Electronic Supplementary Information (ESI)

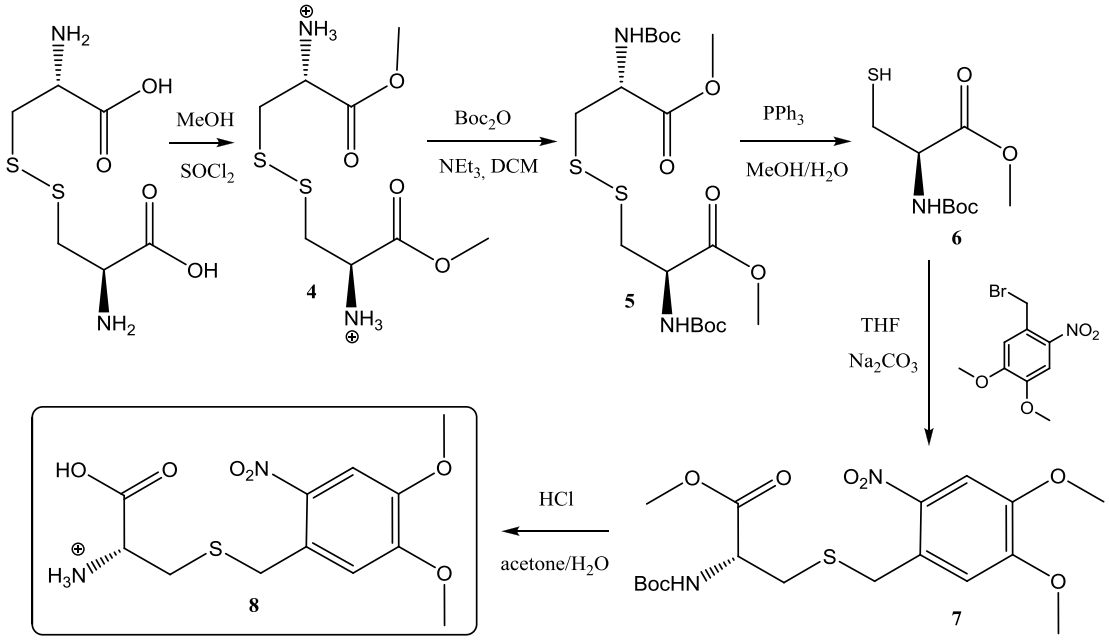
Biosynthetic Selenoproteins with Genetically-Encoded Photocaged Selenocysteines

Rasa Rakauskaitė, Giedrė Urbanavičiūtė, Audronė Rukšėnaitė, Zita Liutkevičiūtė, Robertas Juškėnas, Viktoras Masevičius, and Saulius Klimašauskas*

Institute of Biotechnology, Vilnius University, Vilnius 02241, Lithuania and
Faculty of Chemistry, Vilnius University, Vilnius 03225, Lithuania.

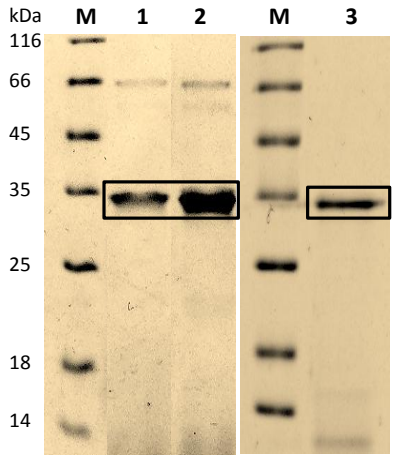
* Correspondence to: saulius.klimasauskas@bti.vu.lt

Supplementary Results

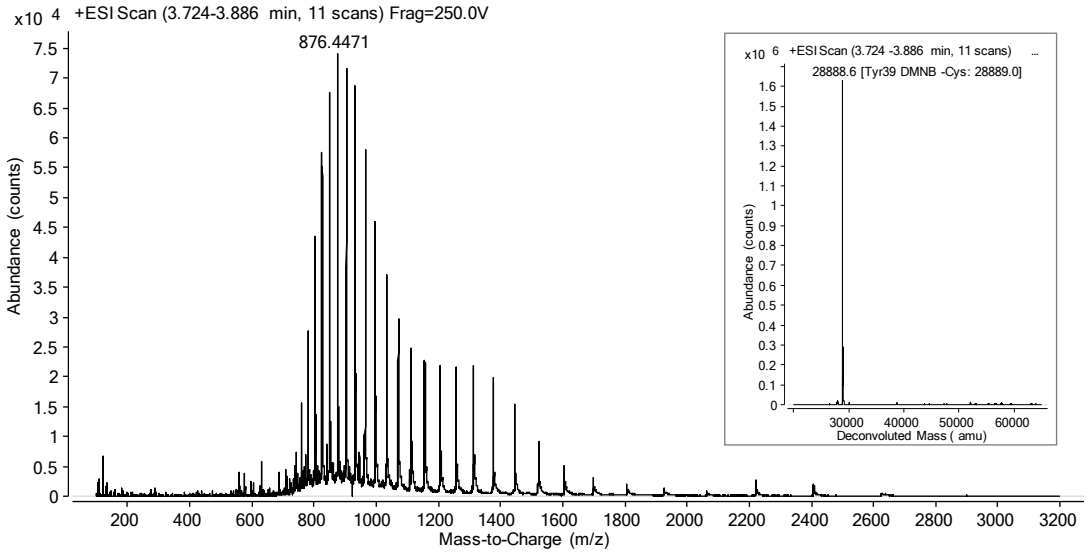


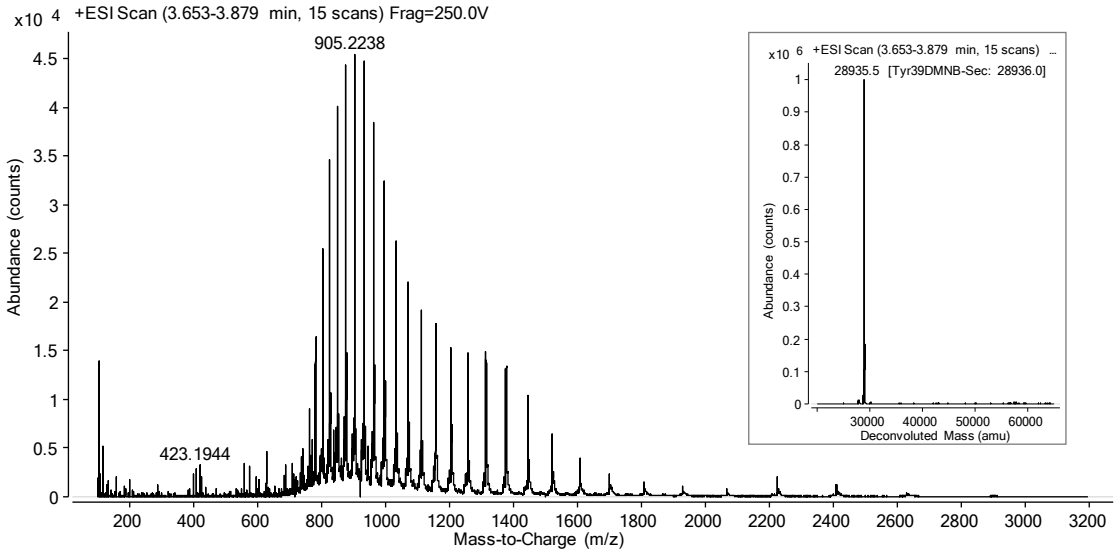
Supplementary Scheme S1. Synthetic route to DMNB-Cys (**8**).

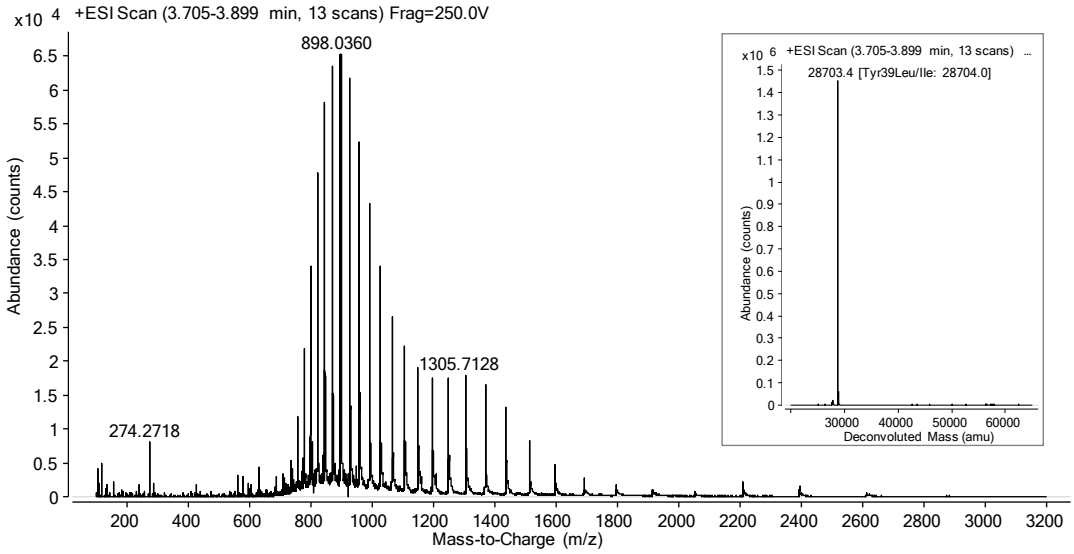
a


kDa
M
1
2
M
3
116
66
45
35
25
18
14

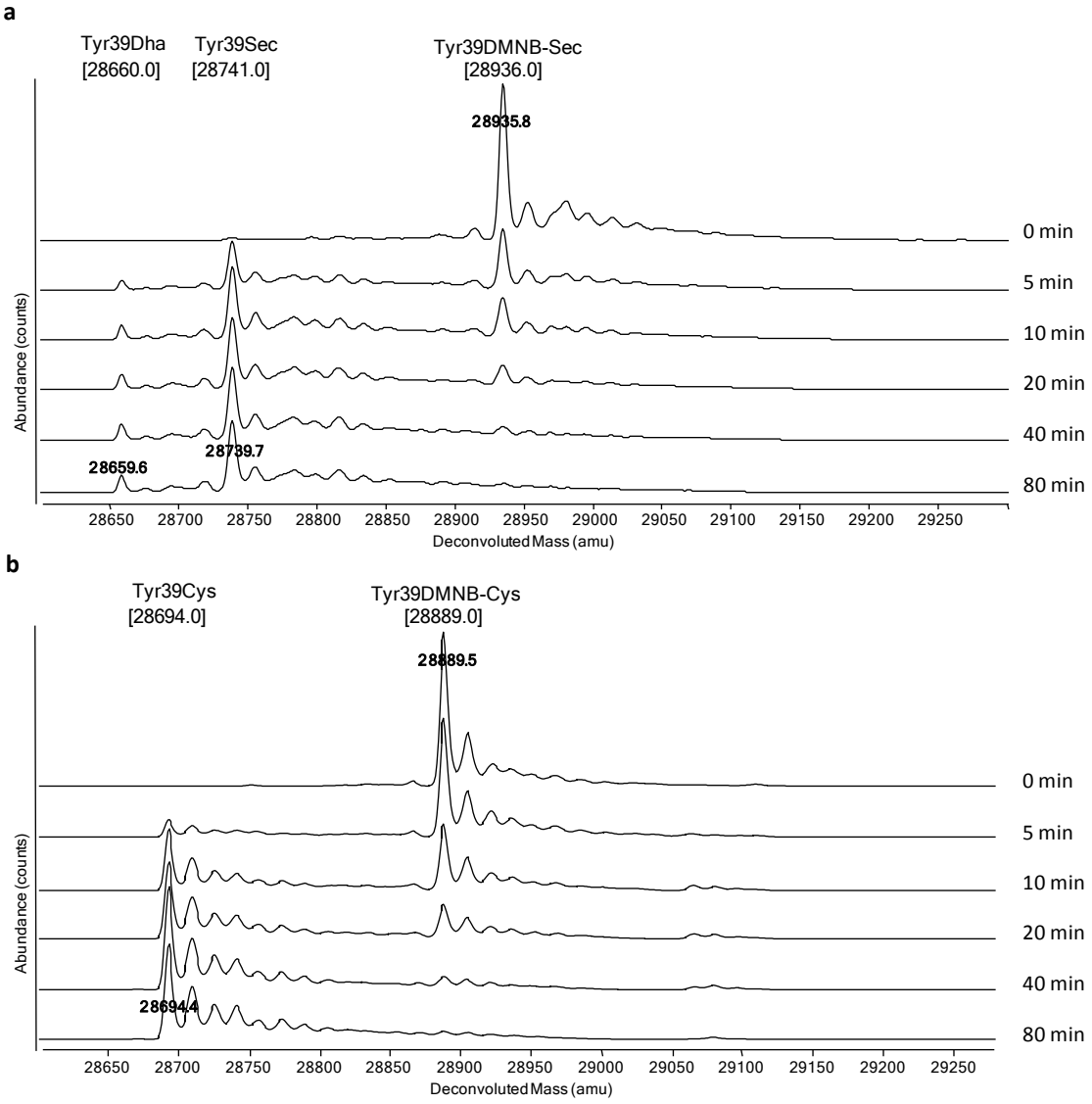
b


x10 4 +ESI Scan (3.724-3.886 min, 11 scans) Frag=250.0V
876.4471
Abundance (counts)
Mass-to-Charge (m/z)
x10 6 +ESI Scan (3.724 -3.886 min, 11 scans)
28888.6 [Tyr39 DMNB -Cys: 28889.0]
Abundance (counts)
Deconvoluted Mass (amu)


x10 4 +ESI Scan (3.653-3.879 min, 15 scans) Frag=250.0V
905.2238
423.1944
Abundance (counts)
Mass-to-Charge (m/z)
x10 6 +ESI Scan (3.653-3.879 min, 15 scans)
28935.5 [Tyr39DMNB-Sec: 28936.0]
Abundance (counts)
Deconvoluted Mass (amu)

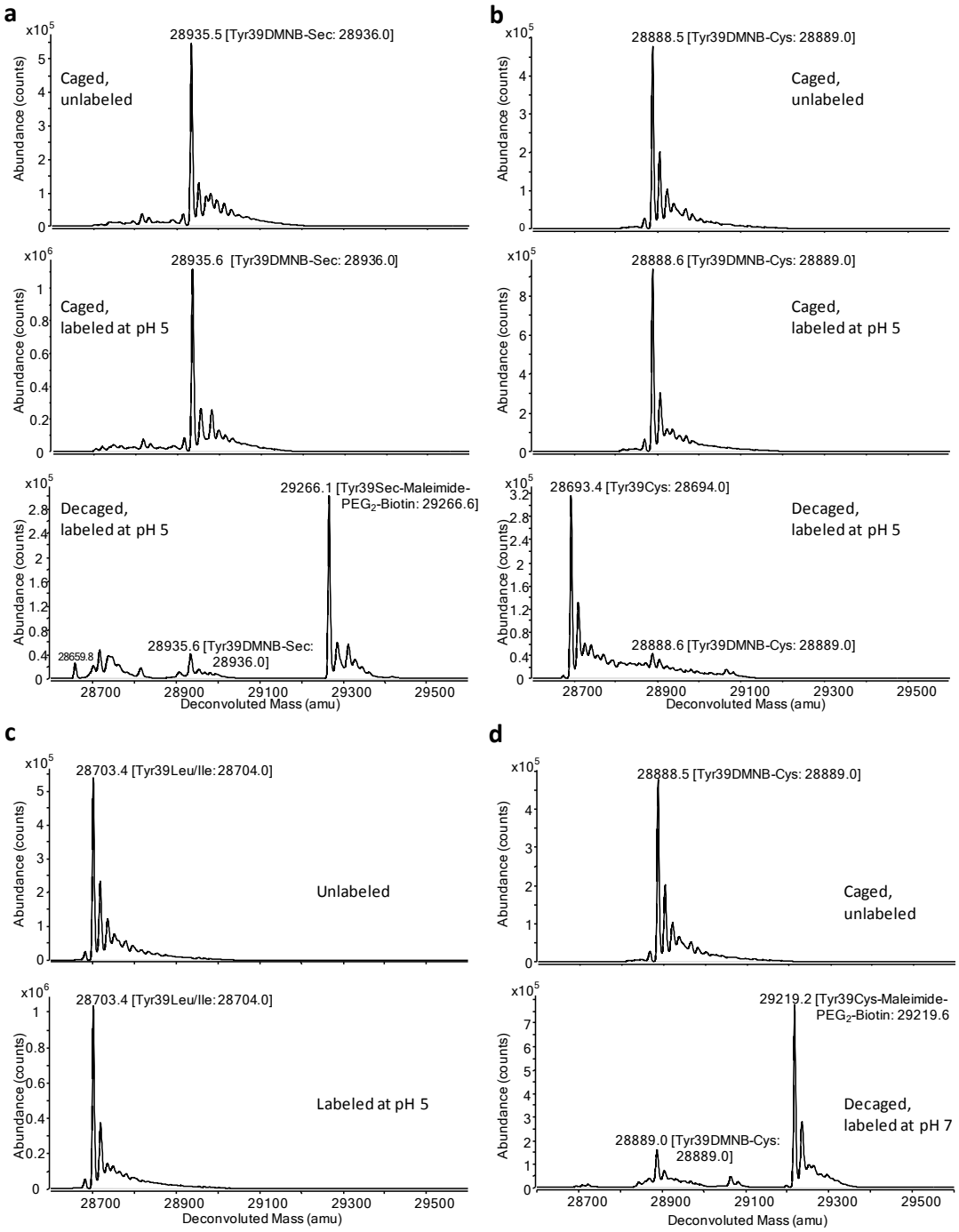


Supplementary Figure S1 . Analysis of affinity chromatography-purified EGFP variants. **(a)** SDS-PAGE analysis. Bands corresponding to fractionated recombinant proteins EGFPTyr39DMNB-Cys (1), EGFPTyr39DMNB-Sec (2), and EGFPTyr39Leu/Ile (3) are highlighted with a frame; M, molecular weight standards. **(b) HPLC/**ESI-MS analysis. Native mass spectra and corresponding deconvoluted mass spectra (inset) of the recombinant proteins EGFPTyr39DMNB-Cys (upper), EGFPTyr39DMNB-Sec (center), and EGFPTyr39Leu/Ile (lower). Theoretical masses of the proteins are shown in brackets.

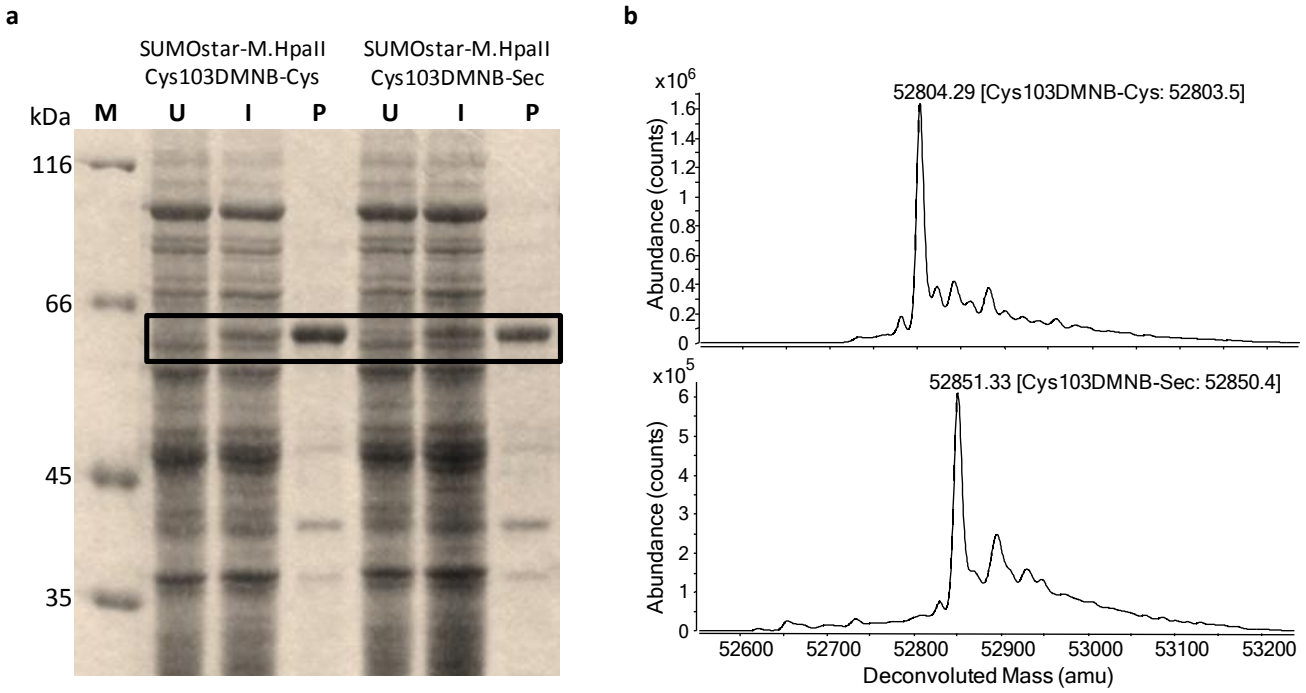


Supplementary Figure S2. Time courses of photochemical decaging of **(a)** EGFPTyr39DMNB-Sec, and **(b)** EGFPTyr39DMNB-Cys proteins, analyzed by HPLC/ESI-MS under reducing conditions (50 mM DTT). Theoretical masses of reaction substrates and expected products are shown in brackets. Deconvoluted spectrograms are scaled to a signal intensity of the most

abundant peak (5×10^5 counts).



Supplementary Figure S3. Labeling of EGFP protein variants with maleimide-PEG_2-biotin. ESI-MS spectra of control and labeling reactions of **(a)** photocaged and decaged EGFPTyr39DMNB-Sec proteins at pH 5, **(b)** photocaged and decaged EGFPTyr39DMNB-Cys proteins at pH 5, **(c)** EGFPTyr39Leu/Ile protein at pH 5, **(d)** decaged EGFPTyr39DMNB-Cys protein at pH 7. Sample treatment conditions prior to HPLC/ESI-MS analysis are indicated for each trace. Theoretical masses (shown in brackets) of corresponding maleimide-PEG_2-biotin coupling reaction products are: 29461.6 Da for **a** (panel 2); 29414.6 Da for **b** (panel 2); 29219.6 Da for **b** (panel 3); and 29229.6 Da for **c** (panel 2). Note that similar patterns of minor peaks following the major peak, manifesting inherent and apparent heterogeneity‡ of the EGFP proteins in solution, are observed regardless of the nature of residue incorporated at position 39 (DMNB-Sec, DMNB-Cys or Leu/Ile).



Supplementary Figure S4. Expression and purification of the SUMOstar-M.HpaII Cys103DMNB-Cys and SUMOstar-M.HpaII Cys103DMNB-Sec proteins. **(a)** SDS-PAGE analysis of total protein extracts of uninduced (U) and $CuSO_4$ induced (I) yeast cells and affinity chromatography-purified (P) proteins. Bands corresponding to fractionated recombinant proteins are highlighted with a frame; M, molecular weight standard. **(b)** HPLC/ESI-MS analysis of SUMOstar-M.HpaII Cys103DMNB-Cys and SUMOstar-M.HpaII Cys103DMNB-Sec proteins. Theoretical masses are shown in brackets.

Supplementary Methods

Synthetic protocols

Analytical TLC was conducted on silica gel plates Silufol 60 F254 (Merck) with detection by UV light or basic potassium permanganate solution. Column flash chromatography was performed on silica gel (Merck, 60 Å, 230-400 mesh). ^{1}H and ^{13}C spectra were recorded on a Bruker 400 (400 MHz and 100 MHz, respectively) spectrometer using residual solvent signals as an internal reference. IR spectra were run on a Perkin-Elmer FT-IR spectrophotometer Spectrum BX II in KBr. HRMS data were obtained using an Agilent 6230 TOF mass spectrometer (ESI). Elemental analyses were performed on a Thermo Scientific Flash 2000 apparatus. Melting points were determined in open capillaries using a digital melting point IA9100 series apparatus (Electrothermal).

Synthesis of DMNB-Sec (Figure 1)

N,N'-bis(*tert*-butoxycarbonyl)seleno-L-cystine **1**

Solution of selenocystine (0.625 g, 1.88 mmol) in 3.8 ml of alkaline water (1 M NaOH) was combined with solution of di-*tert*-butyldicarbonate (1.633 g, 7.49 mmol) in 1.8 ml of dioxane and reaction mixture was stirred for 24 hours at room temperature, concentrated under reduced pressure and washed with ethyl acetate. Water phase was acidified with hydrochloric acid, and product extracted with ethyl acetate. Organic layer washed with brine, dried over sodium sulphate. Ethyl acetate was removed under reduced pressure. Residue was recrystallized from benzene to yield 0.755 g (75.5%) of target product as yellowish solid (m.p. 144–146 °C).

^{1}H NMR (300 MHz, $CDCl_3$): δ = 1.45 (9H, s, 3×CH_3), 3.21 (1H, dd *J* = 12.6, 4.8Hz, SeCH_AH$_B$), 3.46 (1H, dd *J* = 12.6, 4.8Hz, SeC$H_A H_B$), 4.61 (1H, brm, CH), 5.37 (1H, brs, NH) and is consistent with published data.[38] ^{13}C NMR (100 Mz, DMSO-D_6): δ = 28.6, 31.4, 54.7, 78.7, 155.8, 173.0. IR (KBr): 3368 (NH), 1699 (CO), 1684 (CO). Anal. calcd. for $C_{16}H_{28}N_2O_8Se_2$: C, 35.97; H, 5.28; N, 5.24. Found: C, 36.17; H, 5.35; N, 5,08.

N-(*tert*-butoxycarbonyl)-*Se*-(4,5-dimethoxy-2-nitrobenzyl)seleno-L-cysteine **2**

Mixture of *N,N*'-bis(*tert*-butoxycarbonyl)selenocystine (0.755 g, 1.41 mmol) and 4,5-dimethoxy-2-nitrobenzylbromide (1.170 g, 4.24 mmol) in ethanol (30 ml) was cooled to 0 °C, then sodium borohydride (0.161 g, 4.26 mmol) poured to reaction vessel. Reaction mixture was stirred for 75 minutes at room temperature then quenched with water (70 ml). Ethyl acetate was added and reaction mixture was acidified with hydrochloric acid. Organic layer was separated and dried over sodium sulphate. Solvents evaporated under reduced pressure, residue purified by column chromatography ($CHCl_3$, R_f = 0.26). Described procedure yielded 1.133 g (86.5%) of target product as yellow solid (m.p. 76–78 °C).

^{1}H NMR (400 MHz, $CDCl_3$): δ = 1.46 (9H, s, 3×CH_3), 3.03 (1H, dd *J* = 13.6, 4.8Hz, SeC$H_A$$H_B$), 3.09 (1H, dd *J* = 13.6, 4.8Hz, SeC$H_A$$H_B$), 3.95 (3H, s, OCH_3), 4.01 (3H, s, OCH_3), 4.11 (1H, d *J* = 11.6Hz, BnC$H_A$$H_B$Se), 4.21 (1H, d *J* = 11.6Hz, BnC$H_A$$H_B$Se), 4.62 (1H, brm, CH), 5.46 (1H, brm, NH), 6.83 (1H, s, ArH), 7.69 (1H, s, ArH). ^{13}C NMR (100 MHz, $CDCl_3$): δ = 25.9 (2 overlapping signals, it's clear from HSQC data), 28.3, 53.6, 56.4, 56.6, 80.7, 108.9, 113.2, 130.6, 139.7, 147.8, 153.2, 155.6, 174.9. IR (KBr): 3360 (NH), 1712 (CO), 1685 (CO). Anal. calcd. for $C_{17}H_{24}N_2O_8Se$: C, 44.07; H, 5.22. Found: C, 44.13; H, 5.53. HRMS: m/z $[M+Na]^+$ calcd. for $C_{17}H_{24}N_2O_8Se$: 487.0591; Found: 487.0589.

Se-(4,5-dimethoxy-2-nitrobenzyl)seleno-L-cysteine trifluoroacetate **3**

N-(*tert*-butoxycarbonyl)-*Se*-(4,5-dimethoxy-2-nitrobenzyl)selenocysteine (1.13 g, 2.44 mmol) was dissolved in TFA (10 ml) and stirred for 15 min at room temperature. After TFA was evaporated solid residue was washed with DCM, filtered and dried. Crude product was dissolved in methanol, insoluble impurities filtered off. Evaporation of methanol under reduced pressure yielded 0.9 g (77%) of target product as yellow solid (m.p. 155 °C (dec.)).

^{1}H NMR (400 MHz , DMSO-D_6): δ = 2.95-3.11 (2H, m, $SeCH_2$), 3.87 (3H, s, OCH_3), 3.92 (3H, s, OCH_3), 4.14 (1H, d *J* = 12 Hz, BnC$H_A$$H_B$Se), 4.17-4.25 (2H, m, BnC$H_A$$H_B$Se + CH), 7.15 (1H, s, ArH), 7.69 (1H, s, ArH), 8.37 (3H, brs, NH_3). ^{13}C NMR (100 MHz , DMSO-D_6): δ = 23.3, 25.7, 53.0, 56.5, 56.8, 109.3, 114.3, 131.0, 139.6, 147.9, 153.4, 170.2. IR (KBr): 3426 (OH), 3236, 3194, 3133 (NH), 1702 (CO). Anal. calcd. for $C_{14}H_{17}F_3N_2O_8Se$: C, 35.23; H, 3.59; N, 5.87. Found: C, 35.20; H, 3.83, N, 5.82. HRMS: m/z $[M+H]^+$ calcd. for $C_{12}H_{16}N_2O_6Se$: 365.0247; Found: 365.0255.

Synthesis of DMNB-Cys (Supplementary Scheme 1)

L-Cystine dimethyl ester **4**

Suspension of cystine (5 g, 20.8 mmol) in MeOH (100 ml) was cooled to -70 °C and thionyl chloride (3.75 ml, 19.24 mmol) was added to reaction vessel. Reaction mixture was left until room temperature was reached, then refluxed for 6 hours. Methanol evaporated under reduced pressure and residue dissolved in methanol. After addition of diethyl ether product precipitated as white crystals. Procedure gave 6.69g (94%) of target product as hydrochloride salt.

N,N'-bis(*tert*-butoxycarbonyl)-L-cystine dimethyl ester **5**

Reaction was performed following a previously described procedure[25].

N-(*tert*-butoxycarbonyl)-Lcysteine methyl ester **6**

Reaction was performed following a previously described procedure[25].

N-(*tert*-butoxycarbonyl)-*S*-(4,5-dimethoxy-2-nitrobenzyl)-L-cysteine methyl ester **7**

Suspension of *N*-(*tert*-butoxycarbonyl)cysteine methyl ester (1.94 g, 8.25 mmol), 4,5-dimethoxy-2-nitrobenzylbromide (2.40 g, 8.73 mmol) and sodium carbonate (1.51 g, 14.25 mmol) in THF (50 ml) refluxed for 60 hours, then quenched with water (50 ml). THF was evaporated under reduced pressure and product extracted with DCM. DCM evaporated, residue purified by column chromatography (C_6H_6:$CHCl_3$ 1:1, R_f = 0.23). Described procedure yielded 2.16 g (61%) of target product as yellow solid (m.p. 97–99 °C).

^{1}H NMR (400 MHz, $CDCl_3$): δ = 1.47 (9H, s, 3×CH_3), 2.91 (1H, dd *J* = 14, 5.6Hz, SCH_AH$_B$), 2.98 (1H, dd *J* = 14, 5.6Hz, SCH$_A$$H_B$), 3.78 (3H, s, OCH_3), 3.97 (3H, s, ArOCH_3), 4.03 (3H, s, ArOCH_3), 4.09 (1H, d *J* = 13.6Hz, BnCH_AH$_B$S), 4.17 (1H, d *J* = 13.6Hz, BnCH$_A$$H_B$S), 4.56 (1H, brm, CH), 5.35 (1H, d *J* = 7.6Hz, NH), 6.94 (1H, s, ArH), 7.69 (1H, s, ArH). ^{13}C NMR (100 Mz, $CDCl_3$): δ = 28.3, 34.4, 34.5, 52.7, 53.4, 56.4, 56.5, 80.3, 108.8, 113.4, 128.7, 140.5, 148.0, 153.0, 155.2, 174.4. IR (KBr): 3359 (NH), 1731 (CO), 1682 (CO). Anal. calcd. for $C_{18}H_{26}N_2O_8S$: C, 50.22; H, 6.09; N, 6.51; S, 7.45. Found: C, 50.39; H, 6.25; N, 6.53; S, 7.22. HRMS: m/z [M+Na]$^+$ calcd. for $C_{18}H_{26}N_2O_8S$: 453.1302; Found: 453.1310.

S-(4,5-dimethoxy-2-nitrobenzyl)-L-cysteine hydrochloride **8**

To suspension of *N*-(*tert*-butoxycarbonyl)-*S*-(4,5-dimethoxy-2-nitrobenzyl)cysteine methyl ester (0.95 g, 2.21 mmol) in mixture of acetone (25 ml) and water (25 ml) conc. hydrochloric acid (0.92 ml, 11.36 mmol) was added. Reaction mixture was stirred for 3 days at 60 °C (sand bath) temperature, solvents evaporated under reduced pressure, residue washed with acetone. Described procedure yielded 0.74 g (95%) of target product.

^{1}H NMR (400 MHz, DMSO-D_6): δ = 2.95 (1H, dd *J* = 14.6, 6.4Hz, SCH_AH$_B$), 3.01 (1H, dd *J* = 14.6, 4.8, SCH$_A$$H_B$), 3.87 (3H, s, OCH_3), 3.92 (3H, s, OCH_3), 4.12 (1H, d *J* = 13.6 Hz, BnCH_AH$_B$S), 4.15-4.25 (2H, m, BnCH$_A$$H_B$S + CH), 7.18 (1H, s, ArH), 7.70 (1H, s, ArH), 8.40 (3H, brs, NH_3). ^{13}C NMR (100 MHz , DMSO-D_6): δ = 31.3, 33.6, 52.2, 56.5, 56.8, 109.4, 114.7, 128.7, 140.2, 148.1, 153.2, 170.1. IR (KBr): 3425 (OH), 3365, 3209 (NH), 1720 (CO), 1676 (CO). Anal. calcd. for $C_{12}H_{17}ClN_2O_6S$: C, 40.85; H, 4.86; N, 7.94; S, 9.09. Found: C, 40.51; H, 4.90; N, 7.80; S, 8.83. HRMS: m/z [M+H]$^+$ calcd. for $C_{12}H_{16}N_2O_6S$: 317.0802; Found: 317.0805.

Bacterial, yeast strains and plasmid construction

Escherichia coli strain ER2267 (New England Biolabs) was used for plasmid manipulations. Yeast *Saccharomyces cerevisiae* strain LWUPF1Δ[27] was generously provided by L. Wang. Alkali cation method[39] was used to transform yeast cells. Plasmids pSNR-LeuRS[27] and pGFP-TAG[27] were a kind gift from L. Wang. Plasmid pescTrpLeuRSBH5T252A[19] was a kind gift from P. Schultz. Plasmid pRR6 (used to make cells shown in **Fig. 1c** columns 1-3 and 6) carries genes of the orthogonal pair of evolved *Ec*tRNA$^{\mathrm{Leu}}_{\mathrm{CUA}}$ and LeuRSBH5T252A. It was constructed by replacing the LeuRS gene in the pSNR-LeuRS plasmid with the mutated DMNB-Ser specific LeuRSBH5T252A gene from the pescTrpLeuRSBH5T252A plasmid using oligos FW29[27] and FW30[27] and applying standard molecular cloning and sequencing techniques. Plasmid pRR5 (used to make cells shown in **Fig. 1c** column 4) carries only orthogonal *Ec*tRNA$^{\mathrm{Leu}}_{\mathrm{CUA}}$ gene and no LeuRSBH5T252A. It was created by self-ligation of the XhoI, SpeI-digested and purified pSNR-LeuRS plasmid after blunting the XhoI and SpeI ends with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific). Plasmid pRR64 carries His_6-SUMOstar-M.HpaIICys103TAG gene under inducible *CUP*1 promoter and *URA*3 selectable marker. The M.HpaII gene was

PCR-amplified from a plasmid pET15b-hpaII[40] and cloned to pY-SUMOstar vector (LifeSensors) using BsaI and BamHI restriction sites added to the insert gene according to manufacturer's recommendations. The Cys103TAG mutation was introduced into M.HpaII gene by PCR-based site directed mutagenesis using Phusion High-Fidelity DNA polymerase (ThermoFisher Scientific). The original *TRP*1 marker in the construct was replaced with the *URA*3 marker from vector p426GPD[41] using standard genetic engineering techniques, and the final pRR64 construct was verified by sequencing.

Protein expression and isolation

EGFPTyr39DMNB-Sec or EGFPTyr39DMNB-Cys protein expression. Yeast LWUPF1Δ cells were cotransformed with plasmids pGFP-TAG and pRR6. Selected transformants were further cultivated in synthetic drop-out medium lacking Trp and Leu at 30 °C with 220 rpm shaking. Overnight yeast cells from 10 ml cultures were collected, transferred to 30 ml of fresh medium and incubated for 6 h. Cells were collected again, transferred to 30 ml of fresh medium containing 2 mM UAA and incubated in dark with shaking at 30 °C for 18 h.

EGFPTyr39Leu protein expression. Yeast LWUPF1Δ cells were co-transformed with plasmids pGFP-TAG and pSNR-LeuRS. Selected transformants were cultivated in synthetic drop-out medium lacking Trp and Leu at 30 °C with 220 rpm shaking until mid-log phase.

SUMO-M.HpaIICys103DMNB-Cys and SUMO-M.HpaIICys103DMNB-Sec protein expression. Yeast LWUPF1Δ cells were co-transformed with plasmids pRR64 and pRR6. Selected transformants were cultivated in synthetic drop-out medium lacking uracil and Trp (Yeast Nitrogen Base without amino acids and copper (Formedium)) as described for EGFPTyr39DMNB-Sec and EGFPTyr39DMNB-Cys protein synthesis, except that 100 mM $CuSO_4$ was added to induce His_6-SUMOstar-M.HpaII gene expression.

EGFP protein isolation. Cells were collected, washed twice with 0.9 % NaCl and lysed with a 2.5-fold excess of Y-PER reagent (Thermo Scientific) supplemented with Complete, EDTA-free protease inhibitor cocktail (Roche Diagnostics, GmbH) for 20 min at room temperature with orbital rotation. Lysates were clarified by 10 min spinning at x21,000 g. His-tagged proteins were isolated using unpacked Chelating Sepharose Fast Flow (GE Healthcare) according to manufacturer's instructions washing the sorbent with buffer A [20 mM K-PO_4, pH 7.4, 500 mM KCl, 20 mM imidazole, pH 7.4] and eluting the protein with buffer B [20 mM K-PO_4, pH 7.4, 500 mM KCl, 250 mM imidazole, pH 7.4]. To solution of the eluted protein EDTA (5 mM final conc.) was added followed by one buffer exchange with buffer C [20 mM K-PO_4, pH 7.4, 200 mM KCl, 5 mM EDTA] and one with buffer D [20 mM K-PO_4, pH 7.4, 200 mM KCl] using Amicon Ultra 10K centrifugal filter units (Millipore). Protein concentration was determined by 12% SDS-PAGE using BSA standard. Isolated protein solution was aliquoted and stored at -80 °C. Protein identity for all independent isolations was confirmed by HPLC/ESI-MS.

SUMOstar-M.HpaII protein isolation. Cells were collected, washed twice with 0.9 % NaCl and lysed with glass beads, after adding to the wet cell mass equal volume of glass beads and double volume of lysis buffer E [20 mM Na-PO_4, pH 7.4, 500 mM NaCl, 10% glycerol, 4 mM PMSF, 10 mM 2-mercaptoethanol, 2X Complete (EDTA-free protease inhibitor cocktail)]. Lysate was clarified by two spinnings at x21,000 g for 20 min, loaded onto His-Select spin columns (Sigma-Aldrich) and further treated according to manufacturer's instructions. Buffer exchanges of eluted protein, protein identification and storage were same as for EGFP proteins described above.

In vivo fluorescence assay

Yeast LWUPF1Δ cells were transformed with different combinations of plasmids and transformants were grown in appropriate selective media. Overnight cultures for EGFPTyr39DMNB-Sec and EGFPTyr39DMNB-Cys expression were inoculated (1:10) into medium containing 4 mM concentration of a corresponding UAA, other cultures were back diluted (1:150) into original medium. Cells were incubated in dark at 30 °C for 24 h with 220 rpm shaking. Cells were harvested, washed and resuspended in water. Fluorescence measurements (Ex 485 nm; Em 535 nm) were taken from 0.1 ml samples in triplicate using Synergy H4 (BioTek Instruments, Inc.) plate reader. Fluorescence values were normalized according to cell density (A_{595} readings).

Photolysis of photocaged proteins

Protein solution (0.04 mg/ml in buffer D [20 mM K-PO_4, pH 7.4, 200 mM KCl]) in UV-transparent microplates (BD Falcon Microtest 96-Well, Becton-Dickinson) was exposed to light generated by an Olympus IX70 microscope equipped with an excitation filter (330–385 nm bandpass) and a 100 W mercury lamp. Aliquots of 20 µl were typically subjected to HPLC/ESI-MS analysis. For reducing conditions, DTT was added to 50 mM before HPLC analysis. Time course analysis was performed by quantitation of corresponding deconvoluted ESI-MS peaks (Supplementary Figure S2) using GraphPad Prism software.

Semi-native SDS-PAGE

Fluorescent EGFP-Sec protein dimers formed during photochemical reaction were resolved using standard 12% SDS-PAGE gel, omitting the step of thermal sample denaturation. For reducing conditions, DTT was added to samples (50 mM final conc.) prior to loading on a gel. Gel was visualized with an FLA-5100 imaging system (Fujifilm).

Protein labeling

Sec-specific EGFP labeling. Protein solution (0.05 mg/ml) in buffer D [20 mM K-PO_4, pH 7.4, 200 mM KCl] was exposed to UV light for 30 min as described above and treated with DTT (10 mM final conc.) for 5 min. The reaction was then adjusted to pH 5 with NaOAc (100 mM final conc.) and maleimide-PEG_2-biotin(Thermo Scientific) was added at a 360–380-fold molar excess over protein under argon atmosphere and reaction incubated for 10 min at room temperature. Reaction was quenched by addition of 2-mercaptoethanol to 50 mM followed by three exchanges (10-fold concentration-dilution using Amicon Ultra 10K centrifugal filter units (Millipore)) with buffer F [30 mM NaOAc, pH 5, 200 mM NaCl, 50 mM 2-mercaptoethanol], two exchanges with buffer G [20 mM Na-PO_4, pH 7.4, 200 mM NaCl, 50 mM 2-mercaptoethanol] and two exchanges with buffer H [20 mM Na-PO_4, pH 7.4, 200 mM NaCl]. A 20 µl sample of the resulted protein solution was analyzed by HPLC/ESI-MS.

Cys-specific EGFP labeling at pH 7.0 (Control). Protein solution (0.05 mg/ml) in buffer I [20 mM K-PO_4, pH 7.0, 200 mM KCl] was exposed to UV light for 30 min as described. Maleimide-PEG_2-biotin was added at a 360-fold molar excess over protein and reaction was incubated under argon at room temperature for 10 min. Reaction was quenched by addition of 2-mercaptoethanol to 50 mM followed by an exchange with buffer G [20 mM Na-PO_4, pH 7.4, 200 mM NaCl, 50 mM 2-mercaptoethanol] and two exchanges with buffer H [20 mM Na-PO_4, pH 7.4, 200 mM NaCl] as described above.

Protein HPLC/ESI-MS

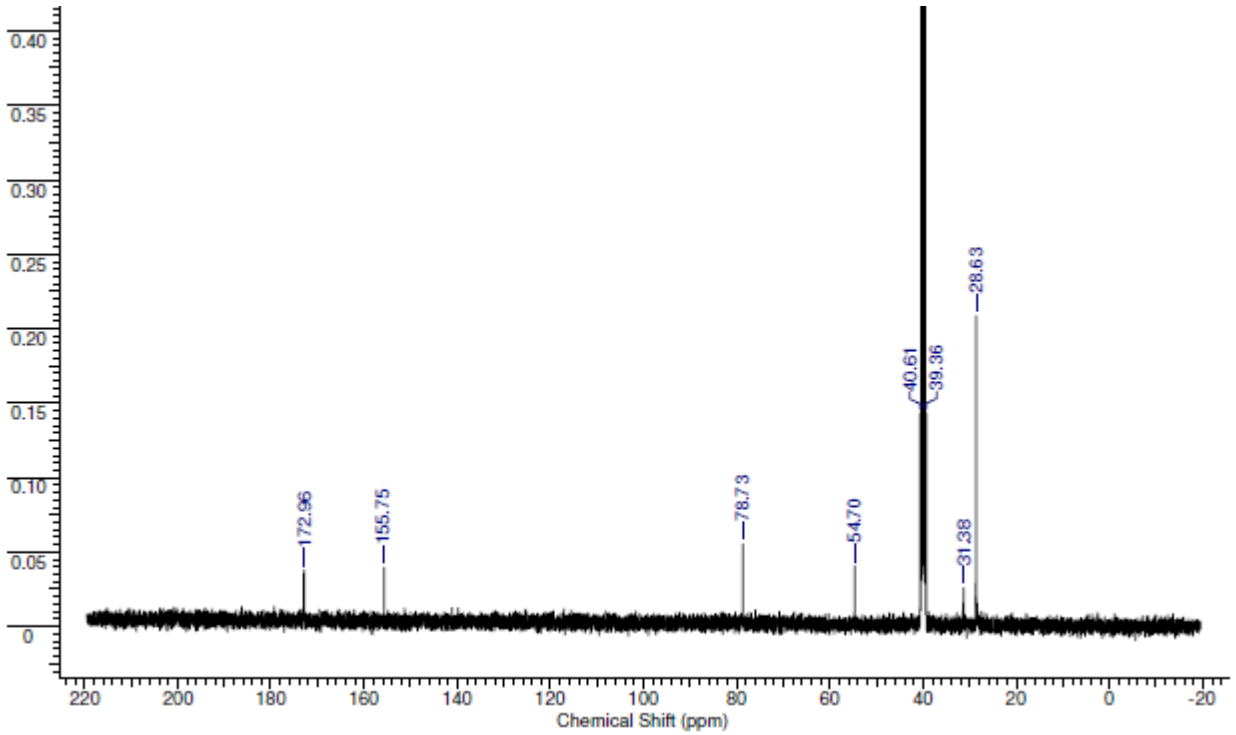
Samples were analyzed on an integrated HPLC/ESI-MS system (Agilent 1290 Infinity) equipped with a Poroshell 300SB-C8 column (2.1x75 mm, 5 µm) by elution with a linear gradient of solvents A (1% formic acid in water)

and B (1% formic acid in acetonitrile) at a flow rate of 0.4 ml/min at 30 °C as follows: 0–1 min, 2% B; 1–6 min, 2–98% B; 6–7 min, 98% B; 7–9 min, 98–2% B; 9–10 min, 2% B. High-resolution mass spectra of protein products were acquired on an Agilent Q-TOF 6520 mass analyzer (100–3200 m/z range, positive ionization mode, sampling window 9.7–13.5 sec, scan rate 1.1/sec). The results were analyzed with Agilent MassHunter Qualitative Analysis software. Protein mass spectra were deconvoluted (maximum entropy) in the range of 20,000–70,000 Da; mass step 0.5 Da; S/N threshold 30; m/z range 700–2200; average mass 90 % peak height; isotope width automatic; minimum consecutive charge states 5; minimum protein fit score 8. Theoretical mass values were obtained using ProtParam tool (web.expasy.org/protparam/) with manual adjustments for EGFP chromophore formation[42], EGFP-Sec dimer formation, and maleimide-PEG_2-biotin coupling with corresponding EGFP variants.

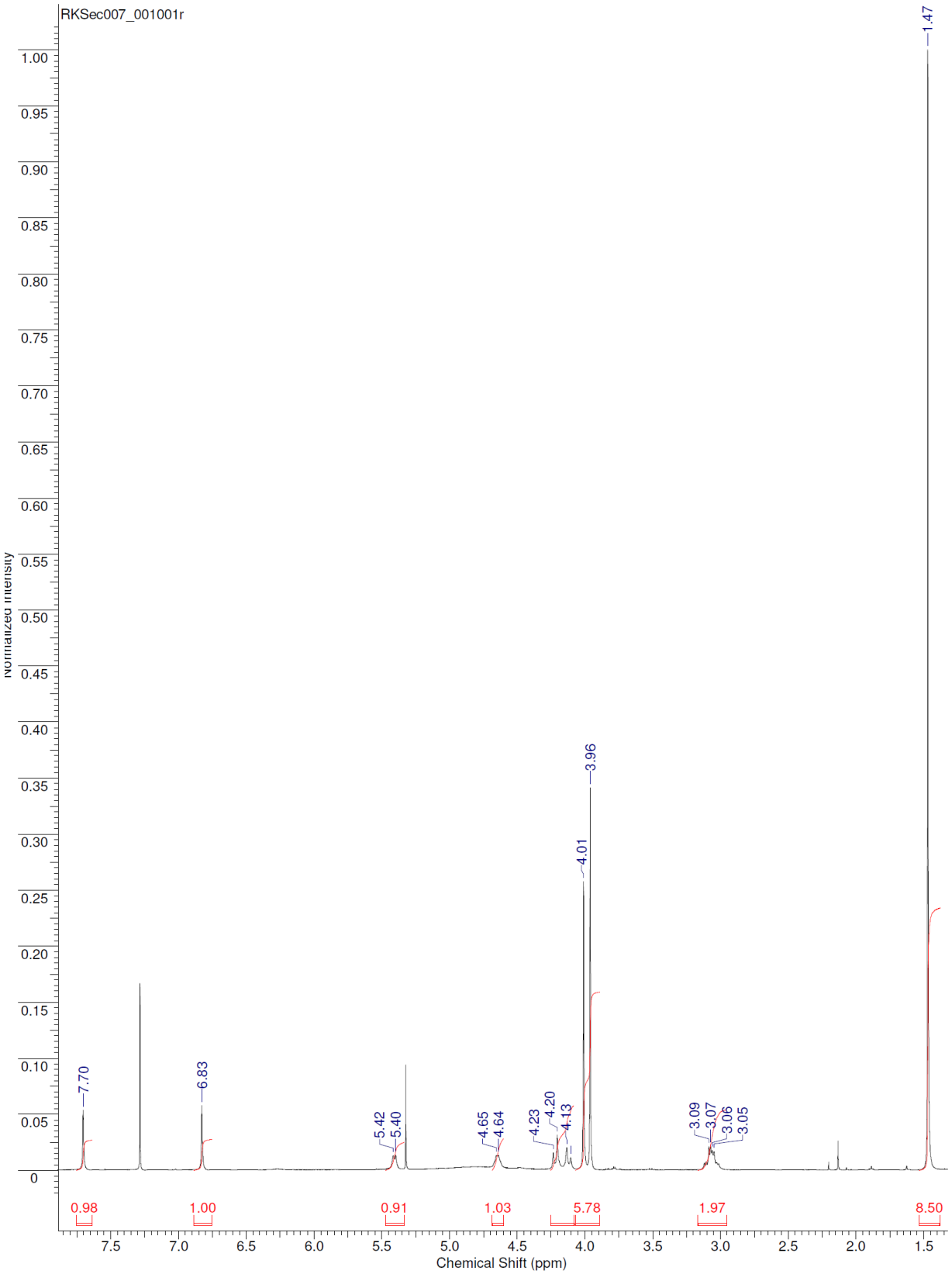
Supplementary References

38. Roosenboom, M., Vermeulen, N.P.E., Andreadou, I. & Commandeur J.N.M. *J. Pharmacol. Exp. Ther.* **294**, 762–769 (2000).
39. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. *J. Bacteriol.* **153**, 163-168 (1983).
40. Lukinavičius, G., Lapinaitė, A., Urbanavičiūtė, G., Gerasimaitė, R. & Klimašauskas, S. *Nucleic Acids Res.* **40**, 11594-602 (2012).
41. Mumberg, D., Müller, R. & Funk, M. *Gene* **156**, 119-22 (1995).
42. Ormö, M. et al. *Science* **273**, 1392-1395 (1996).

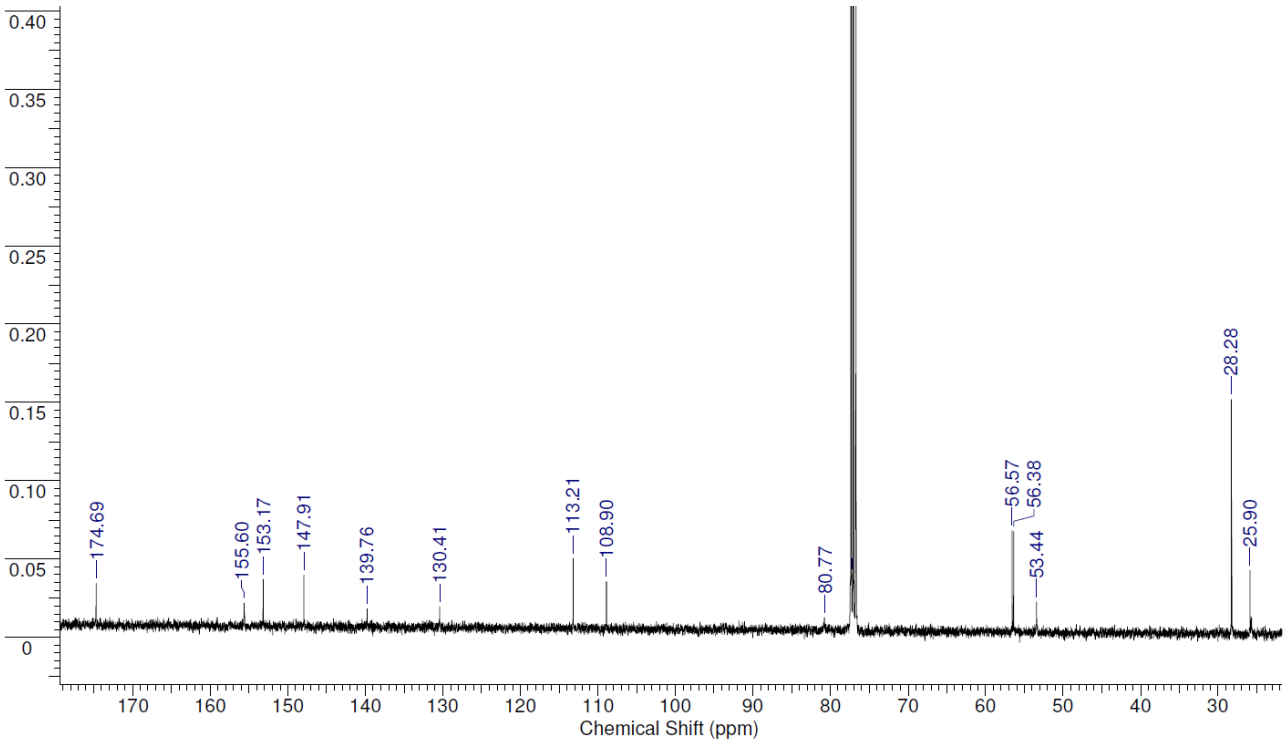
Supplementary Spectra



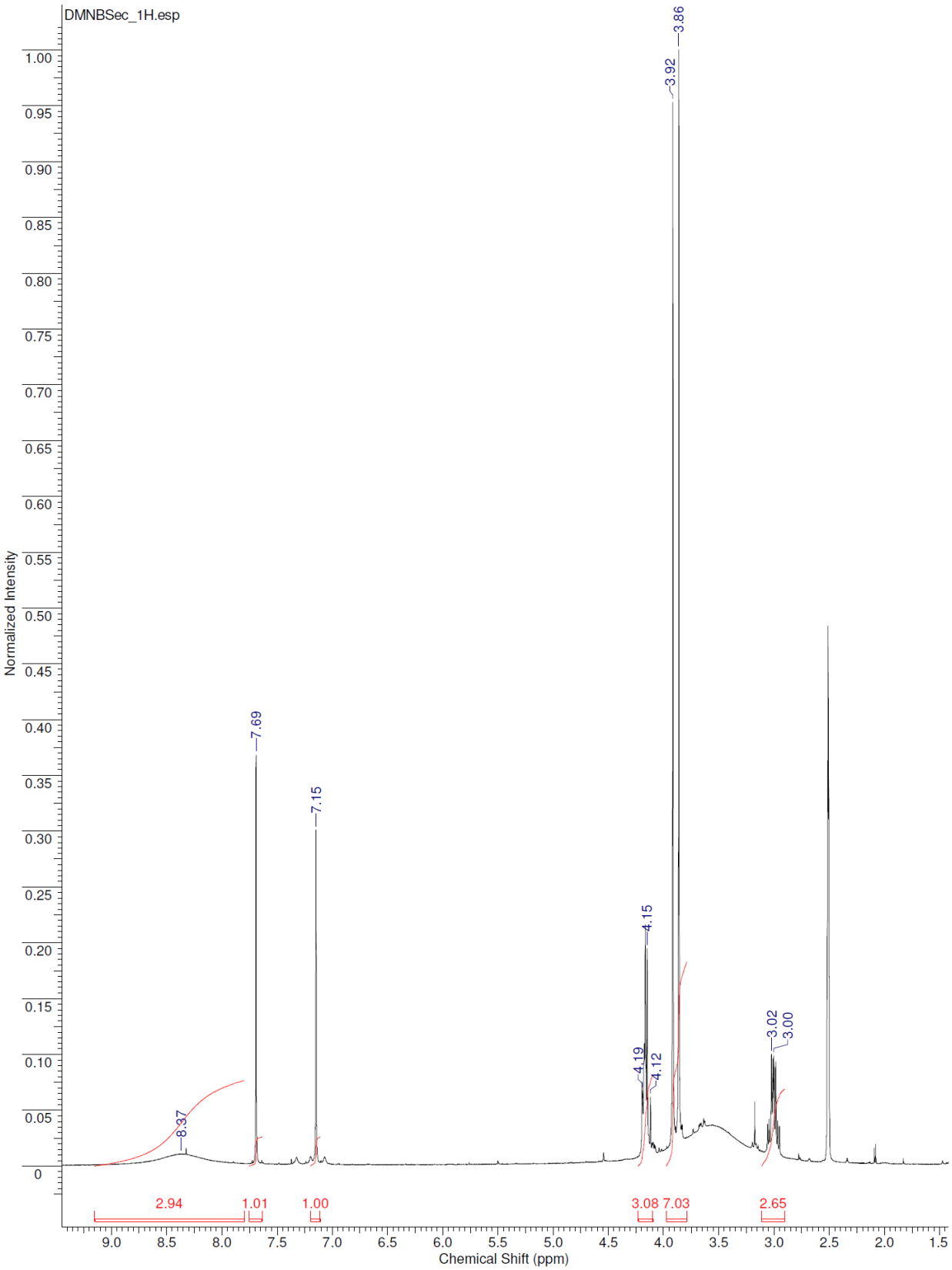
^{13}C NMR of compound **1**



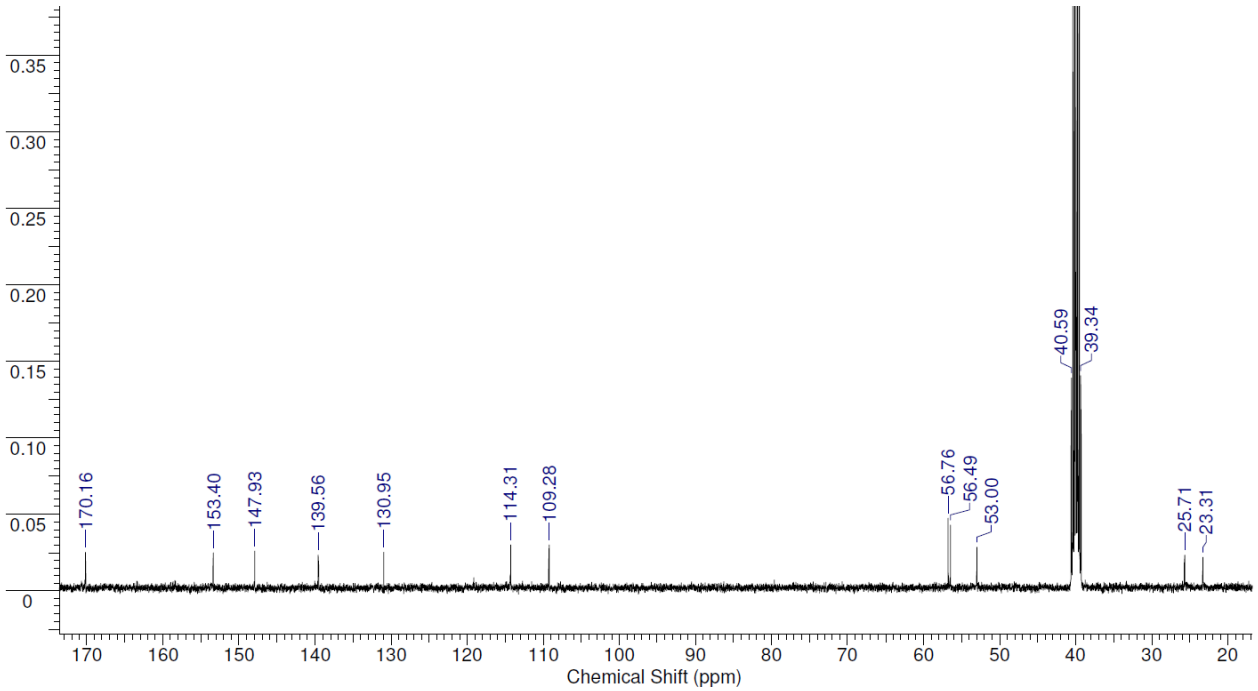
^{1}H NMR of compound **2**



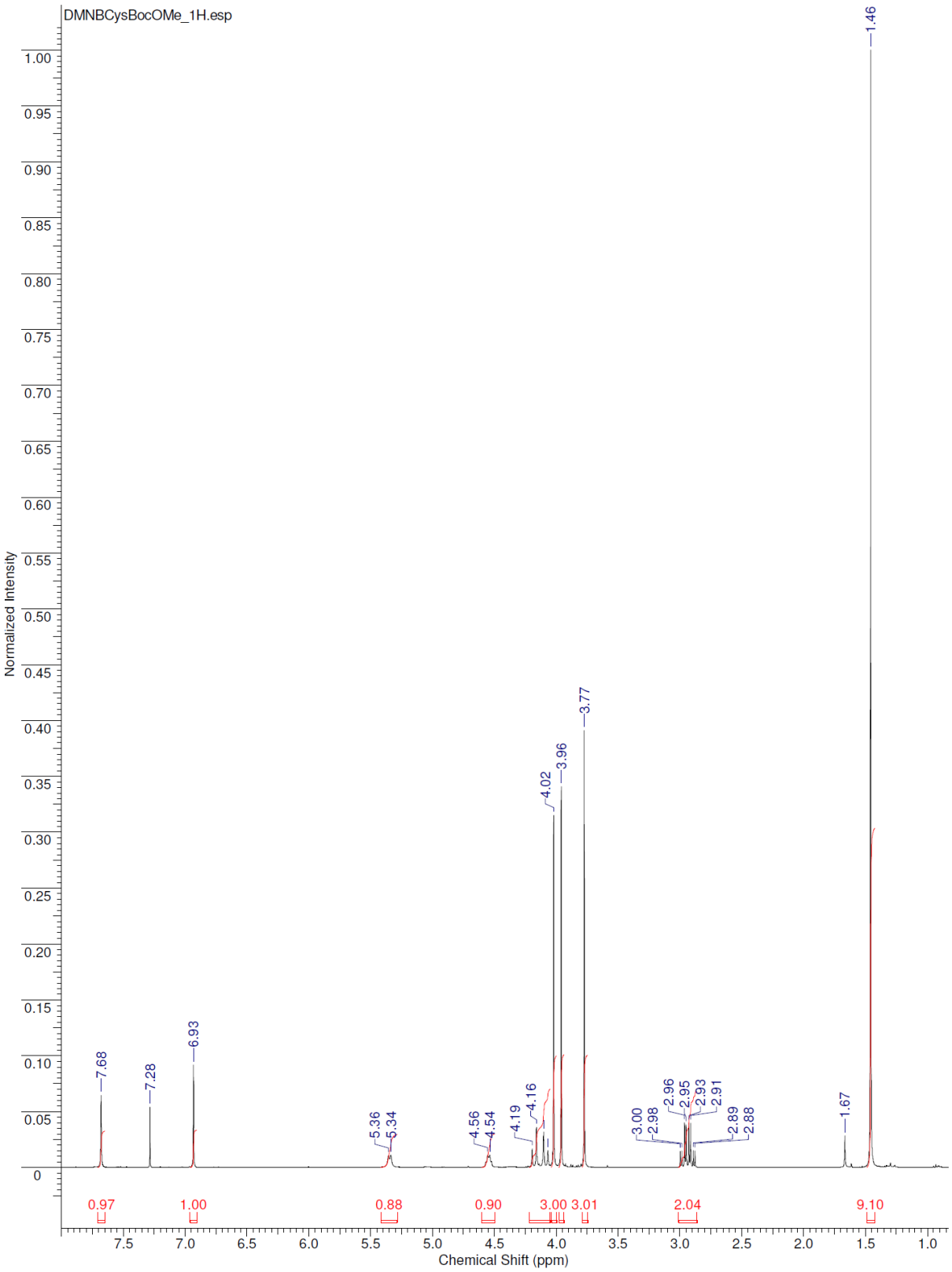
^{13}C NMR of compound **2**



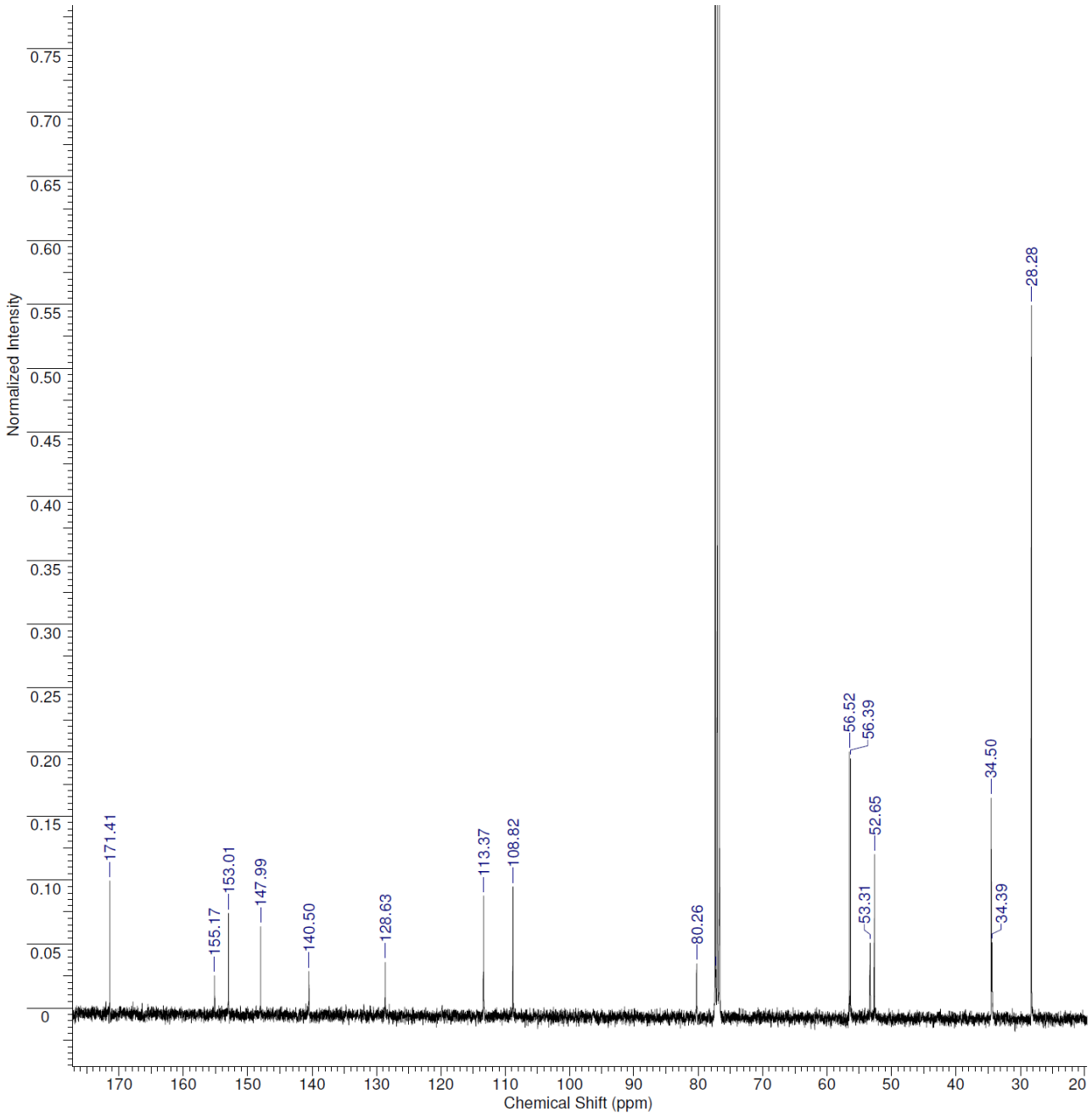
^{1}H NMR of compound **3**



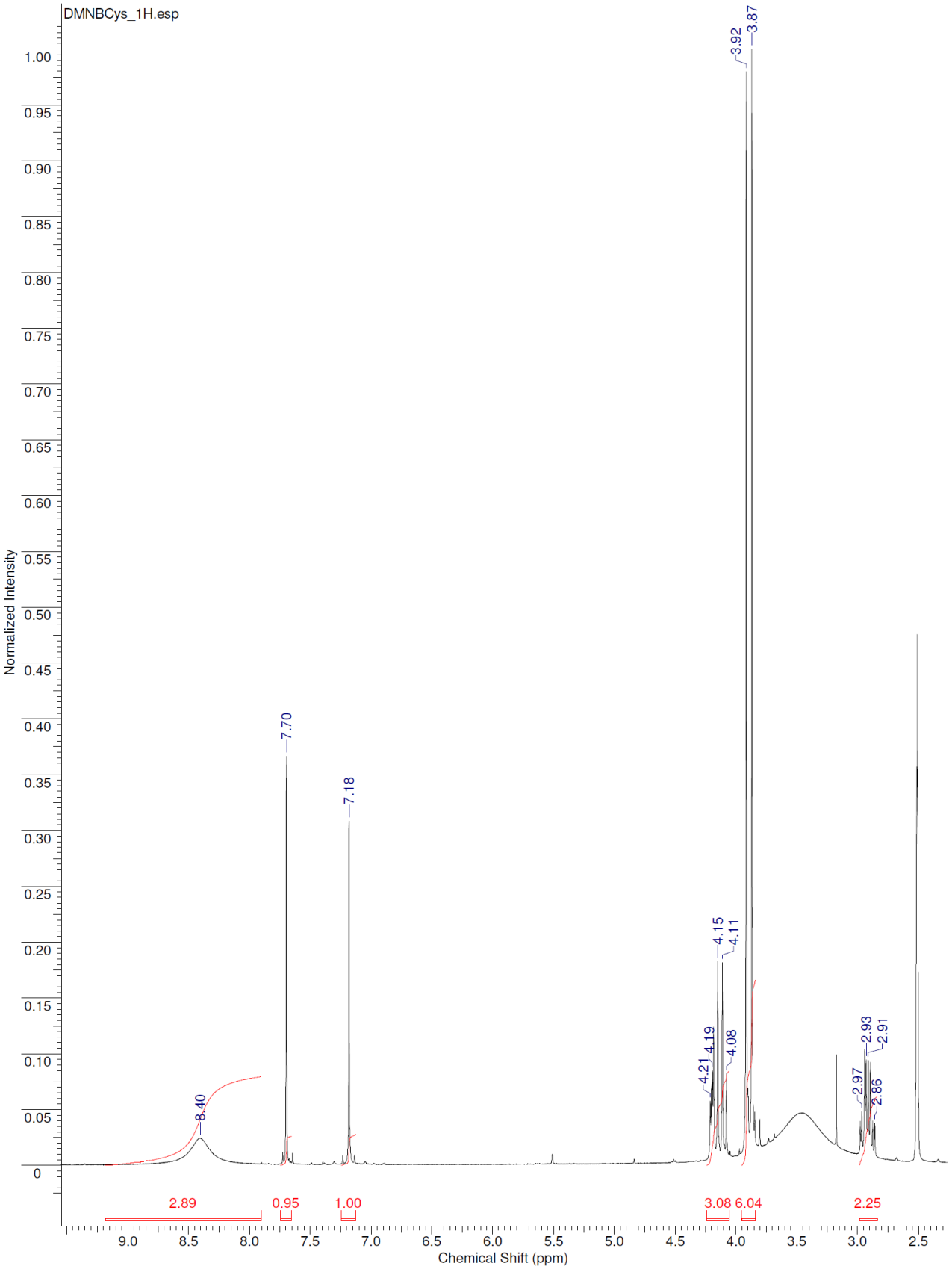
^{13}C NMR of compound **3**



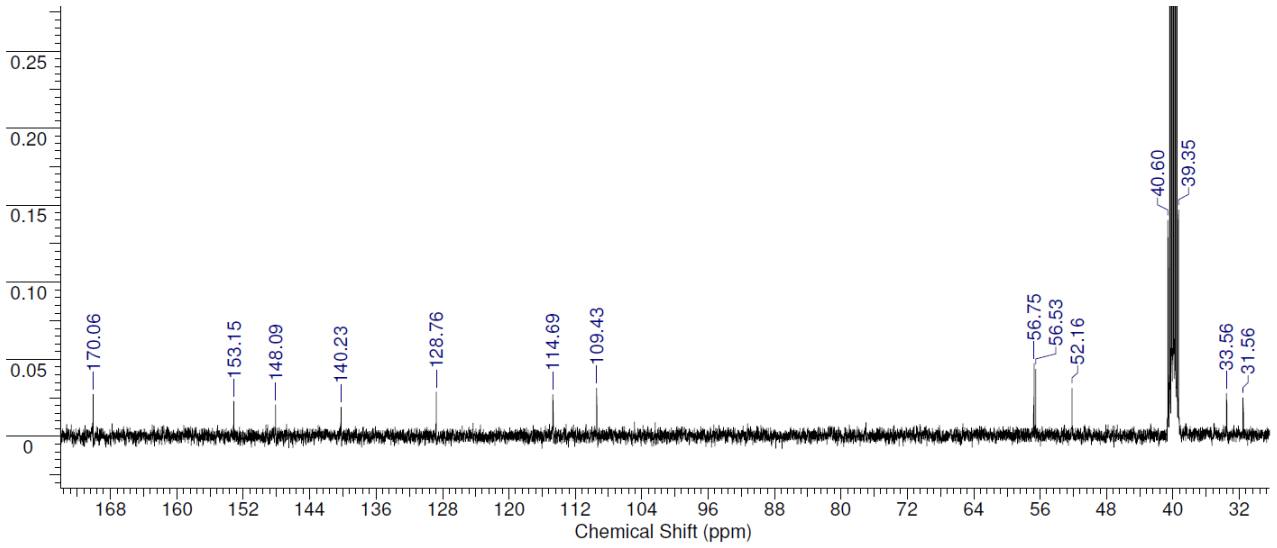
^{1}H NMR of compound **7**



^{13}C NMR of compound **7**



1H NMR of compound **8**



13C NMR of compound **8**