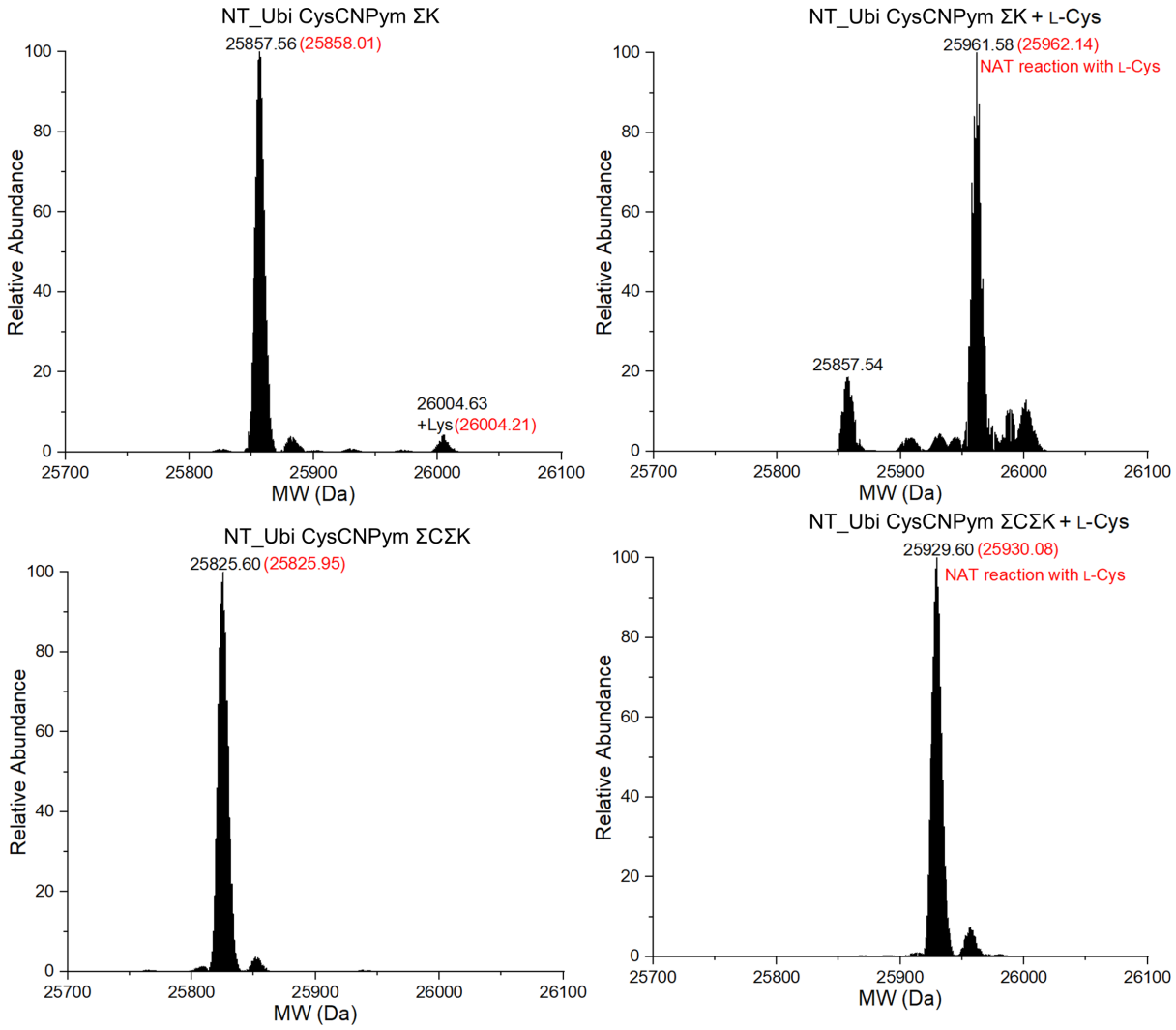
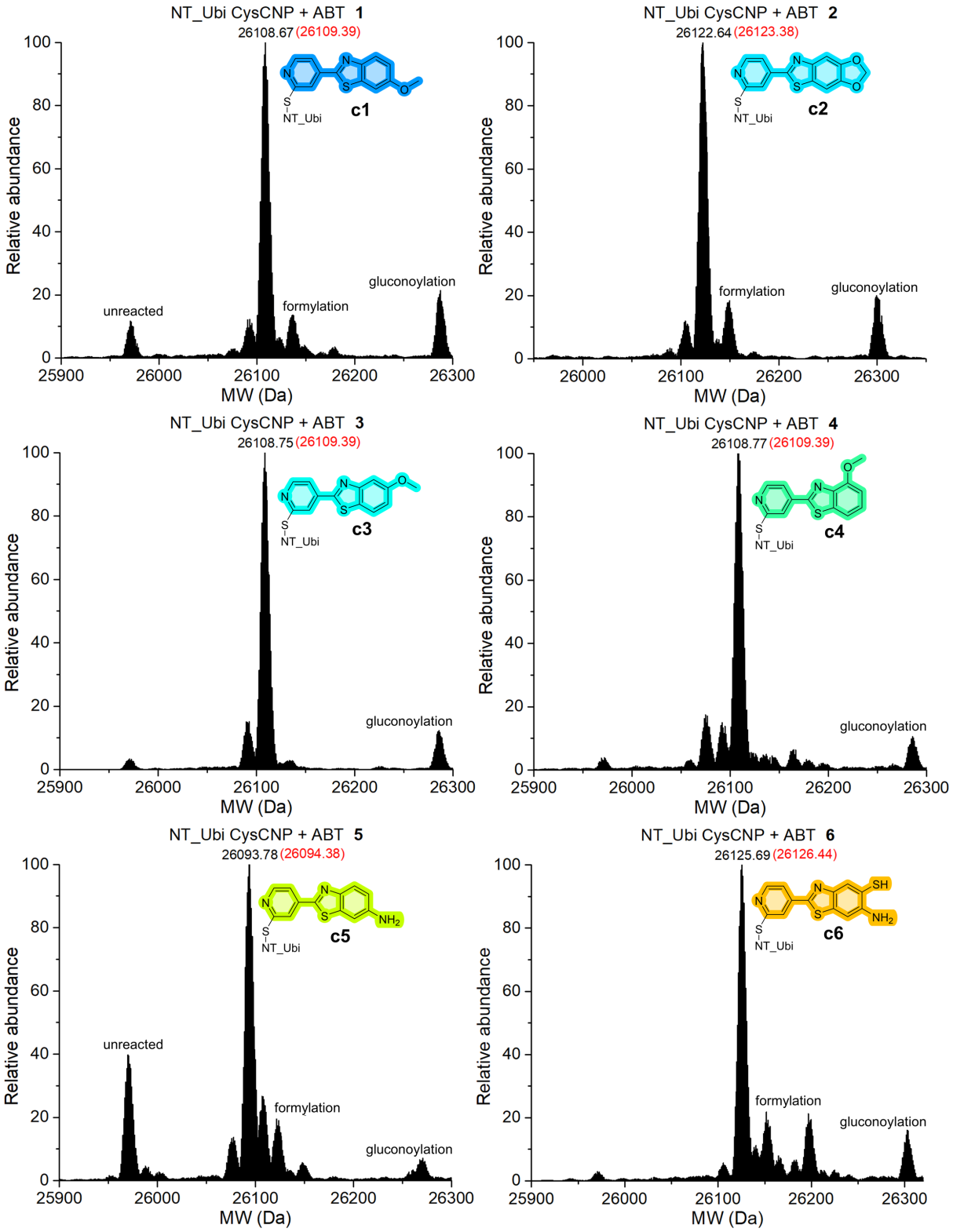
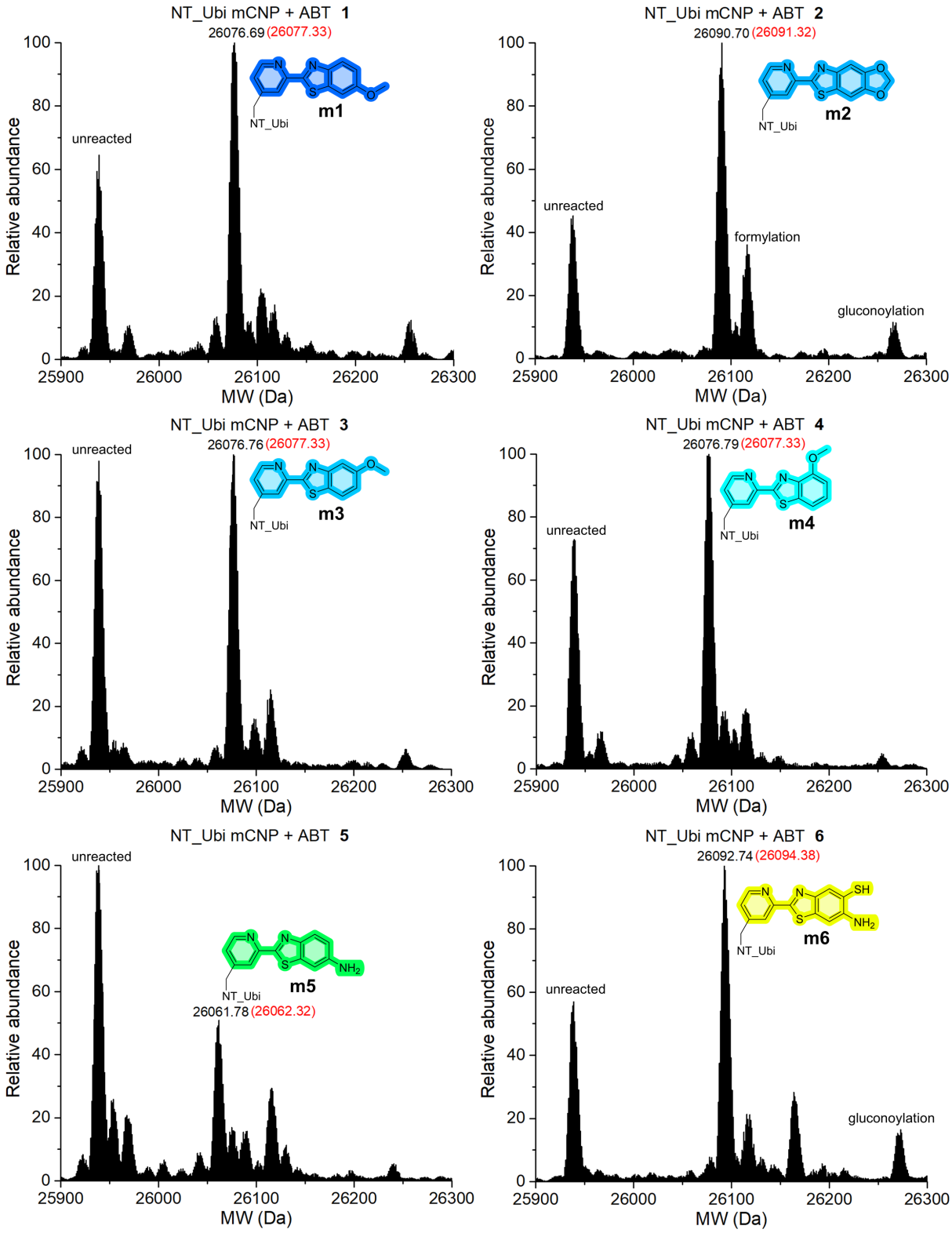
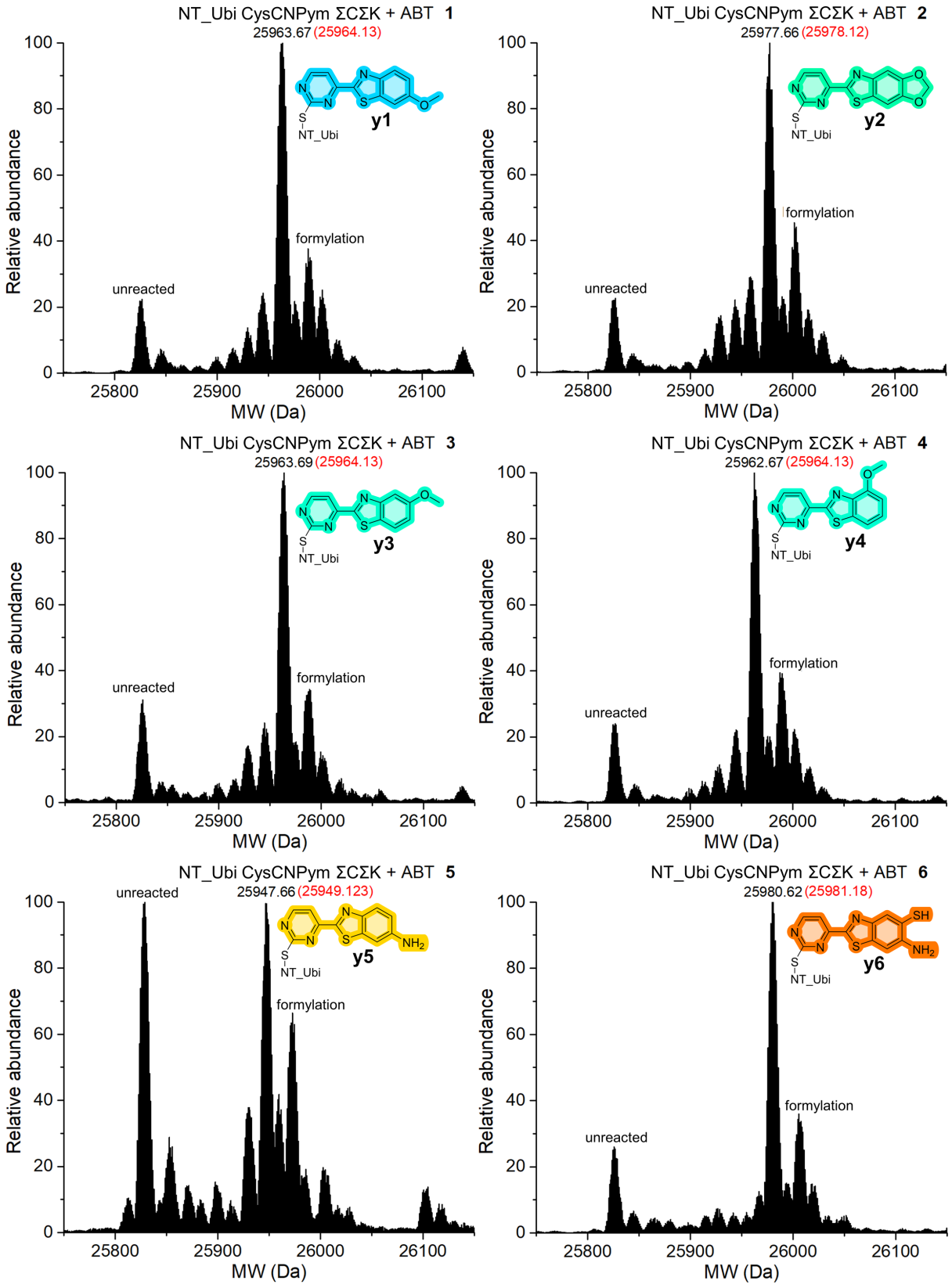


Figure S15. Individual intact protein mass spectra of the proteins shown in **Figures 1 and 2**. All RFP masses indicate the loss of the N-terminal methionine during protein expression. The observed and expected masses are indicated in black and red, respectively.


NT_Ubi CysCNP + ABT 1
26108.67 (26109.39)
c1
NT_Ubi
unreacted
formylation
gluconoylation
Relative abundance
MW (Da)
NT_Ubi CysCNP + ABT 2
26122.64 (26123.38)
c2
NT_Ubi
formylation
gluconoylation
Relative abundance
MW (Da)
NT_Ubi CysCNP + ABT 3
26108.75 (26109.39)
c3
NT_Ubi
gluconoylation
Relative abundance
MW (Da)
NT_Ubi CysCNP + ABT 4
26108.77 (26109.39)
c4
NT_Ubi
gluconoylation
Relative abundance
MW (Da)
NT_Ubi CysCNP + ABT 5
26093.78 (26094.38)
c5
NT_Ubi
unreacted
formylation
gluconoylation
Relative abundance
MW (Da)
NT_Ubi CysCNP + ABT 6
26125.69 (26126.44)
c6
NT_Ubi
formylation
gluconoylation
Relative abundance
MW (Da)


NT_Ubi mCNP + ABT 1
26076.69 (26077.33)
m1
NT_Ubi
unreacted
NT_Ubi mCNP + ABT 2
26090.70 (26091.32)
m2
NT_Ubi
unreacted
formylation
gluconoylation
NT_Ubi mCNP + ABT 3
26076.76 (26077.33)
unreacted
m3
NT_Ubi
NT_Ubi mCNP + ABT 4
26076.79 (26077.33)
m4
NT_Ubi
unreacted
NT_Ubi mCNP + ABT 5
unreacted
m5
NT_Ubi
26061.78 (26062.32)
NT_Ubi mCNP + ABT 6
26092.74 (26094.38)
m6
NT_Ubi
unreacted
gluconoylation
Relative abundance
MW (Da)
0
20
40
60
80
100
25900
26000
26100
26200
26300


NT_Ubi CysCNPym ΣCΣK + ABT 1
25963.67 (25964.13)
y1
NT_Ubi
unreacted
formylation
Relative abundance
MW (Da)
NT_Ubi CysCNPym ΣCΣK + ABT 2
25977.66 (25978.12)
y2
NT_Ubi
unreacted
formylation
Relative abundance
MW (Da)
NT_Ubi CysCNPym ΣCΣK + ABT 3
25963.69 (25964.13)
y3
NT_Ubi
unreacted
formylation
Relative abundance
MW (Da)
NT_Ubi CysCNPym ΣCΣK + ABT 4
25962.67 (25964.13)
y4
NT_Ubi
unreacted
formylation
Relative abundance
MW (Da)
NT_Ubi CysCNPym ΣCΣK + ABT 5
25947.66 (25949.123)
y5
NT_Ubi
unreacted
formylation
Relative abundance
MW (Da)
NT_Ubi CysCNPym ΣCΣK + ABT 6
25980.62 (25981.18)
y6
NT_Ubi
unreacted
formylation
Relative abundance
MW (Da)

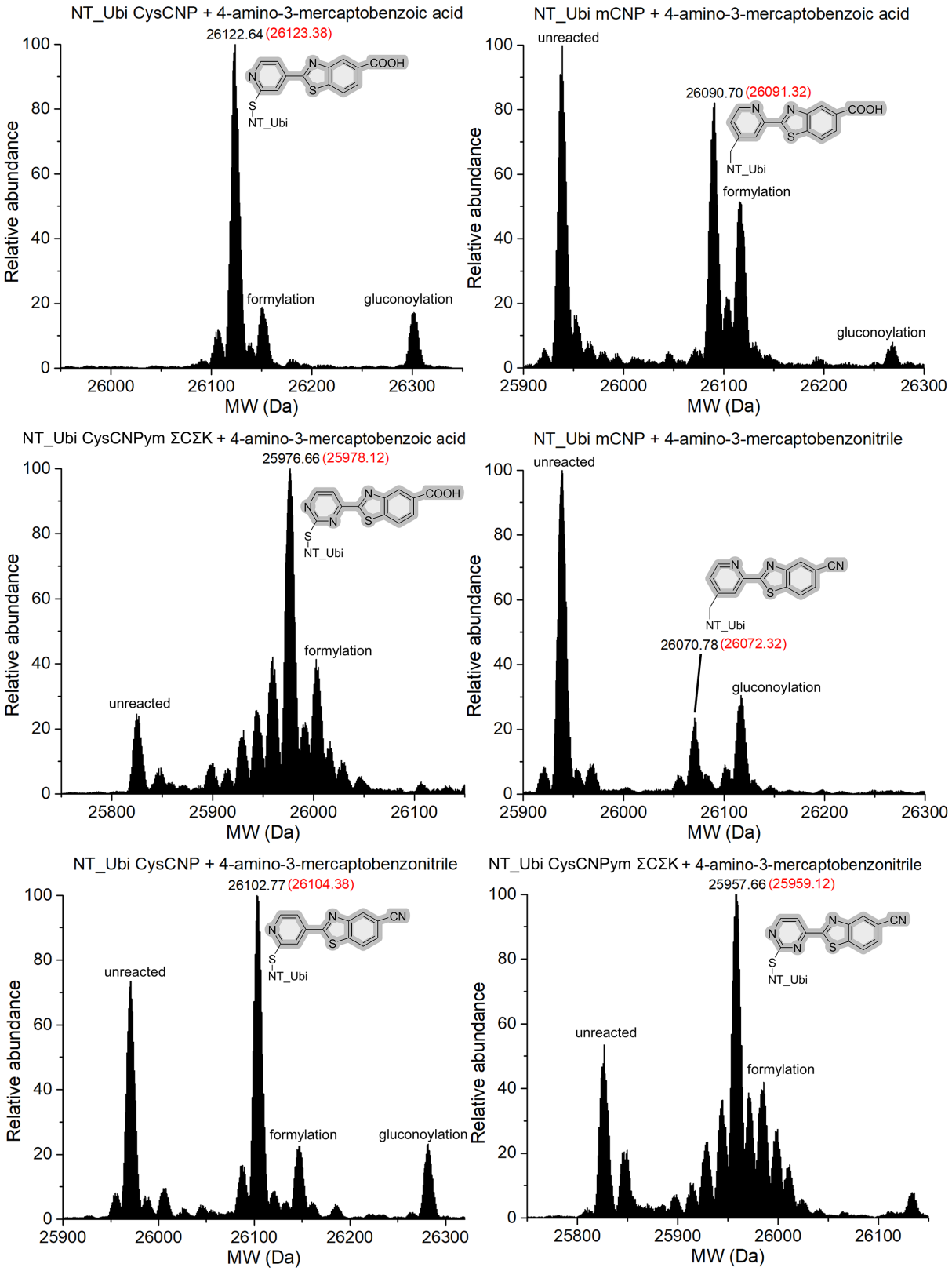


Figure S16. Intact protein mass spectrometry analysis of the arcNAT click reaction between NT_Ubi containing CysCNP (**c**), mCNP (**m**) or CysCNPym (**y**) and various ABTs. The expected chemical structure of each reaction product is depicted with each spectrum. The observed and expected masses are indicated in black and red, respectively.

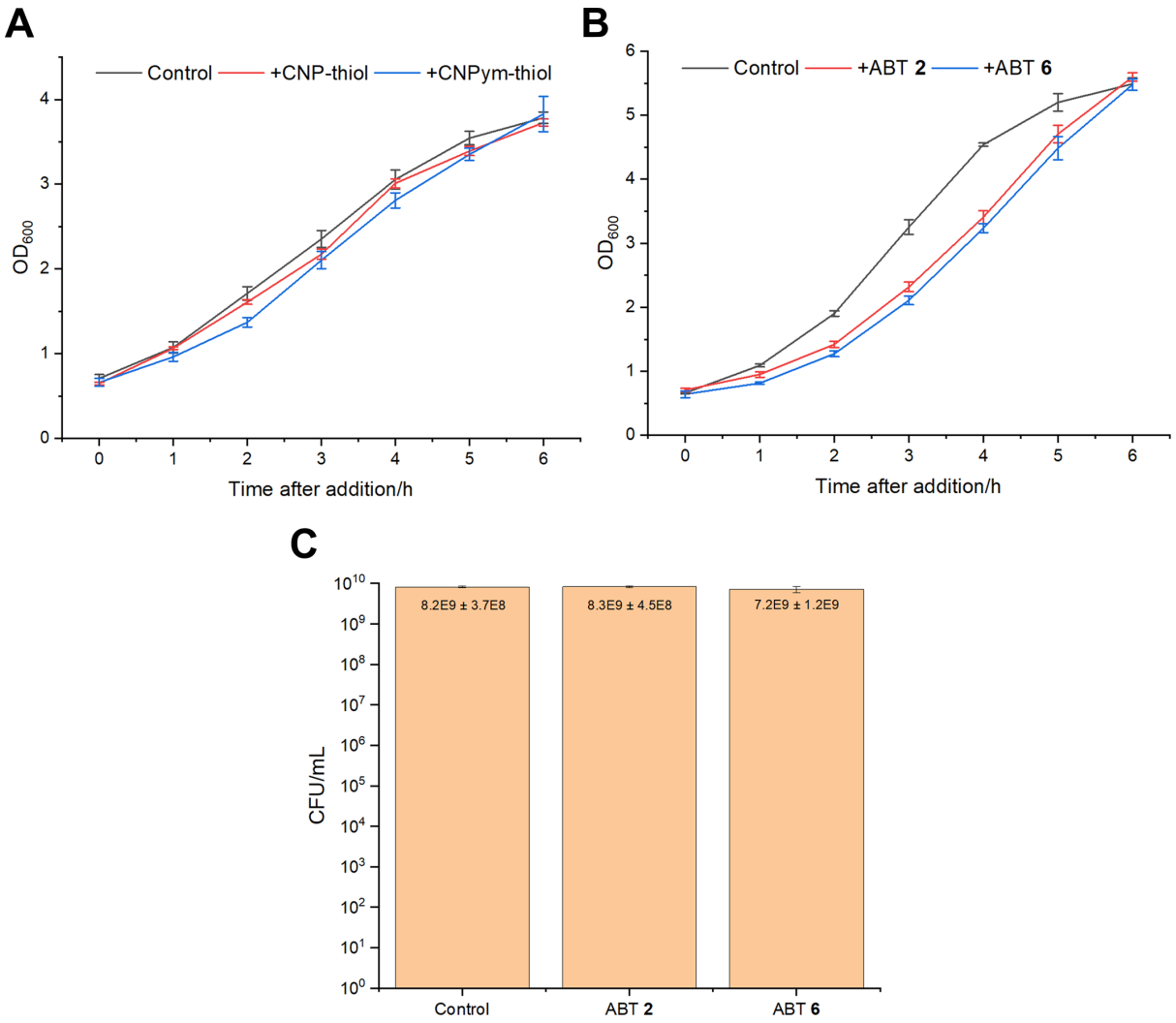


Figure S17. Evaluation of thiol cytotoxicity. **(A)** *E. coli* B-95.ΔAΔ*fabR* cells harbouring the pRSF/pCDF dual plasmid system were grown in LB media containing 25 mg/L kanamycin and 25 mg/L spectinomycin at 37 °C until reaching an OD_{600} of 0.6–0.7. Subsequently, 0.5 mM CNP-thiol or CNPym-thiol were added (DMSO serving as a control) and the cell cultures were shaken at 30 °C. OD_{600} measurements were taken every 1 h for 6 h (mean ± SD, n=3). **(B)** Experiment conducted as in **(A)** but using 0.5 mM ABT **2** or ABT **6**. **(C)** Cell survival following live-cell labelling by the NAT click reaction. *E. coli* B-95.ΔAΔ*fabR* cells expressing NT_Ubi CysCNPym ΣCΣK were subjected to the arcNAT reaction with ABT **2**, ABT **6** or PBS (control). After the reaction, cell suspensions were adjusted to OD_{600} = 2, diluted appropriately and plated on LB agar plates containing 50 mg/L kanamycin and 50 mg/L spectinomycin. Colony forming units (CFU) were determined by scanning the plates with an Epson Perfection V850 Pro scanner (Seiko Epson Corporation, Japan) and colonies were counted using the program OpenCFU [S10] (mean ± SD, n=2).

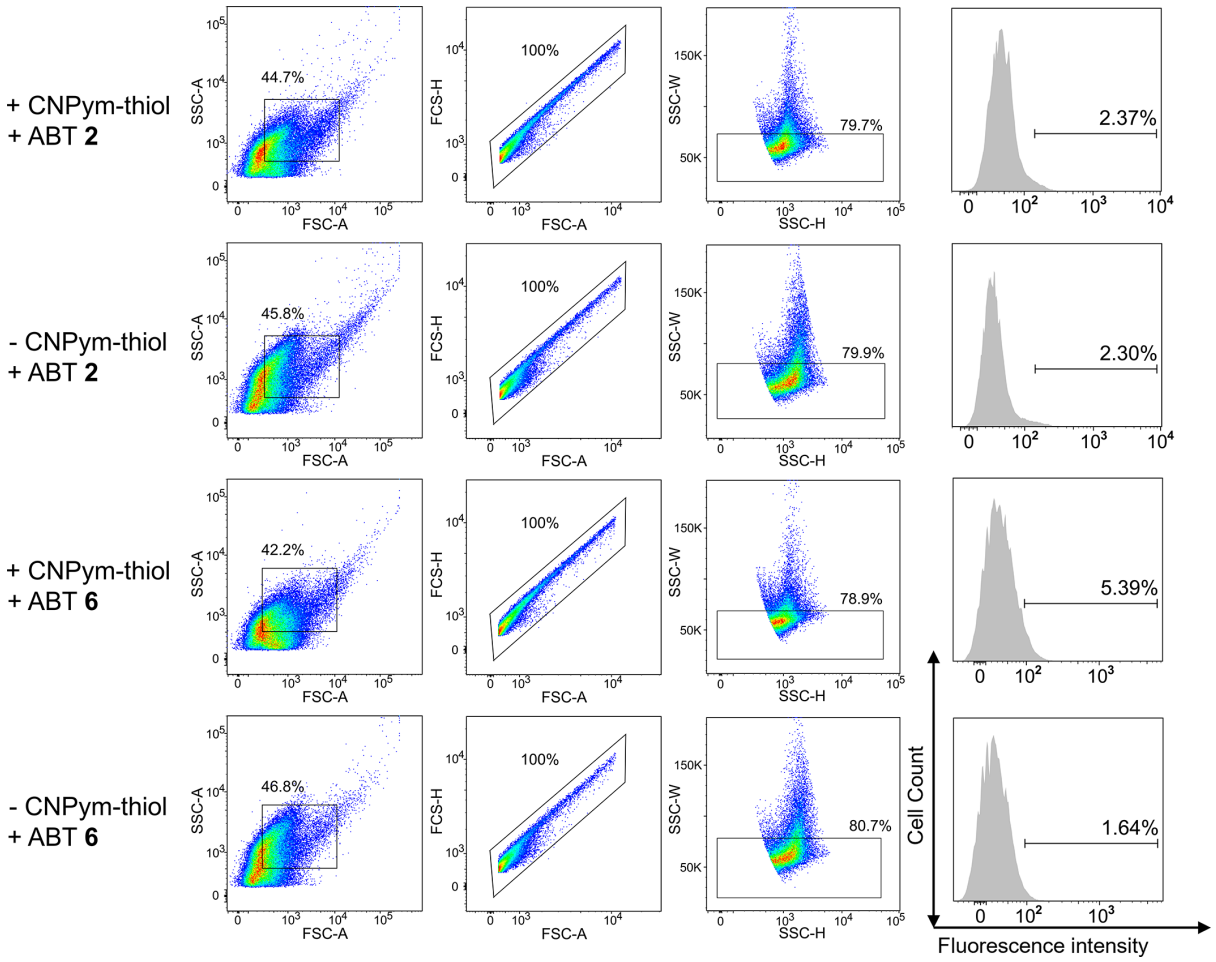


Figure S18. Flow cytometric analysis of *E. coli* cells incubated with CNPym-thiol and different ABTs to assess background fluorescence. Cells were grown in LB medium until reaching OD_{600} = 0.7. At this point, the medium was supplemented with 0.5 mM CNPym-thiol (+CNPym-thiol) or DMSO (-CNPym-thiol). Following 5 h incubation at 30 °C without induction of the protein of interest, the cells were harvested and subjected to the NAT click-reaction protocol of live-cell labelling. The flow cytometry gating of the cell populations was based on forward scatter (FSC) and side scatter (SSC) to discriminate the cells by size and granularity. Debris exclusion was performed using a density plot of SSC area (SSC-A) vs FSC area (FSC-A), and cells were selected by sequential gating with FSC height (FSC-H) vs FSC-A and SSC width (SSC-W) vs height (SSC-H) plots. These morphological assessments revealed no significant differences between the samples. The fluorescence intensities of cells treated with either ABT **2** or ABT **6** were detected using a 405 nm excitation laser and measured using 525/50 and 610/20 bandpass filters, respectively.

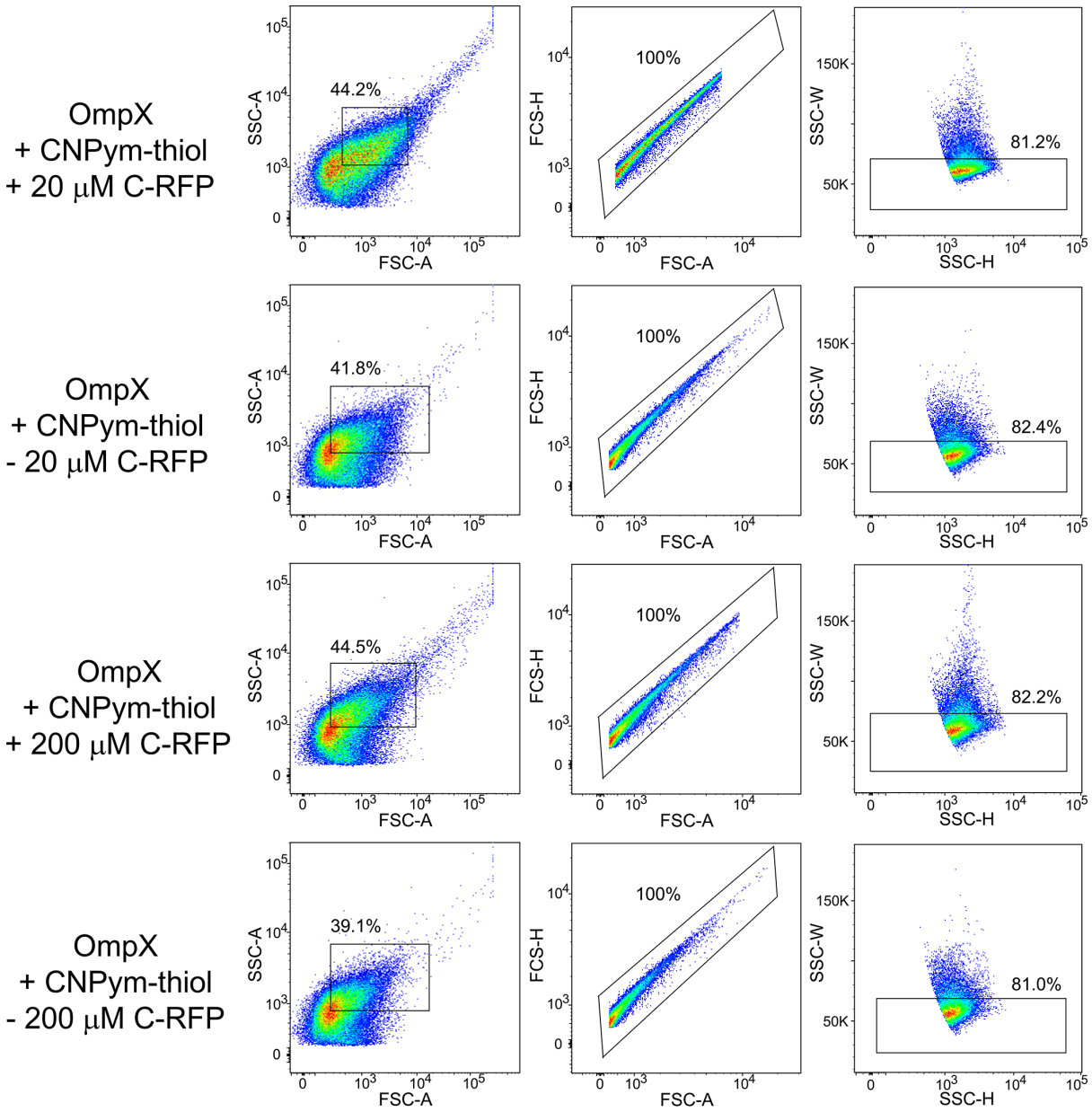


Figure S19. Flow cytometric analysis of the *E. coli* cell samples shown in **Figure 3**. Gating of the cell populations was based on forward scatter (FSC) and side scatter (SSC) to discriminate the cells by size and granularity. Debris exclusion was performed using a density plot of SSC area (SSC-A) vs FSC area (FSC-A), and cells were selected by sequential gating with FSC height (FSC-H) vs FSC-A and SSC width (SSC-W) vs height (SSC-H) plots. These morphological assessments showed no significant differences between the samples.

Video S1. Progression of the fluorogenic NAT click reaction in real time. The video illustrates the speed of the fluorogenic NAT click reaction between 100 μM NT_Ubi CysCNPym ΣCΣK and 10 equivalents of ABT **2** in PBS buffer containing 30 mM TCEP at 25 °C recorded under 365 nm LED illumination using an ordinary mobile phone camera (Samsung Electronics Co., Ltd., South Korea). The timer in the video starts at the mixing point.

Table S1. Mutations found in G1PylRS variants selected through directed evolution for CysCNP recognition[a].

RS Variants	*Randomized Sites*						
*Mm*PylRS-wt	L305	Y306	N346	V348	Y384	V401	W417
G1PylRS-wt	L124	Y125	N165	V167	Y204	A221	W237
CCNP12	S	F	N	F	W	G	Y
CCNP22	S	F	N	F	W	G	Y
CCNP24	S	F	N	F	W	G	Y
CCNP28	L	L	T	F	W	G	K
CCNP37	S	F	N	F	W	G	Y
CCNP47	S	F	N	F	W	G	Y
CCNP49	L	L	T	F	W	G	K

[a] Mutants CCNP12/22/24/37/47 and CCNP28/49 were identified to carry two distinct mutation sets, referred to as RS-I (SFNFWGY) and RS-II (LLTFWGK), respectively. RS-I was named G1(CysCNP)RS and used for all subsequent experiments.

Table S2. Yields of proteins expressed in the present study.

Protein	Expression temperature (°C)[a]	Yield (mg/L cell culture)
RFP wild type	18	103
	25	198
	30	73
RFP CysCNP	18	110
	25	202
	30	52
RFP CysCNPym	18	106
	25	201
	30	79
RFP mCNP	25	100
NT*_GD(CysCNP)	25	84
NT_Ubi CysCNP	25	132
	18	1830[b]
NT_Ubi CysCNPym	30	117
NT_Ubi CysCNPym ΣC	30	88
NT_Ubi CysCNPym ΣK	30	102
NT_Ubi CysCNPym ΣCΣK	30	104
NT_Ubi mCNP	25	156

[a] Expression times were 16, 16 and 5 h at 18, 25 and 30 °C, respectively.
[b] Yield from high cell-density fermentation.

Table S3. Photophysical properties of the chromophores formed in the arcNAT reaction between NT_Ubi mCNP (**m**), NT_Ubi CysCNP (**c**) or NT_Ubi CysCNPym ΣCΣK (**y**) and different ABTs **1–6**.

Chromophore[a]	$^{max}\lambda_{Ex}$ (nm)	$^{max}\lambda_{Em}$ (nm)	Quantum Yield (%)
m1	336	418	56.4 ± 0.7
m2	364	451	45.0 ± 1.9
m3	314	456	2.5 ± 0.1
m4	314	466	4.5 ± 0.1
m5	364	509	23.9 ± 0.2
m6	387	537	15.9 ± 0.0
c1	339	442	31.9 ± 0.1
c2	364	465	39.3 ± 0.2
c3	314	468	2.5 ± 0.0
c4	318	486	5.9 ± 0.1
c5	364	533	20.6 ± 0.1
c6	407	563	12.1 ± 0.1
y1	365	464	17.5 ± 1.0
y2	381	492	17.8 ± 0.5
y3	363	501	1.7 ± 0.1
y4	336	514	2.5 ± 0.2
y5	396	555	12.3 ± 1.3
y6	450	600	7.1 ± 0.1

a

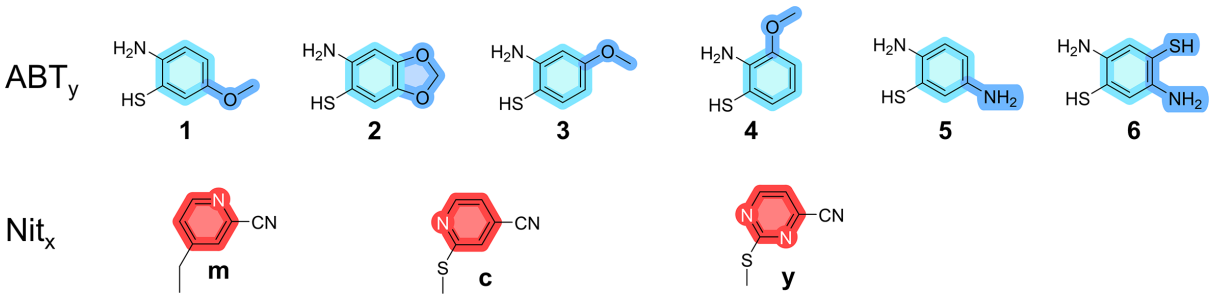
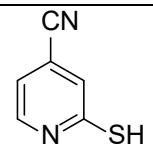
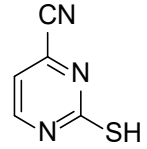
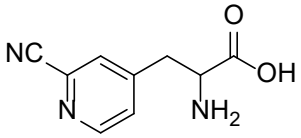
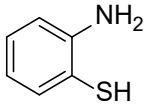
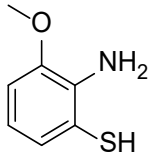
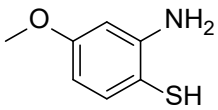
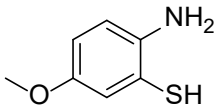
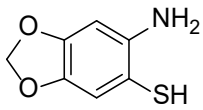
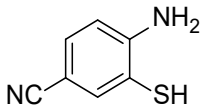
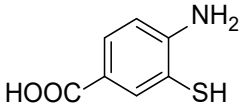
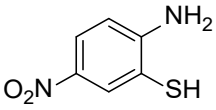
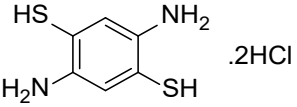
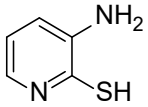
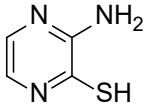
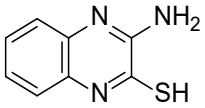
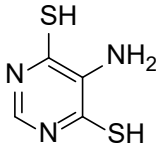
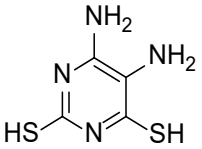


Table S4. List of chemicals and their suppliers used in the current study.

Compound	Catalogue number	Supplier	Cost (USD/g)[a]
 2-sulfanylpyridine-4-carbonitrile (CNP-thiol)	EN300-40040	Enamine Ltd (Kyiv, Ukraine)	110
 2-sulfanylpyrimidine-4-carbonitrile (CNPym-thiol)	EN300-4252097	Enamine Ltd (Kyiv, Ukraine)	430
 3-(2-cyano-4-pyridyl)alanine (mCNP)	CSMB00010457344	Chemspace LLC (Kyiv, Ukraine)	1594
 2-Aminobenzenethiol	274240	Merck KGaA (Darmstadt, Germany)	2
 2-amino-3-methoxybenzenethiol	EN300-1250006	Enamine Ltd (Kyiv, Ukraine)	252
 2-amino-4-methoxybenzenethiol	EN300-102529	Enamine Ltd (Kyiv, Ukraine)	222
 2-amino-5-methoxybenzenethiol	A627990	Toronto Research Chemicals Inc. (Toronto, Canada)	141
	EN300-246455	Enamine Ltd (Kyiv, Ukraine)	319

6-amino-1,3-dioxaindane-5-thiol			
 4-amino-3-mercaptobenzonitrile	EN300-126642	Enamine Ltd (Kyiv, Ukraine)	467
 4-amino-3-mercaptobenzoic acid	EN300-102436	Enamine Ltd (Kyiv, Ukraine)	140
 2-amino-5-nitrobenzenethiol	EN300-102436	Enamine Ltd (Kyiv, Ukraine)	140
 2,5-Diaminobenzene-1,4-dithiol	A159905	Ambeed (Illinois, USA)	12
 3-aminopyridine-2-thiol	EN300-40428	Enamine Ltd (Kyiv, Ukraine)	34
 3-aminopyrazine-2-thiol	EN300-180973	Enamine Ltd (Kyiv, Ukraine)	332
 3-aminoquinoxaline-2-thiol	EN300-1722221	Enamine Ltd (Kyiv, Ukraine)	301
 5-aminopyrimidine-4,6-dithiol	EN300-6744121	Enamine Ltd (Kyiv, Ukraine)	301
 2,4-Dimercapto-5,6-diaminopyrimidine	AABH93DE11BD	Merck KGaA (Darmstadt, Germany)	30

[a] Cost is calculated based of the largest available size, normalized per gram.

Table S5. DNA and corresponding amino acid sequences of the proteins used in the current study.

Gene	DNA sequence	Amino acid sequence[a]
G1(CysCNP)RS	ATGGTGGTGAAATTTACCGATAGCCAGATTCAGCATCTGATGGAA TATGGTGATAATGATTGGAGCGAAGCCGAATTTGAAGATGCAGCA GCACGTGATAAAGAATTTAGCAGCCAGTTTAGCAAACTGAAAAGC GCCAATGATAAAGGCCTGAAAGATGTTATTGCAAATCCGCGTAAT GATCTGACCGATCTGGAAAACAAAATTCGCGAAAAACTGGCAGCC CGTGGTTTTATTGAAGTTCATACCCCGATTTTTGTGAGCAAAAGC GCACTGGCAAAAATGACCATTACCGAAGATCATCCGCTGTTCAAA CAGGTGTTTTGGATTGATGATAAACGTGCACTGCGTCCGATGCAT GCAATGAATAGTTTTAAAGTTATGCGTGAACTGCGCGATCATACC AAAGGTCCGGTTAAAATCTTTGAAATTGGTAGCTGCTTTCGCAAA GAAAGCAAAAGCAGTACCCATCTGGAAGAATTTACCATGCTGAAC CTGTTCGAAATGGGTCCTGATGGTGATCCGATGGAACATCTGAAA ATGTATATTGGCGATATCATGGATGCCGTTGGTGTTGAATATACC ACCAGTCGTGAAGAATCAGATGTTTGGGTTGAAACCCTGGACGTG GAAATTAATGGCACCGAAGTTGCAAGCGGTGGTGTTGGTCCGCAT AAACTGGATCCGGCACATGATGTGCATGAACCGTATGCAGGTATT GGTTTTGGTCTGGAACGTCTGCTGATGCTGAAAAATGGTAAAAGC AATGCACGCAAAACCGGCAAAAGTATTACCTATCTGAATGGCTAC AAACTGGATTAA	MVVKFTDSQIQHLMEYGDN DWSEAEFEDAAARDKEFSS QFSKLKSANDKGLKDVIAN PRNDLTDLENKIREKLAAR GFIEVHTPIFVSKSALAKM TITEDHPLFKQVFWIDDKR ALRPMHAMNSFKVMRELRD HTKGPVKIFEIGSCFRKES KSSTHLEEFTMLNLFEMGP DGDPMEHLKMYIGDIMDAV GVEYTTSREESDVWVETLD VEINGTEVASGGVGPHKLD PAHDVHEPYAGIGFGLERL LMLKNGKSNARKTGKSITY LNGYKLD
RFP amber	ATGGCTTCTATGACCGGTCATCACCATCACCATCACTAGGCCAGT AGTGAAGACGTTATCAAGGAGTTTATGCGTTTCAAAGTACGTATG GAGGGTAGTGTTAACGGACACGAATTTGAGATCGAGGGAGAGGGG GAAGGTCGTCCTTACGAGGGAACTCAAACGGCCAAATTAAAGGTG ACCAAAGGTGGGCCCTTGCCATTCGCGTGGGACATCTTGTCACCC CAGTTCCAGTACGGGTCGAAGGCATACGTAAAACACCCAGCGGAC ATTCCTGACTATCTTAAGTTATCTTTCCCGGAAGGTTTTAAATGG GAACGCGTGATGAACTTTGAGGATGGGGGGGTTGTTACGGTGACA CAAGACTCCTCATTGCAAGATGGAGAGTTTATCTATAAAGTCAAA CTTCGCGGCACCAATTTTCCATCTGACGGTCCTGTAATGCAGAAA AAAACAATGGGCTGGGAAGCCTCCACAGAACGTATGTACCCCGAA GATGGAGCTTTAAAGGGCGAAATTAAAATGCGCTTAAAACTTAAA GACGGCGGCCATTACGACGCCGAAGTGAAAACGACGTATATGGCT AAGAAACCCGTCCAGCTTCCGGGAGCCTATAAAACTGACATCAAA CTGGATATTACATCACACAACGAAGATTATACTATTGTCGAACAG TACGAACGCGCCGAAGGCCGCCATTCAACGGGAGCATAA	MASMTGHHHHHHXASSEDV IKEFMRFKVRMEGSVNGHE FEIEGEGEGRPYEGTQTAK LKVTKGGPLPFAWDILSPQ FQYGSKAYVKHPADIPDYL KLSFPEGFKWERVMNFEDG GVVTVTQDSSLQDGEFIYK VKLRGTNFPSDGPVMQKKT MGWEASTERMYPEDGALKG EIKMRLKLKDGGHYDAEVK TTYMAKKPVQLPGAYKTDI KLDITSHNEDYTIVEQYER AEGRHSTGA
NT* C-terminal tripeptide motif (NT*_GDX)	ATGCATCATCATCATCACCACAGCCATACCACACCGTGGACCAAT CCTGGTCTGGCAGAAAACTTTATGAATAGCTTTATGCAGGGTCTG AGCAGCATGCCTGGTTTTACCGCAAGCCAGCTGGACAAAATGAGC ACCATTGCACAGAGCATGGTTCAGAGCATTCAGAGCCTGGCAGCA CAGGGTCGTACCAGTCCGAATGATCTGCAGGCACTGAATATGGCA TTTGCAAGCAGCATGGCAGAAATTGCAGCAAGCGAAGAAGGTGGC GGTAGCCTGAGCACCAAAACCAGCAGCATTGCAAGCGCAATGAGC AATGCATTTCTGCAGACAACCGGTGTTGTTAATCAGCCGTTTATT AACGAAATTACCCAGCTGGTTAGCATGTTTGCACAGGCAGGTATG AATGATGTTAGCGCAGAAAACCTGTACTTTCAAGGCGATTAGTAA	MHHHHHHSHTTPWTNPGLA ENFMNSFMQGLSSMPGFTA SQLDKMSTIAQSMVQSIQS LAAQGRTSPNDLQALNMAF ASSMAEIAASEEGGGSLST KTSSIASAMSNAFLQTTGV VNQPFINEITQLVSMFAQA GMNDVSAENLYFQGDX

NT_Ubi amber	ATGCATCATCATCATCACCACAGCCATACCACACCGTGGACCAAT CCTGGTCTGGCAGAAAACTTTATGAATAGCTTTATGCAGGGTCTG AGCAGCATGCCTGGTTTTACCGCAAGCCAGCTGGACAAAATGAGC ACCATTGCACAGAGCATGGTTCAGAGCATTCAGAGCCTGGCAGCA CAGGGTCGTACCAGTCCGAATGATCTGCAGGCACTGAATATGGCA TTTGCAAGCAGCATGGCAGAAATTGCAGCAAGCGAAGAAGGTGGC GGTAGCCTGAGCACCAAAACCAGCAGCATTGCAAGCGCAATGAGC AATGCATTTCTGCAGACAACCGGTGTTGTTAATCAGCCGTTTATT AACGAAATTACCCAGCTGGTTAGCATGTTTGCACAGGCAGGTATG AATGATGTTAGCGCAGGTAATAGCGGTGGTGGTGGTAGCGGTGAA AACCTGTATTTTCAGTGCGGCAAACGCAAAAGCTAGGGCGGTGGT GGCAGCATGCAGATCTTCGTGAAGACTCTGACTGGTAAGACCATC ACCCTCGAGGTTGAGCCCAGTGACACCATTGAGAATGTCAAGGCA AAGATCCAAGATAAGGAAGGCATCCCTCCTGACCAGCAGAGGCTG ATCTTTGCTGGAAAACAGCTGGAAGATGGGCGCACCCTGTCTGAC TACAACATCCAGAAAGAGTCCACCCTGCACCTGGTACTCCGTCTC AGAGGTGGATAA	MHHHHHHSHTTPWTNPGLA ENFMNSFMQGLSSMPGFTA SQLDKMSTIAQSMVQSIQS LAAQGRTSPNDLQALNMAF ASSMAEIAASEEGGGSLST KTSSIASAMSNAFLQTTGV VNQPFINEITQLVSMFAQA GMNDVSAGNSGGGGSGENL YFQCGKRKSXGGGGSMQIF VKTLTGKTITLEVEPSDTI ENVKAKIQDKEGIPPDQQR LIFAGKQLEDGRTLSDYNI QKESTLHLVLRLRGG
NT_Ubi ΣC amber	ATGCATCATCATCATCACCACAGCCATACCACACCGTGGACCAAT CCTGGTCTGGCAGAAAACTTTATGAATAGCTTTATGCAGGGTCTG AGCAGCATGCCTGGTTTTACCGCAAGCCAGCTGGACAAAATGAGC ACCATTGCACAGAGCATGGTTCAGAGCATTCAGAGCCTGGCAGCA CAGGGTCGTACCAGTCCGAATGATCTGCAGGCACTGAATATGGCA TTTGCAAGCAGCATGGCAGAAATTGCAGCAAGCGAAGAAGGTGGC GGTAGCCTGAGCACCAAAACCAGCAGCATTGCAAGCGCAATGAGC AATGCATTTCTGCAGACAACCGGTGTTGTTAATCAGCCGTTTATT AACGAAATTACCCAGCTGGTTAGCATGTTTGCACAGGCAGGTATG AATGATGTTAGCGCAGGTAATAGCGGTGGTGGTGGTAGCGGTGAA AACCTGTATTTTCAGGCGGGCAAACGCAAAAGCTAGGGCGGTGGT GGCAGCATGCAGATCTTCGTGAAGACTCTGACTGGTAAGACCATC ACCCTCGAGGTTGAGCCCAGTGACACCATTGAGAATGTCAAGGCA AAGATCCAAGATAAGGAAGGCATCCCTCCTGACCAGCAGAGGCTG ATCTTTGCTGGAAAACAGCTGGAAGATGGGCGCACCCTGTCTGAC TACAACATCCAGAAAGAGTCCACCCTGCACCTGGTACTCCGTCTC AGAGGTGGATAA	MHHHHHHSHTTPWTNPGLA ENFMNSFMQGLSSMPGFTA SQLDKMSTIAQSMVQSIQS LAAQGRTSPNDLQALNMAF ASSMAEIAASEEGGGSLST KTSSIASAMSNAFLQTTGV VNQPFINEITQLVSMFAQA GMNDVSAGNSGGGGSGENL YFQAGKRKSXGGGGSMQIF VKTLTGKTITLEVEPSDTI ENVKAKIQDKEGIPPDQQR LIFAGKQLEDGRTLSDYNI QKESTLHLVLRLRGG
NT_Ubi ΣK amber	ATGCATCATCATCATCACCACAGCCATACCACACCGTGGACCAAT CCTGGTCTGGCAGAAAACTTTATGAATAGCTTTATGCAGGGTCTG AGCAGCATGCCTGGTTTTACCGCAAGCCAGCTGGACAAAATGAGC ACCATTGCACAGAGCATGGTTCAGAGCATTCAGAGCCTGGCAGCA CAGGGTCGTACCAGTCCGAATGATCTGCAGGCACTGAATATGGCA TTTGCAAGCAGCATGGCAGAAATTGCAGCAAGCGAAGAAGGTGGC GGTAGCCTGAGCACCAAAACCAGCAGCATTGCAAGCGCAATGAGC AATGCATTTCTGCAGACAACCGGTGTTGTTAATCAGCCGTTTATT AACGAAATTACCCAGCTGGTTAGCATGTTTGCACAGGCAGGTATG AATGATGTTAGCGCAGGTAATAGCGGTGGTGGTGGTAGCGGTGAA AACCTGTATTTTCAGTGCGGCGCACGCGCGAGCTAGGGCGGTGGT GGCAGCATGCAGATCTTCGTGAAGACTCTGACTGGTAAGACCATC ACCCTCGAGGTTGAGCCCAGTGACACCATTGAGAATGTCAAGGCA AAGATCCAAGATAAGGAAGGCATCCCTCCTGACCAGCAGAGGCTG ATCTTTGCTGGAAAACAGCTGGAAGATGGGCGCACCCTGTCTGAC TACAACATCCAGAAAGAGTCCACCCTGCACCTGGTACTCCGTCTC AGAGGTGGATAA	MHHHHHHSHTTPWTNPGLA ENFMNSFMQGLSSMPGFTA SQLDKMSTIAQSMVQSIQS LAAQGRTSPNDLQALNMAF ASSMAEIAASEEGGGSLST KTSSIASAMSNAFLQTTGV VNQPFINEITQLVSMFAQA GMNDVSAGNSGGGGSGENL YFQCGARASXGGGGSMQIF VKTLTGKTITLEVEPSDTI ENVKAKIQDKEGIPPDQQR LIFAGKQLEDGRTLSDYNI QKESTLHLVLRLRGG

NT_Ubi ΣCΣK amber	ATGCATCATCATCATCACCACAGCCATACCACACCGTGGACCAAT CCTGGTCTGGCAGAAAACTTTATGAATAGCTTTATGCAGGGTCTG AGCAGCATGCCTGGTTTTACCGCAAGCCAGCTGGACAAAATGAGC ACCATTGCACAGAGCATGGTTCAGAGCATTCAGAGCCTGGCAGCA CAGGGTCGTACCAGTCCGAATGATCTGCAGGCACTGAATATGGCA TTTGCAAGCAGCATGGCAGAAATTGCAGCAAGCGAAGAAGGTGGC GGTAGCCTGAGCACCAAAACCAGCAGCATTGCAAGCGCAATGAGC AATGCATTTCTGCAGACAACCGGTGTTGTTAATCAGCCGTTTATT AACGAAATTACCCAGCTGGTTAGCATGTTTGCACAGGCAGGTATG AATGATGTTAGCGCAGGTAATAGCGGTGGTGGTGGTAGCGGTGAA AACCTGTATTTTCAGGCTGGCGCACGCGCGAGCTAGGGCGGTGGT GGCAGCATGCAGATCTTCGTGAAGACTCTGACTGGTAAGACCATC ACCCTCGAGGTTGAGCCCAGTGACACCATTGAGAATGTCAAGGCA AAGATCCAAGATAAGGAAGGCATCCCTCCTGACCAGCAGAGGCTG ATCTTTGCTGGAAAACAGCTGGAAGATGGGCGCACCCTGTCTGAC TACAACATCCAGAAAGAGTCCACCCTGCACCTGGTACTCCGTCTC AGAGGTGGATAA	MHHHHHHSHTTPWTNPGLA ENFMNSFMQGLSSMPGFTA SQLDKMSTIAQSMVQSIQS LAAQGRTSPNDLQALNMAF ASSMAEIAASEEGGGSLST KTSSIASAMSNAFLQTTGV VNQPFINEITQLVSMFAQA GMNDVSAGNSGGGGSGENL YFQAGARASXGGGGSMQIF VKTLTGKTITLEVEPSDTI ENVKAKIQDKEGIPPDQQR LIFAGKQLEDGRTLSDYNI QKESTLHLVLRLRGG
C-RFP[b]	ATGCACCACCATCACCATCACGAAAACCTGTACTTTCAATGCGCC AGTAGTGAAGACGTTATCAAGGAGTTTATGCGTTTCAAAGTACGT ATGGAGGGTAGTGTTAACGGACACGAATTTGAGATCGAGGGAGAG GGGGAAGGTCGTCCTTACGAGGGAACTCAAACGGCCAAATTAAAG GTGACCAAAGGTGGGCCCTTGCCATTCGCGTGGGACATCTTGTCA CCCCAGTTCCAGTACGGGTCGAAGGCATACGTAAAACACCCAGCG GACATTCCTGACTATCTTAAGTTATCTTTCCCGGAAGGTTTTAAA TGGGAACGCGTGATGAACTTTGAGGATGGGGGGGTTGTTACGGTG ACACAAGACTCCTCATTGCAAGATGGAGAGTTTATCTATAAAGTC AAACTTCGCGGCACCAATTTTCCATCTGACGGTCCTGTAATGCAG AAAAAAACAATGGGCTGGGAAGCCTCCACAGAACGTATGTACCCC GAAGATGGAGCTTTAAAGGGCGAAATTAAAATGCGCTTAAAACTT AAAGACGGCGGCCATTACGACGCCGAAGTGAAAACGACGTATATG GCTAAGAAACCCGTCCAGCTTCCGGGAGCCTATAAAACTGACATC AAACTGGATATTACATCACACAACGAAGATTATACTATTGTCGAA CAGTACGAACGCGCCGAAGGCCGCCATTCAACGGGAGCATAA	MHHHHHHENLYFQCASSED VIKEFMRFKVRMEGSVNGH EFEIEGEGEGRPYEGTQTA KLKVTKGGPLPFAWDILSP QFQYGSKAYVKHPADIPDY LKLSFPEGFKWERVMNFED GGVVTVTQDSSLQDGEFIY KVKLRGTNFPSDGPVMQKK TMGWEASTERMYPEDGALK GEIKMRLKLKDGGHYDAEV KTTYMAKKPVQLPGAYKTD IKLDITSHNEDYTIVEQYE RAEGRHSTGA
OmpX amber	ATGAAAAAGATTGCATGTCTGAGCGCACTGGCAGCAGTTCTGGCA TTTACCGCAGGCACCAGCGTTGCAGCAACCAGCACCGTTACCGGT GGTTATGCACAGAGTGATGCACAGGGTCAGATGAATAAAATGGGT GGCTTTAATCTGAAATATCGCTACGAAGAAGATAATAGTCCGCTG GGTGTTATTGGTAGCTTTACCTATACCGAAAAATCAAGAACCGCA AGCGCGGCCGCAGCGTAGGCAGCCAGCGGTGATTATAACAAAAAT CAGTATTATGGCATTACAGCCGGTCCGGCATATCGTATTAATGAT TGGGCAAGCATTTATGGTGTTGTTGGTGTTGGTTATGGCAAATTT CAGACCACCGAATATCCGACCTATAAACATGATACCAGCGATTAT GGTTTTAGCTATGGTGCAGGTCTGCAGTTTAATCCGATGGAAAAT GTTGCACTGGATTTCAGCTATGAACAGAGCCGTATTCGTAGCGTT GATGTTGGCACCTGGATTGCCGGTGTGGGTTATCGTTTTCACCAT CACCATCACCATTAA	MKKIACLSALAAVLAFTAG TSVAATSTVTGGYAQSDAQ GQMNKMGGFNLKYRYEEDN SPLGVIGSFTYTEKSRTAS AAAAXAASGDYNKNQYYGI TAGPAYRINDWASIYGVVG VGYGKFQTTEYPTYKHDTS DYGFSYGAGLQFNPMENVA LDFSYEQSRIRSVDVGTWI AGVGYRFHHHHHH*

[a] X indicates the positions of the ncAA.

[b] Sequence of C-RFP with the N-terminal His_6 tag and TEV recognition sequence. The amino acid sequence of C-RFP after TEV cleavage is underlined.

References

[S1] Qianzhu, H., Abdelkader, E.H., Otting, G., and Huber, T. (2024). Genetic encoding of fluoro-L-tryptophans for site-specific detection of conformational heterogeneity in proteins by NMR spectroscopy. J. Am. Chem. Soc. *146*, 13641–13650. 10.1021/jacs.4c03743.

[S2] Abdelkader, E.H., Qianzhu, H., George, J., Frkic, R.L., Jackson, C.J., Nitsche, C., Otting, G., and Huber, T. (2022). Genetic encoding of cyanopyridylalanine for in-cell protein macrocyclization by the nitrile–aminothiol click reaction. Angew. Chem. Int. Ed. *61*, e202114154. 10.1002/anie.202114154.

[S3] Abdelkader, E.H., Qianzhu, H., Huber, T., and Otting, G. (2023). Genetic encoding of 7-aza-L-tryptophan: isoelectronic substitution of a single CH-group in a protein for a nitrogen atom for site-selective isotope labeling. ACS sens. *8*, 4402–4406. 10.1021/acssensors.3c01904.

[S4] Qianzhu, H., Abdelkader, E.H., Herath, I.D., Otting, G., and Huber, T. (2022). Site-specific incorporation of 7-fluoro-L-tryptophan into proteins by genetic encoding to monitor ligand binding by ^{19}F NMR spectroscopy. ACS sens. *7*, 44–49. 10.1021/acssensors.1c02467.

[S5] Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., Yokoyama, S., and Sakamoto, K. (2015). Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. Sci. Rep. *5*, 9699. 10.1038/srep09699.

[S6] Klopp, J., Winterhalter, A., Gébleux, R., Scherer-Becker, D., Ostermeier, C., and Gossert, A.D. (2018). Cost-effective large-scale expression of proteins for NMR studies. J. Biomol. NMR *71*, 247–262. 10.1007/s10858-018-0179-0.

[S7] Würth, C., Grabolle, M., Pauli, J., Spieles, M., and Resch-Genger, U. (2013). Relative and absolute determination of fluorescence quantum yields of transparent samples. Nat. Protoc. *8*, 1535–1550. 10.1038/nprot.2013.087.

[S8] Skinner, S.O., Sepúlveda, L.A., Xu, H., and Golding, I. (2013). Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent *in situ* hybridization. Nat. Protoc. *8*, 1100–1113. 10.1038/nprot.2013.066.

[S9] Aon, J.C., Caimi, R.J., Taylor, A.H., Lu, Q., Oluboyede, F., Dally, J., Kessler, M.D., Kerrigan, J.J., Lewis, T.S., Wysocki, L.A., and Patel, P.S. (2008). Suppressing posttranslational gluconoylation of heterologous proteins by metabolic engineering of *Escherichia coli*. Appl. Environ. Microbiol. *74*, 950–958. 10.1128/AEM.01790-07.

[S10] Geissmann, Q. (2013). OpenCFU, a new fee and open-source software to count cell colonies and other circular objects. PLOS ONE *8*, e54072. 10.1371/journal.pone.0054072.