Supporting Information

The semisynthesis of nucleolar human selenoprotein H

Rebecca Notis Dardashtia,[a] Shay Laps,[a] Jacob S. Gichtin[a] and Norman Metanis[a,b,c,*]

[a]The Institute of Chemistry, [b]Casali Center for Applied Chemistry, [c]The Center for Nanoscience and Nanotechnology, The Hebrew University of Jerusalem Edmond J. Safra Campus, Givat Ram, Jerusalem 9190401 (Israel)

* Correspondence author, E-mail: metanis@mail.huji.ac.il

Materials and Methods

Materials

Buffers were prepared using MilliQ water (Millipore, Merck). Any buffers containing TCEP·HCl or sodium ascorbate were prepared fresh each day of use. Ultrapure guanidinium chloride (Gn·HCl, Apollo Scientific, Manchester, UK) was used in all ligation reactions. $Na_2HPO_4 \cdot 12H_2O$, tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl), ethanedithiol (EDT), triisopropylsilane (TIPS), *D,L*-dithiothreitol (DTT), 2,2'-Dithiobis(5-nitropyridine) (DTNP), sodium ascorbate, and 4-mercaptophenylacetic acid (MPAA) were purchased from Merck (formerly Sigma-Aldrich, Rehovot, Israel). 2-chlorotrityl chloride resin (1.5 mmol/g) and all Fmoc-amino acids were obtained from Chem-Impex Intl. Inc, amino acids were purchased with the following side chain protecting groups: Arg(Pbf), Asp(OtBu), Glu(OtBu), Ser(tBu), Thr(tBu), Cys(Trt), Lys(Boc), Tyr(tBu), Asn(Trt). (Pbf = 2,2,4,6,7- pentamethyl-2,3-dihydrobenzofuran-5-sulfonyl).

1-[Bis(dimethylamino)methylen]-5- chlorobenzotriazolium 3-oxide hexafluorophosphate, *N,N,N',N'*-Tetramethyl-O-(6- chloro-1*H*-benzotriazol-1-yl)uronium hexafluorophosphate (HCTU) and Ethyl cyano(hydroxyimino)acetate (OxymaPure) were purchased from Luxembourg Biotechnologies Ltd. (Rehovot, Israel). All solvents: *N,N*-dimethylformamide (DMF), dichloromethane (DCM), acetonitrile (ACN), *N,N*-diisopropylethyl amine (DIEA), Trifluoroacetic acid (TFA), and piperidine (Pip) were purchased from BioLab (Jerusalem, Israel) and were peptide synthesis, HPLC, or ULC-grade.

Instruments

Analytical and semi-preparative reversed-phase (RP) HPLC analyses were performed on a Waters Alliance HPLC with UV detection (220 nm and 280 nm) using an XSelect C18 column (3.5 μm, 130 Å, 4.6 × 150 mm), XBridge C4 column (3.5 μm, 4.6 × 150 mm) or XBridge

BEH300 C4 column (5 µm, 19 × 150 mm). Preparative and semi-preparative RP-HPLC were performed on a Waters 150Q LC system using a XSelect C18 column (5 µm, 30 × 250 mm), Xbridge C4 column (5 µm, 19 × 250 mm), or XBridge C4 column (5 µm, 10 × 150 mm).
Linear gradients of ACN (with 0.1 % TFA, buffer B) in water (with 0.1 % TFA, Buffer A) were used for all systems to elute bound peptides. The flow rates were 1 mL/min (analytical, column heated at 30 °C), 3.4 mL/min (C4 semi-preparative, column heated to 30 °C), and 20 mL/min (C18 preparative).
ESI-MS was performed on LCQ Fleet Ion Trap mass spectrometer (Thermo Scientific). Peptide masses were calculated from the experimental mass-to-charge (m/z) ratios from the observed multiply-charged species of a peptide.
The HR-MS were recorded on a Q-ExactivePlus Orbitrap mass spectrometer (Thermo Scientific) with an ESI source and 140'000 FWHM, in a method with AGC target set to 1E6, and scan range of 400-2800 m/z.
Deconvolution of the experimental MS data was performed with the help of MagTran v1.03 software.
Anaerobic reactions were performed in an anaerobic chamber from Coy Labs in an atmosphere of 95% N_2:5% H_2 with 0 ppm oxygen as monitored with a gas detector.
Protein purifications were performed using an Äkta Explorer FPLC system (GE Healthcare).
DNA-protein interactions were measured using biolayer interferometry on a Gator® prime instrument (Gator Bio, Inc.).

General SPPS Protocols Used

Direct loading of amino acid onto resin: 0.25 mmol resin was washed with DMF and DCM, then swelled in 10 mL of 1:1 DMF:DCM for 1 h and drained. 1 mmol (4 equiv) of Fmoc-protected amino acid was dissolved in 5 mL DMF, to which 39.4 µL DIEA (1 mmol, 4 equiv) was added. The amino acid solution was added to the swelled resin, which was shaken for 1.5 hours, then drained and washed well with DMF. Unreacted groups were capped by shaking for 10 min with 50% MeOH in DMF, after which the resin was again washed well with DMF. Substitution number of the resin was then calculated, as described below.
Hydrazide activation of resin: 2-chlorotrityl chloride resin was activated with hydrazide as a thioester precursor according to the method of Liu et al..[6] In short, 0.25 mmol resin was washed with DMF and DCM, then swelled in 10 mL 1:1 DMF:DCM for 1 h and drained. 4 mL of 5% v/v NH_2NH_2 in DMF was added and shaken with resin for 30 min and drained. A fresh portion of 4 mL 5% v/v NH_2NH_2 was added and shaken for additional 30 min. Following draining and

washing, the unreacted sites on resin were capped by shaking for 10 min with 4 mL of 5% v/v MeOH in DMF. Resin was washed well and taken directly to double-coupling with the appropriate amino acid according to standard methods (see below) after which the substitution number was calculated.

Calculating substitution number of pre-loaded resin: After the first Fmoc-amino acid was coupled to resin, the resin was washed well with DCM and dried overnight on the lyophilizer. 3 ~5 mg portions of the dried peptidyl-resin were measured in 3 separate eppendorf tubes and the exact weights of each sample were recorded. Each was deprotected in 1 mL 20% piperidine in DMF for 20 min and the eppendorfs were spun down. 100 µL of supernatant was diluted in 2 mL DMF. Absorbance at 301 nm was recorded for each sample and substitution was calculated for each aliquot according to the following:

Substitution = $21(A_{301})/7.8$(weight of resin(mg))

From this resin, the appropriate amount for the desired scale was used for peptide synthesis.

SPPS: Standard Fmoc-SPPS was performed at 0.25 mmol scale unless otherwise noted. Fmoc-deprotections were performed twice in 20% piperidine in DMF for five minutes and then ten minutes. For automated synthesis, Fmoc-amino acids (2 mmol, 8 equiv) were activated for 3 min with HATU or HCTU (2 mmol, 8 equiv in 5 mL DMF) and DIEA (4 mmol, 16 equiv in 5 mL DMF), then shaken to couple for 30 minutes. Prior to coupling amino acids containing Se, peptidyl resin was dried and divided into two 0.125 mmol portions.

Se-containing amino acids coupling was performed on 0.125 mmol scale: Fmoc-Sec(Mob)-OH or Fmoc-Sez-OH (0.19 mmol, 1.5 equiv) and OxymaPure (26.7 mg, 0.19 mmol, 1.5 equiv) were dissolved in 6 mL 1:1 DMF:DCM, cooled to 0°C, activated with DIC (27.1 µL, 0.175 mmol, 1.4 equiv) for 5 minutes, and then shaken with peptidyl-resin to couple for 3 hours. Reaction efficiency was monitored using small cleavage followed by HPLC and MS and, when necessary, double-coupling was performed.

0.125 mmol of dried peptidyl-resin was cleaved in 10 mL standard cleavage cocktail (95% TFA, 2.5% TIPS, 2.5% H_2O) or, if Mob-protecting group was present, modified cleavage cocktail (2.5% H_2O, 2.5% TIPS, and 95% TFA with 2 equiv DTNP[7] and 5 equiv $MeONH_2 \cdot HCl$) for 3 h, after it was found that addition of $MeONH_2 \cdot HCl$ prevented unwanted formylation of Sec.[1] After the TFA was evaporated through N_2 bubbling, peptide was precipitated with cold ether, centrifuged, resuspended in 30% ACN in H_2O, and lyophilized to dryness.

All peptides were purified via preparative HPLC. SELENOH(91-122)(A91U) was treated with 50 mg TCEP and 120 mg sodium ascorbate to reduce any selenosulfide bonds between the free Sec and TNP prior to HPLC purification.

Synthetic Strategy for 3-Segment Assembly of SELENOH

Below is the sequence of SELENOH, obtained from the UniProt database:

```
                10                  20                  30                  40                  50
MAPRGRKRKA  EAAVVAVAEK  REKLANGGEG  MEEATVVIEH  CTSURVYGRN
                60                  70                  80                  90                 100
AAALSQALRL  EAPELPVKVN  PTKPRRGSFE  VTLLRPDGSS  AELWTGIKKG
               110                 120
PPRKLKFPEP          QEVVEELKKY LS
```

Ligation sites are in bold and underlined.

Methods used to obtain each segment are provided below, in detail. In brief:

The N-terminal segment, SELENOH(2-43), was obtained using recombinant protein expression, with an N-terminal fusion to SUMO and C-terminal fusion to intein MxeGyrA. Subsequent to purification, this segment was converted in buffer to a thioester.

The middle segment, SELENOH(44-90), was produced as a C-terminal thioester from C-terminal hydrazide using the methods developed by Flood et al.[2] Sec44 was replaced with Fmoc-Sez as a protecting group.

The C-terminal segment, SELENOH(91-122), was synthesized on 2-chlorotrityl chloride resin, and Ala91 was temporarily replaced with Sec.

SELENOH(91-122)(A91U)

481 mg SELENOH(91-122)(A91U) was purified in four batches via RP-HPLC in a C18 column with gradient of 31%-56% ACN in H_2O with 0.1% TFA over 55 minutes. Purified peptide was characterized using analytical HPLC and ESI-MS(Figure S1). 35.6 mg total (8.0% yield) was obtained.

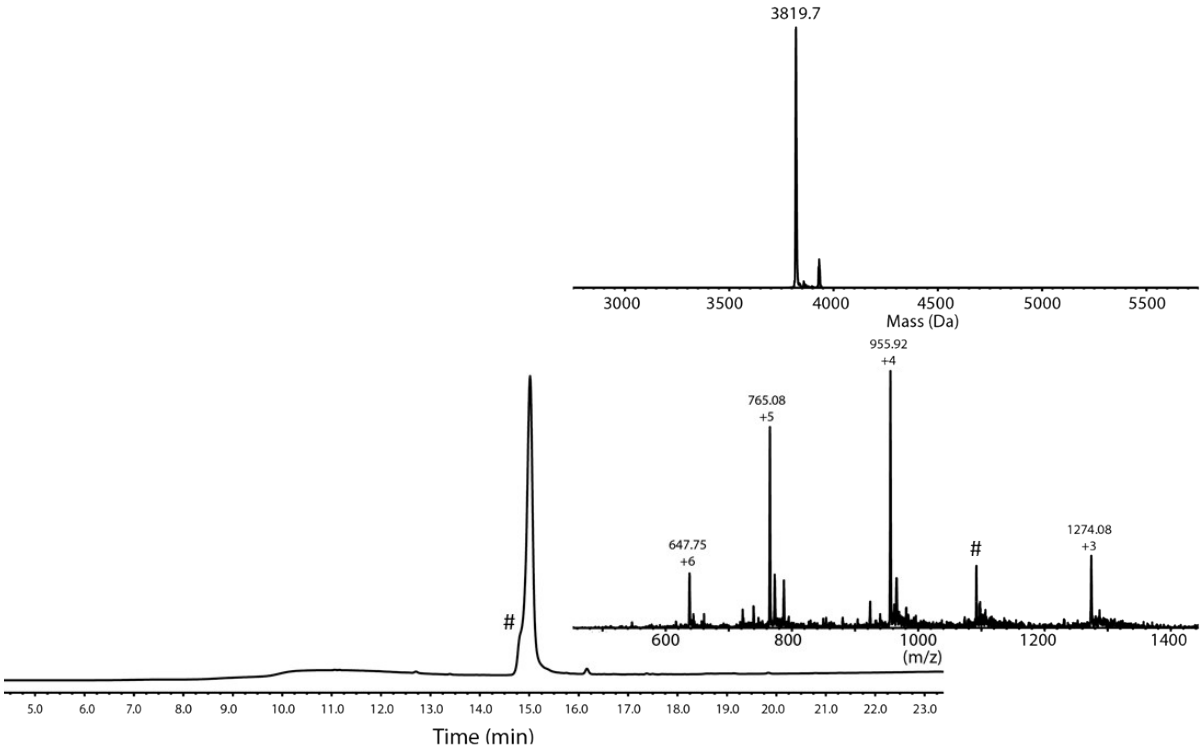

Figure S1. HPLC trace and ESI-MS spectrum of purified SELENOH(91-122)(A91U). Calc. mass 3818.4 Da, obs. mass 3819.7 Da. # shows an impurity which was removed in subsequent ligation.

SELENOH(44-90)(U44Sez)-COSR

91.1 mg crude SELENOH(44-90)(U44-Fmoc-Sez)-$CONHNH_2$ was converted to MPAA thioester before purification. To convert hydrazide to thioester,[2] crude peptide was dissolved in 8.4 mL thioesterification buffer (6 M GdmCl, 0.2 M sodium phosphate, pH 3.3). 283 mg MPAA was added to make a suspension (supposed final concentration 200 mM, final pH 2), and the reaction was initiated with the addition of 17.2 µL acetyl acetone (10 equiv) and incubated at room temperature overnight with stirring. The reaction mixture treated with 950 mg TCEP to reduce any oxidized MPAA (which coeluted with product in HPLC), and was then centrifuged to pellet solid MPAA. Supernatant was filtered and injected to preparative RP-HPLC in a C4 column with an initial isocratic flow of 10% ACN for 20 min to remove excess MPAA, then a gradient of 30%-50% ACN in H_2O with 0.1% TFA over 40 minutes. 8.7 mg pure (9.1% yield) was obtained, and characterized with ESI-MS and analytical HPLC (Figure S2).

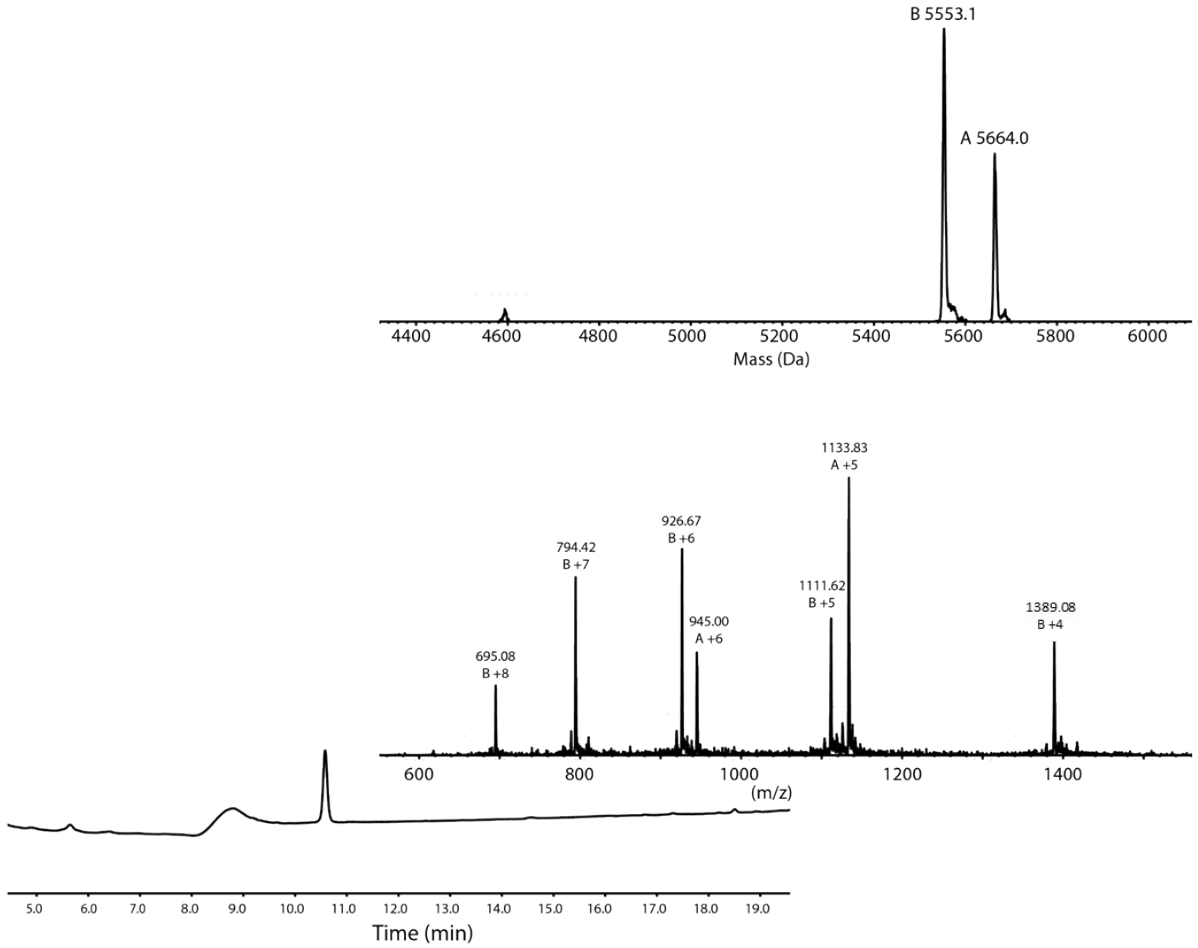

Figure S2. HPLC trace and ESI-MS spectrum of purified SELENOH(44-90)(U44-Fmoc-Sez)-MPAA. Calc. mass 5553.02 Da, obs. mass 5553.1 Da, 5564.0 Da (+TFA salt).

SELENOH(44-122): One-pot ligation, deselenization, Fmoc-deprotection and Sez opening

In an anaerobic chamber, 5.0 mg SELENOH(91-122)(A91U) (1.3 µmol) and 8.7 mg Fmoc-SELENOH(44-90)(U44Fmoc-Sez)-MPAA (1.6 µmol) were dissolved in 414 µL selenoligation buffer (6 M GdmCl, 200 mM sodium phosphate, pH 6.5). 19 µL of 0.01 mg/µL TCEP was added (0.7 µmol, 0.5 equiv) to initiate thioesterification, for a final peptide concentration of 3 mM. The reaction was monitored via HPLC and an additional 0.5 equiv TCEP was added after several hours to enable the reaction to continue. Once ligation was complete, deselenization was performed by adding 18.6 mg TCEP (65 µmol, 50 equiv) to the reaction and raising pH to 6.5 with 70 µL of 2 M NaOH. After 14 hours, deselenization was complete as determined by analytical HPLC. Reaction was removed from anaerobic chamber and 14 mg sodium ascorbate was added to protect Sez from deselenization as it opened. 211 µL of a stock solution containing 68% piperidine in water at pH 11 was added to the reaction to give a final concentration of 20% piperidine. The reaction was adjusted to pH 10 with 200 µL of 6 M HCl

to initiate Sez opening, and then incubated at 37 °C. After 4 hours, all Fmoc-Sez had been converted to Sec, as determined by analytical HPLC and ESI-MS (Figure 1). The reaction was purified using semi-prep C4 in a 20%-50% gradient. In total, 2.5 mg of SELENOH(44-122) was obtained (21.6% yield from one-pot reactions), as determined by A_{280} (ε_{280} = 8250 $M^{-1}cm^{-1}$).

SUMO3-SELENOH(2-43)-MxeGyrA Expression and Purification

A gene containing the sequence for SELENOH(2-43)-MxeGyrA was ligated into a pETM-SUMO3-GFP vector via Gibson assembly.

We designed a DNA sequence encoding SELENOH(2-43)-MxeGyrA and ordered it from IDT as a gBlock® Gene Fragment. The SELENOH sequence was obtained from the PDB, while MxeGyrA sequence was replicated from previous works by Shah et al..[8]

Designed amino acid sequence:

SELENOH(2-43)

MxeGyrA Intein

APRGRKRKAEAAVVAVAEKREKLANGGEGMEEATVVIEHCTS*CITGDALVALPE GESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTA NHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVP GLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGF VSHATEA*

DNA sequence (optimized for E.coli):

GCA CCG CGT GGC CGT AAA CGC AAA GCG GAA GCC GCG GTA GTT GCA GTA GCG GAA AAG CGC GAG AAG TTA GCT AAT GGC GGT GAG GGG ATG GAG GAA GCT ACC GTA GTT ATT GAA CAC TGC ACT TCG TGC ATT ACT GGT GAT GCG T TA GTA GCA CTT CCG GAA GGC GAG AGT GTC CGC ATC GCA GAC ATT GTA CC G GGC GCG CGT CCT AAC TCG GAC AAC GCA ATT GAC TTG AAA GTG TTG GA C CGC CAT GGG AAT CCG GTG TTG GCT GAT CGT CTT TTT CAT TCT GGC GAG CAT CCC GTC TAT ACG GTC CGC ACT GTG GAA GGT CTT CGT GTG ACG GGT A CT GCT AAC CAT CCT TTA CTG TGC TTG GTA GAT GTT GCC GGT GTG CCC ACG TTG CTG TGG AAG TTA ATT GAC GAA ATT AAG CCA GGA GAC TAT GCC GTC ATC CAA CGC TCA GCT TTT AGC GTA GAC TGT GCA GGA TTC GCG CGT GGT A AG CCT GAA TTC GCT CCC ACC ACG TAT ACC GTG GGG GTC CCT GGG TTG GT C CGT TTC TTG GAG GCA CAT CAT CGT GAC CCC GAC GCT CAG GCA ATT GCA GAT GAA CTG ACG GAT GGA CGT TTC TAC TAC GCT AAA GTG GCG TCG GTA A CA GAT GCA GGC GTT CAG CCG GTG TAC AGC TTA CGT GTT GAC ACG GCA G AC CAT GCC TTT ATC ACG AAC GGT TTT GTA AGC CAC GCA ACT GAG GCC

The DNA was amplified by PCR with the following primers:

5′- GCACCGCGTGGCCGTAAACGC-3’,
5′- GGCCTCAGTTGCGTGGCTTAC-3′.

Simultaneously, the pETM-SUMO3-GFP vector was linearized via PCR with primers that introduced overlapping ends with the desired gene:

5'- CACGCAACTGAGGCCtgaaagcttgcggccgcactcgagc-3'
5'- GTTTACGGCCACGCGGTGCggatccaccggtctgttgctg-3'

The designed gene was inserted into the linearized plasmid via Gibson assembly and the product was transformed into competent Escherichia coli XL-10 cells; bacterial colonies were screened for the presence of the gene using PCR. Positive colonies were further verified by sequencing. Following Gibson assembly, the vector contained a gene to express the following protein:

His_6-SUMO3
SELENOH(2-43)
MxeGyrA Intein

MKHHHHHHPMSDYDIPTTENLYFQGAMGNDHINLKVAGQDGSVVQFKIKRHTPLS
KLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGS**APR**
GRKRKAEAAVVAVAEKREKLANGGEGMEEATVVIEHCTS*CITGDALVALPEGESV*
RIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTANHPL
LCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPGLVR
FLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSHA
TEA

Site-directed mutagenesis was performed to obtain the S43A and S43T variants of the above protein.

S43A: The following primers were used to obtain the S43A variant:

5'-CACTGCACTgCGTGCATTACTGGTGATG-3'
5'- GTAATGCACGcAGTGCAGTGTTCAATAAC-3'
Mutagenic PCR products were verified with genetic sequencing. The resulting amino acid sequence was obtained:

MKHHHHHHPMSDYDIPTTENLYFQGAMGNDHINLKVAGQDGSVVQFKIKRHTPLS
KLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGS**APR**
GRKRKAEAAVVAVAEKREKLANGGEGMEEATVVIEHCTA*CITGDALVALPEGES*
VRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTANHP
LLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPGLV
RFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSH
ATEA

S43T: The following primers were used to obtain the S43T variant:

5'-CACTGCACTaCGTGCATTACTGGTGATG-3'
5'-GTAATGCACGtAGTGCAGTGTTCAATAAC-3'

Mutagenic PCR products were verified with genetic sequencing. The resulting amino acid sequence was obtained:

MKHHHHHHPMSDYDIPTTENLYFQGAMGNDHINLKVAGQDGSVVQFKIKRHTPLS
KLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGS**APR**
GRKRKAEAAVVAVAEKREKLANGGEGMEEATVVIEHCTT*CITGDALVALPEGESV*
RIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTANHPL
LCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPGLVR
FLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSHA
TEA

For all variants, His6-SUMO3-SELENOH(2-43)-MxeGyrA was expressed in C41 cells. Bacteria were grown in 2XYT medium containing 1x NPS and 1% glucose at 37 ^{o}C until OD_{600}=0.6. Expression was induced by addition of 0.1 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and incubation at 22 ^{o}C for 18 h. Following centrifugation (10,000 rpm, 15 min, 4 ^{o}C), the pellets were stored at -20 ^{o}C until lysis.

Cell pellets (1 L expression) were resuspended in lysis buffer (150 mL of 50 mM Tris, 0.5 M NaCl, 10% glycerol, 10 mM $MgCl_2$, 1 mM PMSF, 0.25 mg/mL lysozyme, and 25 ug/mL DNAase). Cells were lysed by microfluidizer and the suspension was centrifuged at 16000 g for 10 m at 4 ^{o}C.

The soluble fraction was loaded onto a 4 mL Ni-Sepharose column and eluted with Buffer B (50 mM Tris, 0.5 M NaCl, 10% glycerol, 250 mM imidazole). Fractions were analyzed using SDS-PAGE and a dye containing no thiols. The purest fractions of protein were pooled, diluted to 50 mM NaCl concentration, and loaded onto a Resource 15Q Anion Exchange Column (4 mL). Protein was eluted with anion exchange buffer (25 mM Tris with 1 M NaCl, pH 7.5). Purest fractions were pooled and transferred into storage buffer (100 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, pH 7.2) using either dialysis or desalting column. Fractions were again analyzed on SDS-PAGE with a dye containing no thiols (Figure S3), and the cleanest fractions were pooled for thioesterification, and characterized via HPLC and ESI-MS (Figure S4).

Total yield (quantified using ε_{280} = 17210 M^{-1} cm^{-1}):

Wild-type: 57 mg/L

S43A: 14.3 mg/L

S43T: 30.1 mg/L

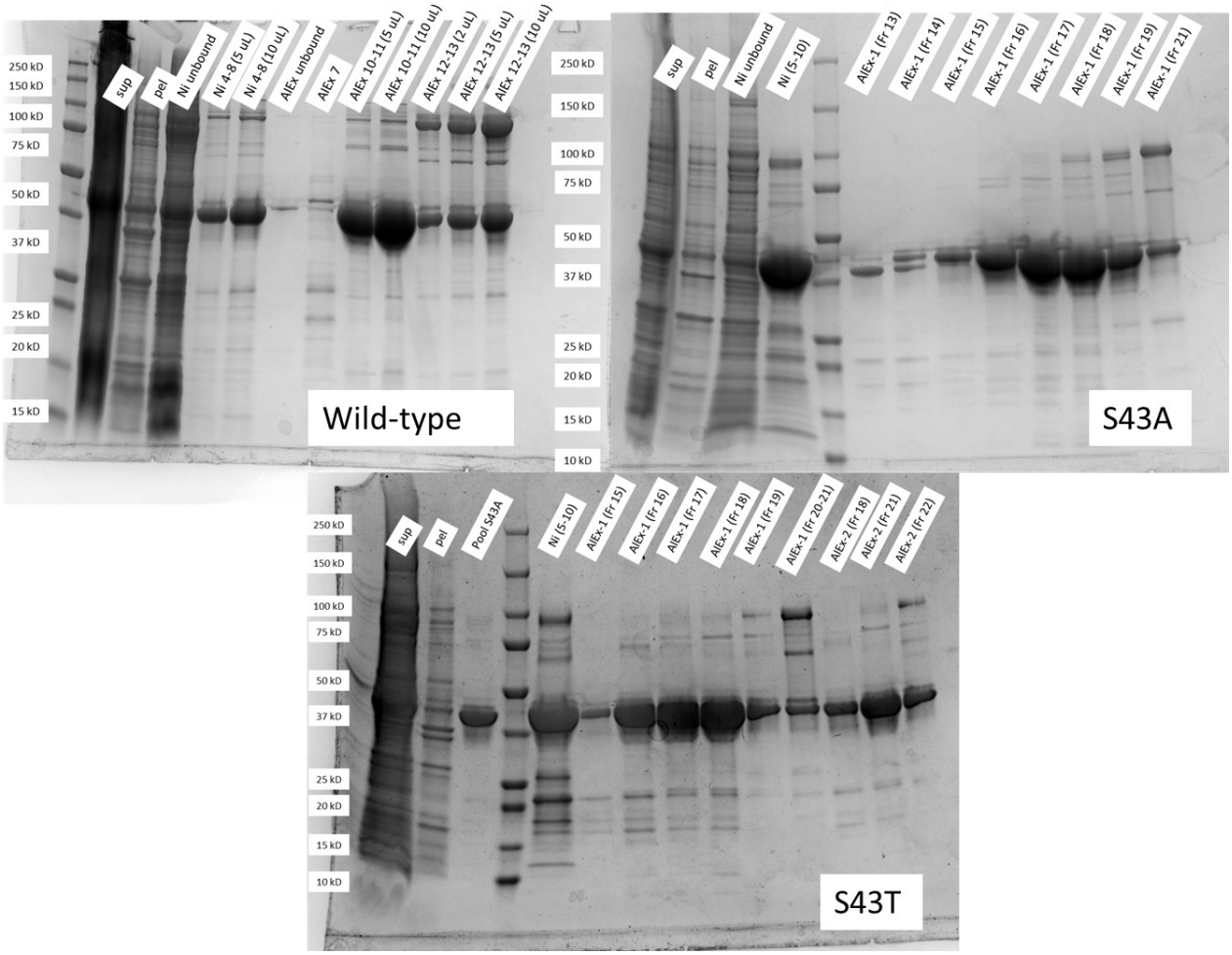

Figure S3. SDS-PAGE of fractions from Ni and AIEx purifications for SUMO-SELENOH(2-43)-MxeGyrA protein as wild-type (top left), S43A mutant (top right), and S43T mutant (bottom).

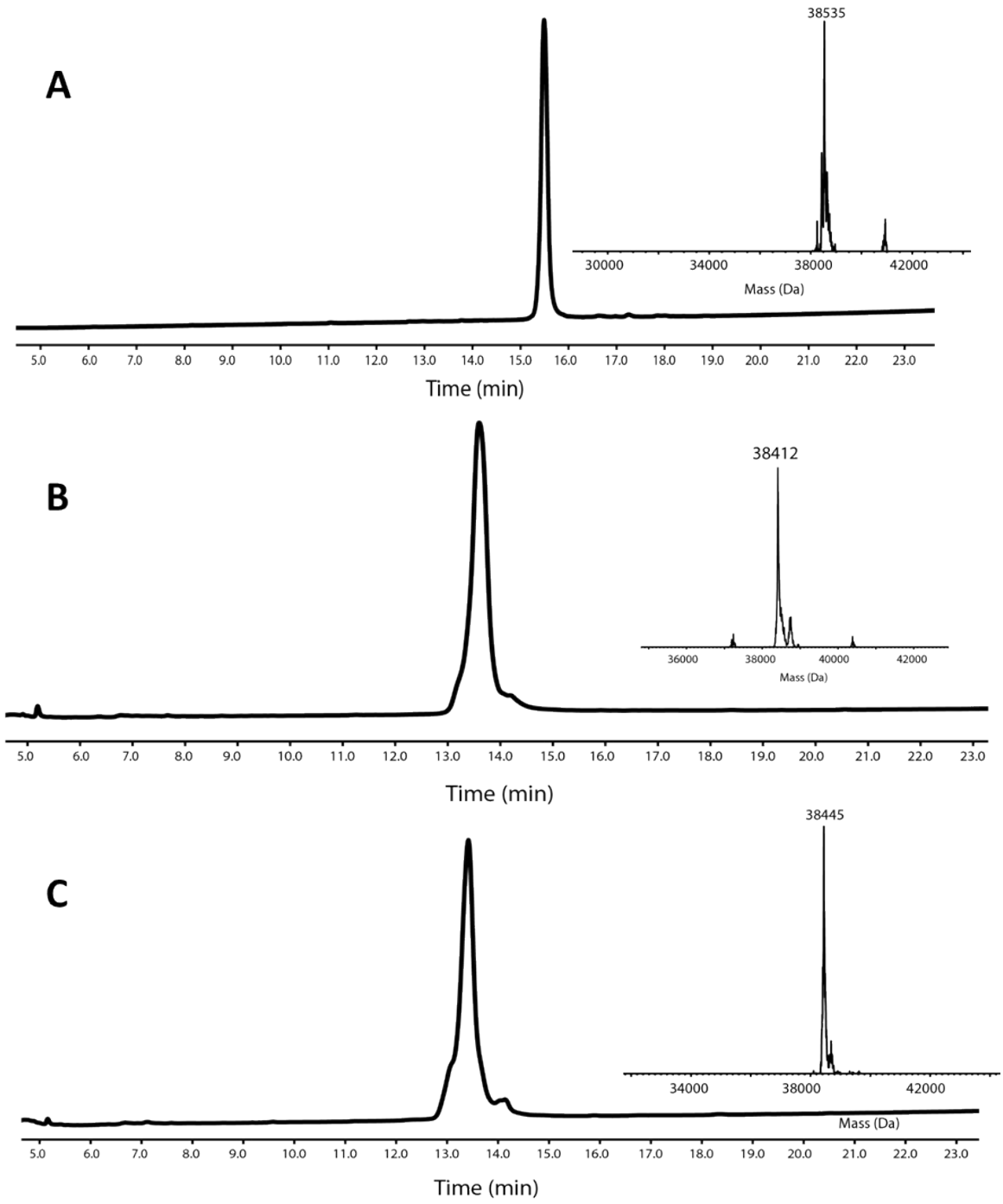

Figure S4. HPLC and ESI-MS data for **A**. Expressed SUMO-SELENOH(2-43)-MxeGyrA calc. mass 38438.2 Da, observed 38535 Da; **B**. Expressed SUMO-SELENOH(2-43)(S43A)-MxeGyrAcalc. mass 38422.2 Da, observed 38412 Da; **C**. Expressed SUMO-SELENOH(2-43)(S43T)-MxeGyrAcalc. mass 38452.2 Da, observed 38445 Da

Expressed Protein Ligation to full SELENOH

SUMO-SELENOH(2-43)-MxeGyrA Thioesterification

S43A variant: 3.65 mL of purified protein in storage buffer (10.1 mg, 2.76 mg/mL) were pooled and 5% hydrazine (v/v) was added (Figure 3).[3] After 2 hours, the reaction was purified via prep

C4 in a 30-55% gradient. Purified SUMO-SELENOH(2-43)(S43A) -$NHNH_2$was lyophilized; 3.2 mg protein was obtained (74.5% yield).

S43T variant: 4 mL of purified protein in storage buffer (20.8 mg, 5.2 mg/mL) were pooled and 5% hydrazine (v/v) was added (Figure 3).[3,4] After 2 hours, the reaction was purified via prep C4 in a 30-55% gradient. Purified SUMO-SELENOH(2-43)(S43T)-$NHNH_2$was lyophilized; 5.1 mg protein was obtained (57.7% yield).

Formation of MPAA thioester for both variants was performed at lower concentrations (<200 μM protein) with 200 equiv MPAA, 10 equiv of acetylacetone, pH 2).[2] Reaction rate was limited due to the dilution of the protein; the samples were inclubated overnight, then purified in a 25%-50% gradient on C4 semi-prep. 1.7 mg of SUMO-SELENOH(2-43)(S43A)-MPAAwas obtained (46.4% yield), and 2.6 mg of SUMO-SELENOH(2-43)(S43T)-MPAAwas obtained (76.9% yield)

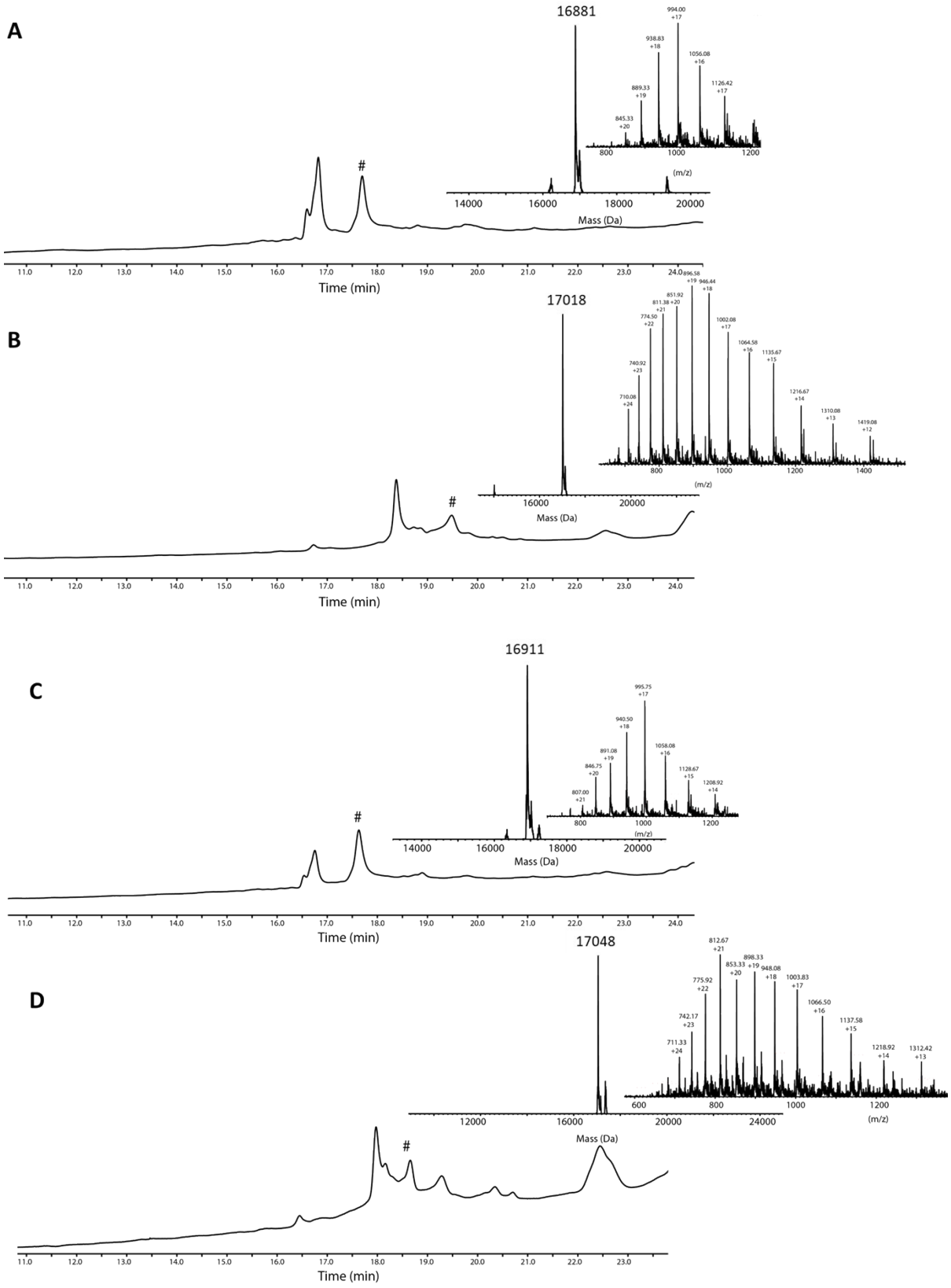

Figure S5. HPLC traces and ESI-MS spectrum for **A**. SUMO-SELENOH(2-43)(S43A)-$NHNH_2$ calc. mass 16883 Da, observed 16881 Da; **B**. SUMO-SELENOH(2-43)(S43A)-MPAA calc. mass 17019 Da, observed 17018 Da; **C**. SUMO-SELENOH(2-43)(S43T)-$NHNH_2$ calc. mass 16913 Da, observed 16911 Da; **D**. SUMO-SELENOH(2-43)(S43T)-MPAA calc. mass 17049 Da, observed 17048 Da.

SUMO-SELENOH(2-43)(S43A/T)-COSR and SELENOH(44-122) Ligation

S43A variant (Figure 4): 1.7 mg (0.10 µmol) of SUMO-SELENOH(2-43)-MPAA and 1.1 mg (0.12 µmol) SELENOH(44-122) were dissolved in 200 µL selenoligation buffer (0.5 mM concentration, 6 M GdmCl, 200 mM sodium phosphate, 100 mM sodium ascorbate, 50 mM TCEP, pH 6.5). Reaction was monitored via analytical HPLC. Reaction was purified after 18 h and 0.54 mg protein was obtained, as calculated from A_{280} (ε_{280} = 12090 $M^{-1}cm^{-1}$).

S43T variant (Figure 4): 2.6 mg (0.15 µmol) of SUMO-SELENOH(2-43)-MPAA and 1.1 mg (0.12 µmol) SELENOH(44-122) were dissolved in 200 µL selenoligation buffer. Reaction was purified after 18 h and 0.92 mg protein was obtained, as calculated from A_{280} (ε_{280} = 12090 $M^{-1}cm^{-1}$).

For both reactions: After 4 h, the N-terminal peptide-thioester had self-cyclized (presumably with Cys41, Figure 3). Further incubation overnight did not show further progress in the reaction, and so the product was purified on analytical C4 in a 25%-45% gradient.

SUMO removal from full SELENOH

0.26 mg of SUMO-SELENOH(S43T) was dissolved in 100 µL dissolving buffer (8 M Urea, 5 mM DTT, 50 mM Tris, 150 mM NaCl, 2 mM $MgCl_2$, 0.1% Triton X-100, pH 8) and diluted to 400 µL total to lower urea content (final concentration 2 M Urea, 5 mM DTT, 50 mM Tris, 150 mM NaCl, 2 mM $MgCl_2$ 0.1% Triton X-100, pH 8). ULP enzyme was added in 1:500 ratio to the protein and the reaction was incubated for 1 h at room temperature, after which a fresh portion of ULP was added, and the mixture was again shaken for 1 h at room temperature. The reaction was shaken for 1 h at 4 °C with 50 µL equilibrated Ni-sepharose beads in a dolphin-nosed tube. The beads were spun down at 3000 RPM for 3 min, and the eluent containing cut SELENOH(S43T) was removed. Portions were taken from all stop points of the reaction to show successful digest and isolation and analyzed on SDS-PAGE (Figure 5).

Expression of SELENOH(U44C)

A gene encoding SELENOH(U44C) flanked with His_6, TEV and restriction sites was ordered as a gBlock® from IDT.

gtacaTATGTATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCAAC
GACCGAAAACCTGTATTTTCAGGGCGCCCCACGTGGTCGCAAACGTAAAGCTGA
GGCGGCGGTAGTTGCCGTTGCCGAAAAGCGTGAAAAACTGGCTAACGGTGGCGA
AGGCATGGAGGAGGCCACTGTAGTGATCGAACACTGCACCTCTTGCCGTGTCTAT
GGCCGCAATGCGGCGGCACTGAGTCAGGCCTTACGCCTGGAGGCTCCTGAGCTG

CCCGTTAAGGTGAACCCGACTAAGCCTCGCCGTGGTTCTTTTGAAGTGACGCTGC
TTCGCCCAGATGGTTCTTCAGCAGAGCTTTGGACAGGTATCAAGAAAGGGCCTCC
GCGTAAGTTAAAATTTCCTGAGCCACAAGAGGTAGTAGAGGAGCTTAAGAAGTA
TCTTTCTtaaCTCGAGtatg

However, we chose to ligate only part of the gene (absent TEV and restriction sites) into a pETM11 vector containing SUMO3.

The gBlock® was amplified using the following primers that introduced overlapping ends with the desired plasmid:

5'-ccagcaacagaccggtggaGCCCCACGTGGTCGCAAACG-3'
5'-gagtgcggccgcaagctttcaAGAAAGATACTTCTTAAGCTCCTCTACTACCTCTTG-3'

Simultaneously, the pETM11 plasmid was linearized:

5'-aagcttgcggccgcactcgagc-3'
5'-tccaccggtctgttgctggaac-3'

The designed gene was inserted into the linearized plasmid via Gibson assembly and the product was transformed into competent Escherichia coli XL-10 cells; bacterial colonies were screened for the presence of the gene using PCR. Positive colonies were further verified by sequencing. Following Gibson assembly, the vector contained a gene to express the following protein:

$\underline{His_6}$
SUMO3
SELENOH(U44C) (Cys44 is bold and underlined)

$\underline{HHHHHH}$**PMSDYDIPTTENLYFQGAMGNDHINLKVAGQDGSVVQFKIKRHTPLS**
KLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG*A*
*PRGRKRKAEAAVVAVAEKREKLANGGEGMEEATVVIEHCTS**$\underline{C}$**RVYGRNAAALSQALRLEA*
PELPVKVNPTKPRRGSFEVTLLRPDGSSAELWTGIKKGPPRKLKFPEPQEVVEELKKYLS

His_6-SUMO-SELENOH(U44C) was expressed in BL21 cells. Bacteria were grown in 2XYT medium containing 1x NPS and 1% glucose at 37 ^oC until OD_{600}=0.6. Expression was induced by addition of 0.1 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and incubation at 17 ^oC for 18 h. Following centrifugation (6,000 rpm, 10 min, 4 ^oC), 500 mL pellets were stored at -20 ^oC until lysis.

Cell pellets (500 mL expression) were resuspended in lysis buffer (40 mL of 50 mM Tris, 0.5 M NaCl, 10% glycerol, 10 mM $MgCl_2$, 1 mM PMSF, 0.25 mg/mL lysozyme, and 25 ug/mL

DNAase). Cells were lysed by microfluidizer and the suspension was centrifuged at 16000 g for 10 m at 4 °C.

The soluble fraction was loaded onto a Ni-Sepharose column and eluted with Buffer B (50 mM Tris, 0.5 M NaCl, 10% glycerol, 250 mM imidazole). Fractions were analyzed using SDS-PAGE and a dye containing no thiols. The purest fractions (4-16) of protein were pooled and the amount of protein was quantified using UV absorbance. The protein was subjected to SUMO digest with 1:500 equivalents of ULP (ubiquitin-like protease) and shaking at room temperature for 1 h. Imidazole was removed using a HiTrap desalting column and the eluent was pooled and run again through a Ni-Sepharose column, this time to remove His_6-SUMO and ULP protease (which also has a His-tag). The eluted protein did not appear to be pure using SDS-PAGE (Figure S6, Channel 8). The protein was pooled and salt content was reduced to 50 mM NaCl using HiTrap desalting column. The protein was purified on cation exchange column and eluted in two pulses (0.5 M NaCl and 1 M NaCl) (Figure S6, Channels 11-12). The pulses were stored separately.

Total yield (quantified using $\varepsilon_{280} = 8250\ M^{-1}\ cm^{-1}$):

Pool 2 (0.5 M NaCl elution): 3.1 mg (23.0 uM)

Pool 3 (1 M NaCl elution): 4.4 mg (66.2 uM)

Following isolation, the proteins were dialyzed in their separate pools into a buffer containing 150 mM NaCl, 25 mM sodium phosphate, pH 7.5.

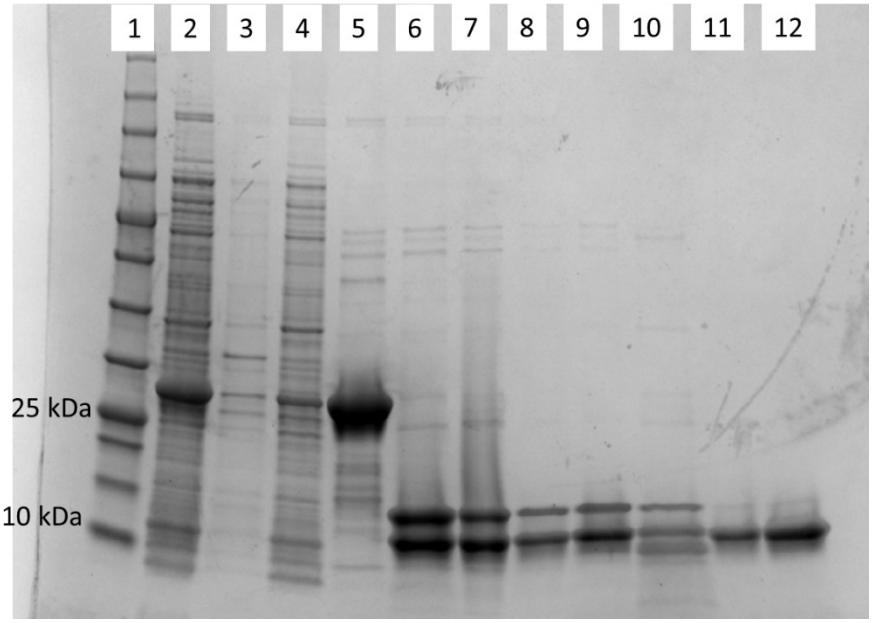

Figure S6. SDS-PAGE for purification of SELENOH(U44C). 1 = ladder, 2 = sup, 3 = pellet, 4 – Ni unbound, 5 = first Ni column pooled fractions 4-16, 6 = after SUMO cut, 7 = after desalting, 8 = negative Ni fr 4, 9 = after desalting, 10 = CIEX pulse 1, 11 = CIEX pulse 2, 12 = CIEX pulse 3

Structural determination by Gel Filtration and Circular Dichroism (CD)

One possible explanation for the separation of Pool 2 and Pool 3 proteins, which have the same MW, at different NaCl concentrations, is different oligomerization states. To this end, we ran analytical gel filtration/size-exclusion chromatography samples (Superdex S75 column) to determine the size of the purified proteins. Pool 2 and 3 were eluted in the same volume, indicating that their oligomerizations states were identical, and this was not the source for their different elutions in cation exchange (Figure S7).

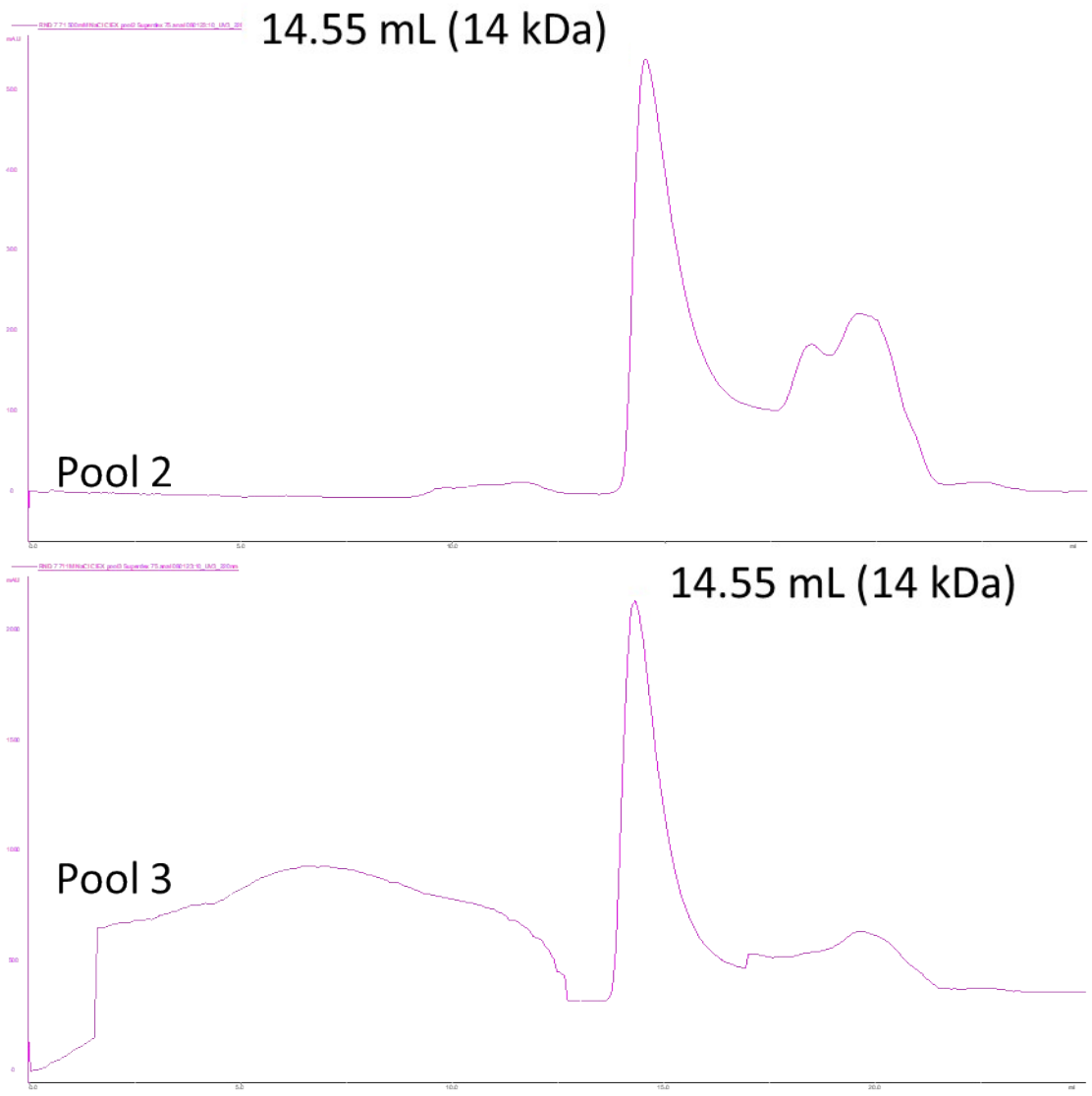

Figure S7. Gel Filtration/Size-exclusion chromatogram shows that the proteins of Pool 2 and Pool 3 are identical oligomerization states.

To further explore the differences in these proteins, their masses were determined via HR-MS to see if their oxidation states differed (Figure S8B). Both proteins had identical masses of reduced protein. Finally, the secondary structures of the folded proteins were characterized using CD spectroscopy, in the far-UV range (190 nm to 250 nm). CD measurement was

performed at 25 °C in BLI buffer (150 mM NaCl, 25 mM sodium phosphate, pH 7.5) in a 0.1 cm cuvette using MOS-500 spectropolarimeter (BioLogic). Protein concentrations were 40.4 μM for Pool 3 (reduced disulfide) and 16.6 μM for Pool 2 (reduced disulfide). The spectra were recorded by averaging 5 wavelength scans in 1 nm steps, with a signal averaging time of 0.05 s and a bandwidth of 2 nm (Figure S8A). The resulting curves showed that Pool 2 contained misfolded protein with a minimum in the disordered region (200 nm), while Pool 3 protein was correctly folded with an α-helix (minima at 208 nm and 222 nm). With this information in hand, we concluded that the different elutions during cation exchange were due to different folding states of the protein, and only used Pool 3 (folded) proteins during DNA binding experiments.

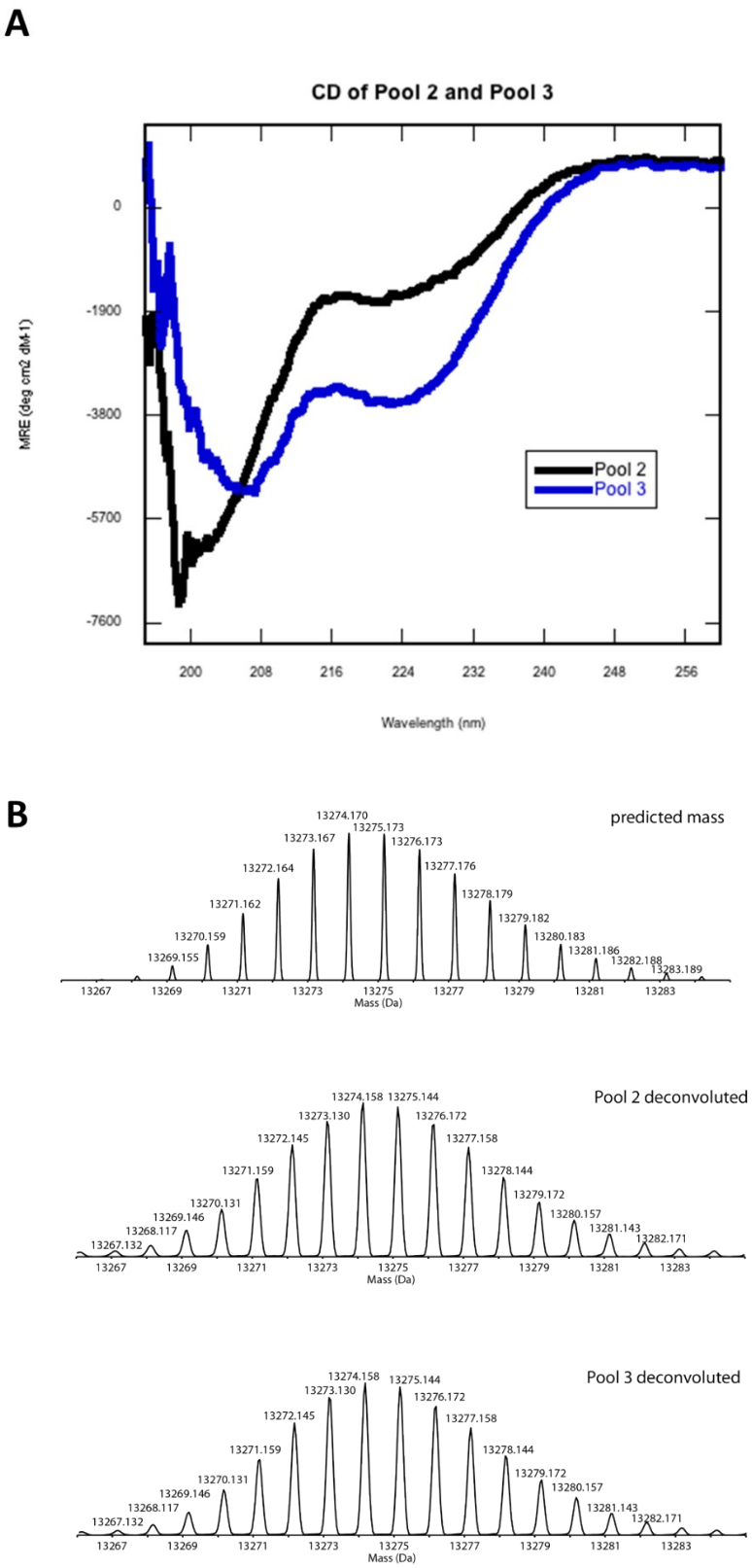

Figure S8. **A**. Structural characterization shows that Pool 3 protein has a more defined structure than Pool 2. **B.** Both Pools 2 and 3 have identical mass as determined by HRMS.

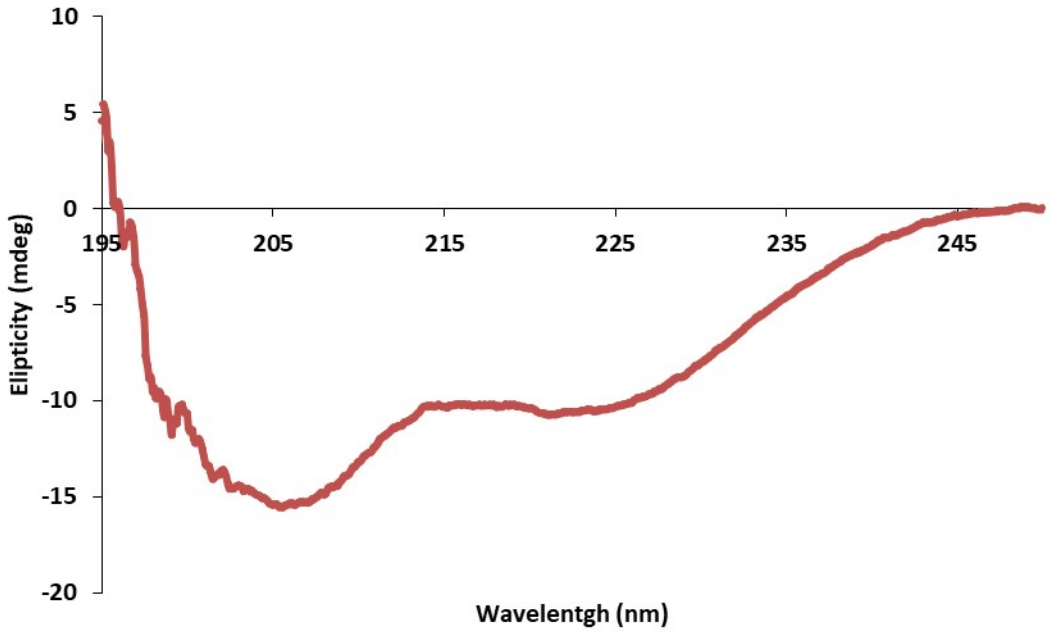

Figure S9. CD analysis for isolated folded SELENOH(U44C).

Protein-DNA affinity binding: Biolayer Interferometery (BLI)

DNA Preparation

To prepare dsDNA, each ssDNA was dissolved to 100 µM concentration in ddw. 10 µL of each component (top and bottom strand) was combined in a PCR tube (20 µL total in a tube). The solution was heated to 95 °C to denature any structures and then slowly cooled over 30 min to room temperature, annealing as a dsDNA. The resulting 100 µM solutions of dsDNA were stored at -20 °C and diluted as needed into BLI buffer (150 mM NaCl, 25 mM sodium phosphate, pH 7.5)

BLI Assay Protocol

Gator Flex SA Probes were regenerated with the Flex SA Reagent Kit prior to each usage. Probes were dipped in BLI buffer (150 mM NaCl, 25 mM sodium phosphate, pH 7.5) and then into a solution of 3 µg/200 µL of relevant dsDNA as a loading step. The sensors were returned to buffer to measure a baseline, and then dipped into increasing concentrations of protein (0.2, 0.3, 0.5, 1, and 2 µM) as an association step, with intermediate washes in buffer for dissociation steps. K_D's were calculated using steady-state analysis (explained below).

Steady State Analysis

Signal shift (response) of the probe after reaching steady-state in binding was plotted against concentration. A curve was fit according to the following equation:

$$Response = \frac{R_{max} * conc}{K_D + conc}$$

DNA type	K_D (nM)	Error
HSE	1200	±400
STRE	1000	±200
scrambled	700	±100
Scrambled AT-poor	1700	±800

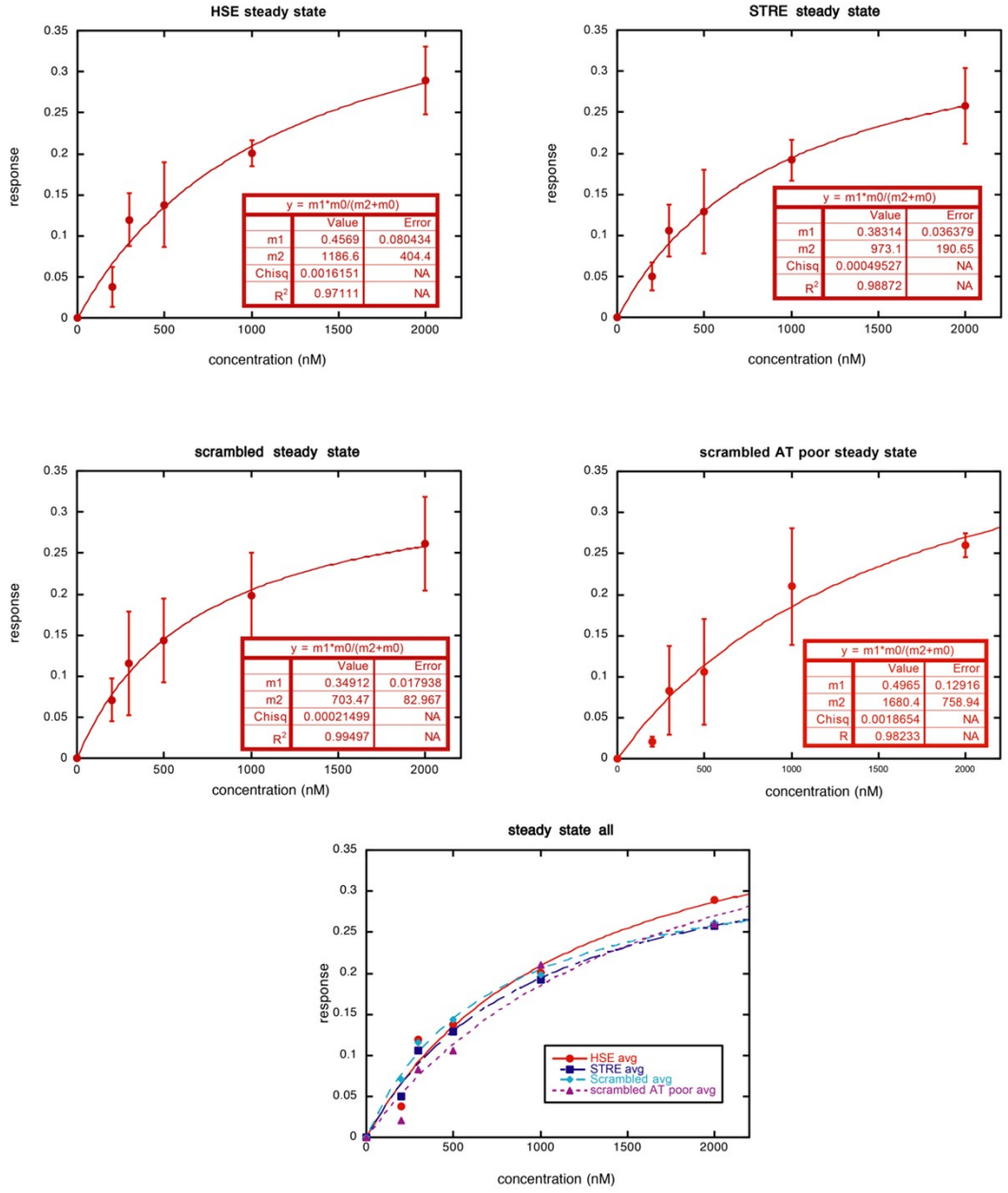

Figure S10. Signal shift of probe at equilibrium was plotted against concentration and fitted to give K_D.

(1) Dardashti, R. N.; Kumar, S.; Sternisha, S. M.; Reddy, P. S.; Miller, B. G.; Metanis, N. Selenolysine: A New Tool for Traceless Isopeptide Bond Formation. *Chem. – Eur. J.* **2020**, *26* (22), 4952–4957. https://doi.org/10.1002/chem.202000310.
(2) Flood, D. T.; Hintzen, J. C. J.; Bird, M. J.; Cistrone, P. A.; Chen, J. S.; Dawson, P. E. Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for Use in Native Chemical Ligation Zuschriften. *Angew Chem Int Ed Engl* **2018**, *130*, 11808–11813. https://doi.org/10.1002/ange.201805191.
(3) Thom, J.; Anderson, D.; McGregor, J.; Cotton, G. Recombinant Protein Hydrazides: Application to Site-Specific Protein PEGylation. *Bioconjug. Chem.* **2011**, *22* (6), 1017–1020. https://doi.org/10.1021/bc2001374.
(4) Kulkarni, S. S.; Watson, E. E.; Maxwell, J. W. C.; Niederacher, G.; Johansen-Leete, J.; Huhmann, S.; Mukherjee, S.; Norman, A. R.; Kriegesmann, J.; Becker, C. F. W.; Payne, R. J. Expressed Protein Selenoester Ligation. *Angew. Chem.* **2022**, *134* (20), e202200163. https://doi.org/10.1002/ange.202200163.