# Site-specific incorporation of selenocysteine by genetic encoding as a photocaged unnatural amino acid

Adarshi P. Welegedara,[1] Luke A. Adams,[2] Thomas Huber,[1] Bim Graham,[2] Gottfried Otting[1*]

[1] Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[2] Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia

[*]corresponding author

phone: +61 2 61256507

fax: +61 2 61250750

e-mail: gottfried.otting@anu.edu.au

## ABSTRACT

Selenocysteine (Sec) is a naturally occurring amino acid that is also referred to as the 21[st] amino acid. Site-specific incorporation of Sec into proteins is attractive, because the reactivity of a selenol group exceeds that of a thiol group and thus allows site-specific protein modifications. It is incorporated into proteins by an unusual enzymatic mechanism which, in *E. coli* and other organisms, involves the recognition of a selenocysteine insertion sequence (SECIS) in the mRNA of the target protein. Reengineering of the natural machinery for Sec incorporation at arbitrary sites independent of SECIS elements, however, is challenging. Here we demonstrate an alternative route, whereby a photocaged selenocysteine (PSc) is incorporated as an unnatural amino acid in response to an amber stop codon, using a mutant *Methanosarcina mazei* pyrrolysyl-tRNA synthetase, *Mm* PCC2RS, and its cognate $tRNA_{CUA}$. Following decaging by UV irradiation, proteins synthesized with PSc are readily tagged, e.g. with NMR probes to study ligand binding by NMR spectroscopy. The approach provides a facile route for genetically encoded Sec incorporation. It allows the production of pure selenoproteins and the Sec residue enables site-specific covalent protein modification with reagents that would usually react first with naturally occurring cysteine residues. The much greater reactivity of Sec residues allows their selective alkylation in the presence of highly solvent-exposed cysteine residues.

## INTRODUCTION

Site-specific protein modification is a fundamental tool for a myriad of applications in protein chemistry, including protein immobilization on surfaces, enhancement of protein solubility by PEGylation, production of therapeutic conjugates with drugs, mimicking of natural post-translational modifications, and introduction of tags for structural studies by fluorescence resonance energy transfer (FRET), electron paramagnetic resonance (EPR) spectroscopy, or nuclear magnetic resonance (NMR) spectroscopy.[1–8] Cysteine is the most convenient target for protein modification as it is the most nucleophilic and oxidation-sensitive among the 20 canonical amino acids. Thus site-directed modifications of proteins are often carried out on cysteine residues with the formation of disulfide or thioether bonds. To be site-selective, however, this approach requires that only a single cysteine residue is amenable to the chemical reaction, which is difficult when the protein of interest contains other cysteine residues that are essential for the structure or function of the protein. In principle, this problem can be solved by genetic encoding of photocaged cysteine (PCC) as an unnatural amino acid (UAA),[9] which enables a protocol involving the chemical protection of the native cysteine residues, decaging of the PCC residue and chemical modification and deprotection of the native cysteine residues.[10] Charging of the cognate amber suppressor $tRNA_{CUA}$ with the PCC amino acid is elegantly achieved by a mutant *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase (*Mb* PylRS), named PCC2RS,[9] but the number of steps involved in selective cysteine labeling renders the overall process unattractive. A more straightforward alternative is to incorporate a Sec residue, which is chemically more reactive than cysteine, obviating the need for protection of any native cysteine residues. Sec is chemically similar to cysteine but has a significantly lower p$K_a$ value (5.2 versus 8.5).[11–13] This can be exploited for selective modifications of Sec residues in the presence of native cysteine residues at slightly acidic pH values.[14–17] A recent, very attractive application harnesses the exceptional reactivity of Sec residues for expressed protein-ligation.[17–19]

Site-specific incorporation of Sec is challenging. The natural mechanism employs a highly specialized translation system, relying on a specific tRNA ($tRNA^{Sec}$), selenophosphate synthase (SelD), Sec synthase (SelA), a special EF-Tu-independent elongation factor (SelB), and a conserved mRNA sequence (selenocysteine insertion sequence or SECIS) in the vicinity of the target UGA codon. The requisite mRNA stem-loop structure comprised by the SECIS element restricts the sequence context where Sec residues can be positioned.[11,13] In an *E. coli* strain devoid of the release factor 1 (RF1), the SECIS/SelB system has also successfully been harnessed for the synthesis of pure selenoproteins using a UAG (amber) codon instead of the usual UGA codon but, in the absence of the SECIS element, this approach results in full-length protein due to misincorporation of natural amino acids.[20] Site-selective Sec incorporation at an amber stop codon independent of SelB and the SECIS element has also been

demonstrated, using a $tRNA^{Sec}$ engineered for EF-Tu–mediated Sec incorporation.[21–24] At present, this approach is limited to low yields of selenoproteins (<0.5 mg/L cell culture), and further optimization of the engineered $tRNA^{Sec}$ is difficult to conceive as it needs to be recognized both by SelA and EF-Tu.[25] Poor expression yields thus have prevented the use of this approach for subsequent applications that require larger protein quantities, such as NMR spectroscopy or expressed protein ligation. The problem of low yield is circumvented by a new system for EF-Tu mediated selenocysteine incorporation at amber codons, which uses *Aeromonas salmonicida* selenocysteine synthase (SelA) to convert Ser-allo-tRNA into Sec-allo-tRNA which is efficiently recognized by EF-Tu.[26] The system can be used to obtain selenoproteins in high yields, but it is prone to incorporation of serine in place of selenocysteine due to incomplete conversion of Ser to Sec by SelA. This makes the strategy critically dependent on optimization of the relative expression levels of allo-tRNA and SelA as well as the expression temperatures, which are different for different proteins.[26]

The wide interest in Sec incorporation has triggered the development of a range of alternative strategies for the production of selenoproteins, in particular focusing on expression in *E. coli*, which is the most readily used organism for protein production. In a straightforward way, cysteine residues can globally be replaced by Sec by misloading the cysteinyl-tRNA with Sec in a growth medium supplemented with Sec, but incorporation of cysteine cannot be completely avoided in this approach and site-selectivity is achieved only if the native protein sequence contains only a single cysteine residue,[27] or a Sec-containing fragment is post-translationally ligated with a separately produced polypeptide.[17] As an alternative, it has been shown that a photocaged Sec, 4,5-dimethoxy-2-nitrobenzyl-Sec (DMNB-Sec), can be incorporated as a UAA in yeast by genetic encoding based on an orthogonal LeuRS/$tRNA_{CUA}$ pair, with subsequent decaging providing a straightforward route to a site-specific Sec residue.[16] This system, however, requires high concentrations of DMNB-Sec (3 mM) to suppress erroneous incorporation of leucine and isoleucine,[16] and cannot be used in *E. coli*, which limits its applications, e.g., making it difficult to produce isotope-labeled proteins. The misincorporation of leucine and isoleucine appears to be intrinsic, as it has also been observed in the corresponding system established for DMNB-Ser.[28] A cost-effective general method is yet to be established that works in *E. coli* and produces selenoproteins in sufficient yield and with reproducible purity (>90%) to render Sec incorporation attractive for site-specific chemical modifications for studies by NMR spectroscopy.

The present study is based on the hypothesis that the pyrrolysine system developed for the incorporation of photocaged cysteine, PCC, at amber stop codons could be adapted for the incorporation of the seleno-analogue of PCC (Figure 1A). Incorporation of selenocysteine as a photocaged unnatural amino acid (PSc) not only provides the required specificity, but also minimizes losses arising from the high reactivity of free selenol groups, as decaging of the PSc residue can be performed on the purified

protein and the product can be subjected to further reactions immediately and under controlled conditions. In our hands, the wild-type pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* (*Mm*) had performed well in different experiments for the incorporation of Boc-lysine at amber stop codons. Therefore, we transferred the mutations reported for the *Mb* PCC2RS enzyme to the corresponding synthetase from *Methanosarcina mazei* to produce *Mm* PCC2RS, and tested the incorporation of PSc in two different proteins. The strategy involves the expression of the target protein in a two-plasmid system in *E. coli*, with the suppressor tRNA and aminoacyl-tRNA synthetase on one vector and the target protein on the other, followed by protein purification, decaging and reaction of the selenocysteine with an alkylating agent (Figure 1B). We demonstrate that this strategy yields selenocysteine proteins with very high purity and in sufficient yield for NMR studies. Finally, we demonstrate that this approach allows quantitative and selective covalent protein modification with reagents, which would usually react with solvent-exposed cysteine residues. Due to the much greater reactivity of selenocysteine, competing reactions of solvent-exposed cysteine residues were undetectable.

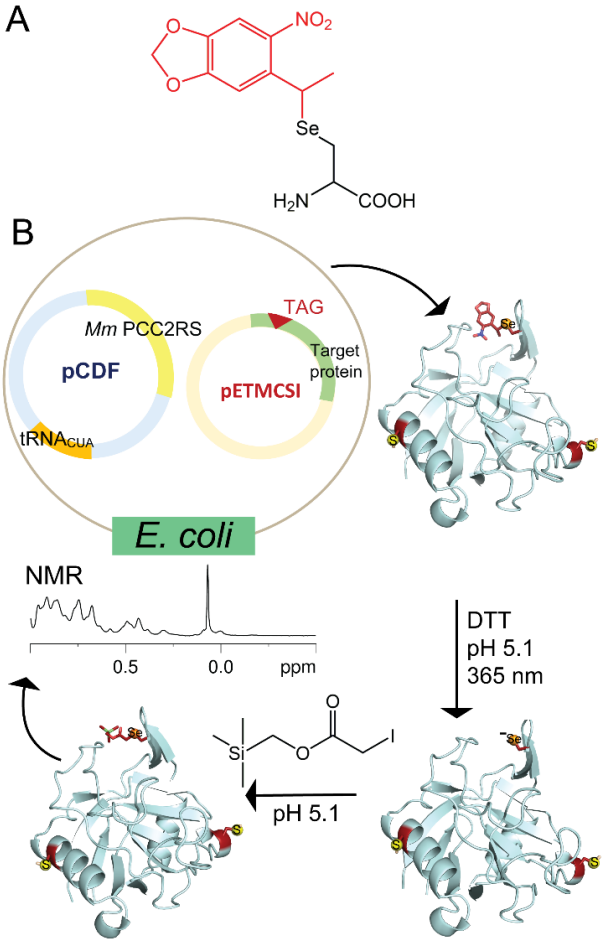



**Figure 1.** Genetically encoded incorporation of PSc into proteins in *E. coli*. (A) Chemical structure of photocaged selenocysteine (PSc). The red moiety is the caging group that is released upon UV irradiation. (B) Strategy for the production and site-specific alkylation of seleno-proteins followed in the present work.

## RESULTS AND DISCUSSION

### PSC incorporation into proteins

The gene of the *Mm* PCC2RS enzyme was cloned into a pCDF vector that also contained the gene of the cognate $tRNA_{CUA}$.[29] The amino acid residues lining the substrate binding pockets of *Mm*PylRS and *Mb*PylRS are highly conserved between both enzymes, so that the specificity-changing mutations in *Mb* PCC2RS were easily transferred to make *Mm* PCC2RS. Denoting the unnatural amino acid PSc or Sec as "U", the mutant H147U of *E. coli* peptidyl-prolyl *cis-trans* isomerase B (PpiB H147U) and the mutant V36U of the Zika virus NS2B-NS3 protease (ZiPro V36U) were selected as model proteins to demonstrate the incorporation of PSc by the *Mm* PCC2RS/$tRNA_{CUA}$ pair. Both wild-type proteins contain cysteine residues and the cysteine residues in ZiPro are fully solvent exposed (Figure S1). Production of full-length PpiB H147U and ZiPro V36U in *E. coli* BL21(DE3) depended on the addition of PSc (Figure 2B). As the amber stop codon is recognized both by $tRNA_{CUA}$ and RF1, truncated proteins are an unavoidable side product. To obtain pure full-length proteins, we expressed them with a C-terminal $His_6$-tag and purified using a Ni-NTA column (Figure 2B).

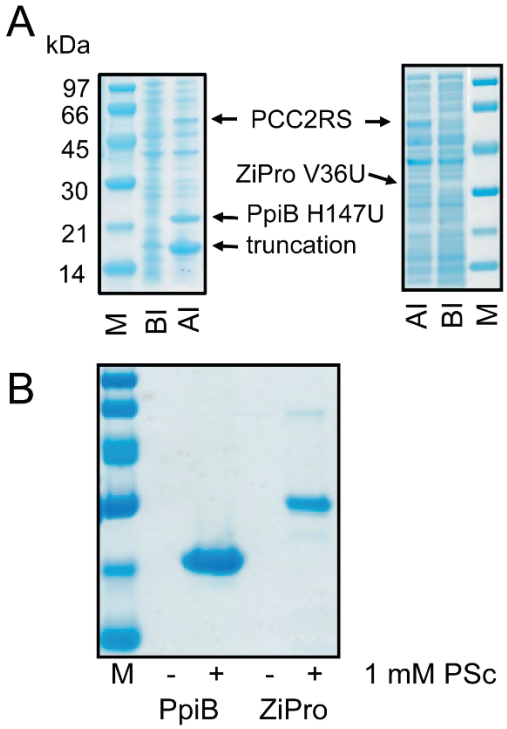



**Figure 2.** SDS-PAGE analysis of the expression and purification of PpiB H147U and ZiPro V36U, where U stands for PSc. (A) Expression of PpiB H147U (left panel) and ZiPro V36U (right panel). M: marker lane; BI and AI: before and after induction with IPTG, showing the total cellular protein of *E. coli*. The bands of full-length protein, truncated protein and *Mm* PCC2RS are marked. Truncated ZiPro (9.5 kDa) is not visible in the gel. (B) PpiB H147U and ZiPro V36U expressed in the presence and absence of PSc, after purification on a Ni-NTA column.

High-resolution mass spectrometry of full-length proteins confirmed the incorporation of PSc (Figure 3A and B) and in-gel tryptic digestion followed by MS/MS peptide mapping confirmed PSc incorporation at the amber codon (Figure S2). The yields of PpiB H147U and ZiPro V36U were, respectively, about 4.5 mg and 1.5 mg per liter cell culture, using PSc concentrations of only 1 mM in the expression medium.

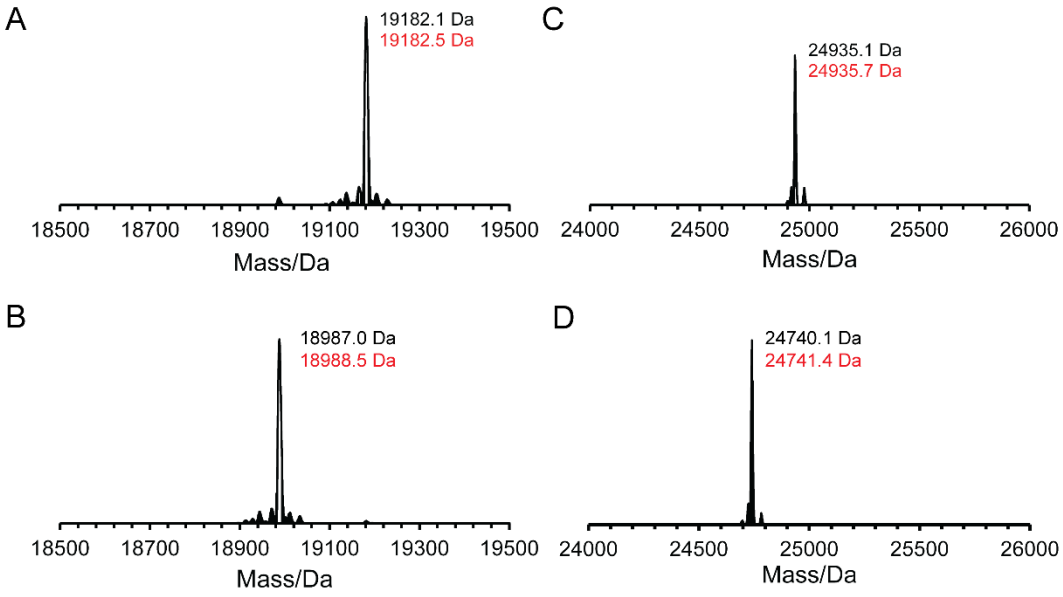



**Figure 3.** Mass spectra of intact proteins containing PSc and, after decaging, Sec. Observed and expected masses are indicated in black and red, respectively. The expected masses were calculated based on the masses observed for wild-type PpiB and ZiPro (Figure S3). (A) PpiB H147PSc. (B) ZiPro V36PSc. (C) PpiB H147Sec. (D) ZiPro V36Sec.

**Photolysis and tagging**

High-resolution mass spectra confirmed the complete conversion of PSc into Sec upon illumination of the protein solution at 365 nm at 4 °C (Figure 3C and D). Importantly, dehydroalanine or serine by-products were not observed after photolysis, even though these undesired species have often been observed as a consequence of selenoprotein oxidation followed by *syn*-β-elimination of selenenic acid.[16,30]

To demonstrate the use of Sec-specific labeling, we alkylated the Sec residue with the trimethylsilyl (TMS) tags TMS-1 and TMS-2 (Figure 4A). These tags form thioether bonds with solvent exposed cysteine residues and are not suitable for tagging specific cysteine residues in the presence of other thiols.[31,32] Reaction of these tags with the Sec residue in PpiB H147U and ZiPro V36U, however, proceeded quantitatively within 5 minutes at room temperature and pH 5.1. Complete ligation was confirmed by high-resolution mass spectrometry (Figure 4B–D). Most importantly, the solvent-exposed

free cysteine residues in ZiPro V36PSc at this pH did not react in this short time as thiol groups are less reactive.[31,32] To probe the reactivity of cysteine residues towards alkylation with TMS-1, we first applied the same reaction conditions to wild-type PpiB, wild-type ZiPro, and ZiPro V36U prior to decaging. As mass spectra revealed no change in mass, we prolonged the reaction time for the wild-type proteins to 12 h and increased the pH to 7.5. Under these conditions, the solvent exposed cysteine residues of ZiPro were quantitatively alkylated, whereas the buried cysteine residues in PpiB remained intact (Figure S3). To probe the reactivity of cysteine residues at the incorporation sites of Sec residues in PpiB H147U and ZiPro V36U, we prepared the mutants PpiB H147C and ZiPro V36C/C80S/C143S and reacted the proteins with TMS-1 overnight at pH 5.1. Even after 16 h reaction time at room temperature, mass spectra and NMR spectra revealed only a limited degree of reaction in PpiB H147C (Figures S4 and S5) and no evidence of reaction for the ZiPro mutant (Figure S4). This result highlights the much greater reactivity of Sec over Cys even at sites, where alkylation is relatively disfavoured. The pH is very important, however, as 5 minutes reaction of decaged ZiPro V36U with TMS-1 at pH 7.5 and room temperature resulted in an additional $^{1}$H NMR signal of the TMS group at about 15% intensity of the main peak.

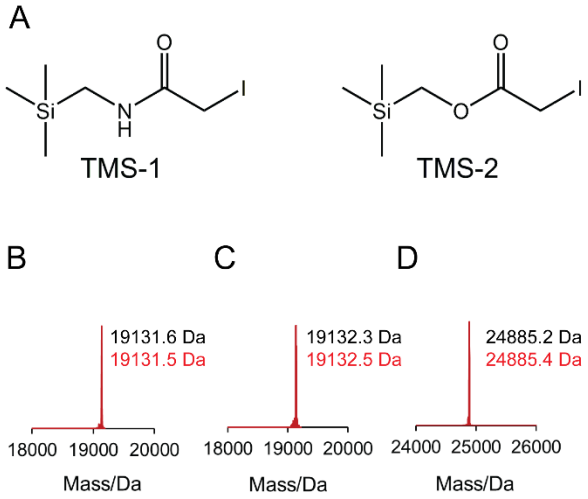



**Figure 4.** Selective Sec labeling with alkylating TMS tags. Observed and expected masses are indicated in black and red, respectively. The expected mass increases with TMS-1 and TMS-2 are 143 and 144 Da, respectively. Compare with Figure 3C and D for the masses of the untagged selenoproteins. (A) Structures of TMS-1 and TMS-2. (B) Mass spectrum of PpiB H147U labeled with TMS-1. (C) Mass spectrum of PpiB H147U labeled with TMS-2. (D) Mass spectrum of ZiPro V36U labeled with TMS-2.

### 1D $^{1}$H NMR and ligand binding studies

The TMS groups of the TMS-1 and TMS-2 tags produce a narrow, intense peak in 1D $^{1}$H NMR spectra near 0 ppm (Figure 5A and B).[31,32] In this spectral region, proteins display few $^{1}$H NMR resonances, making it easy to identify the TMS signal without assigning the NMR spectrum of the entire protein.[32] As shown previously, this can be exploited in ligand-binding studies to measure intermolecular NOEs[31] and

ligand binding affinities.[32] The TMS-1 tag is an amide whereas the TMS-2 tag is an ester. Figure 5A shows that the TMS-1 tag produces a small additional signal which we attribute to *cis-trans* isomerization of the amide bond, as it is absent from the spectrum with the TMS-2 tag.

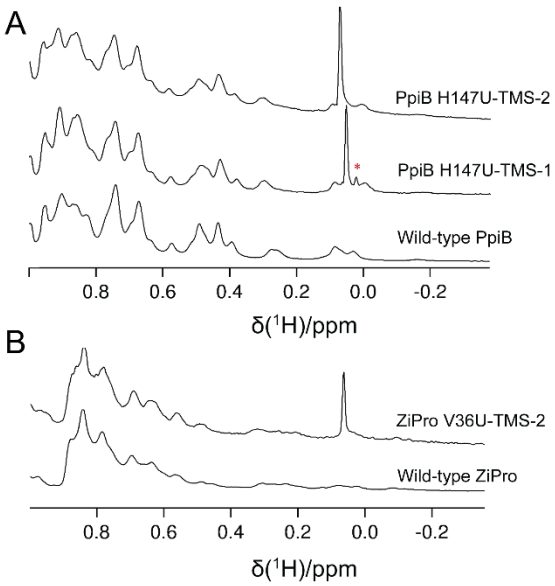



**Figure 5.** 1D $^{1}$H NMR spectra of PpiB and ZiPro labeled with TMS tags. All spectra were measured at 25 °C. (A) $^{1}$H NMR spectrum of wild-type PpiB and PpiB H147U tagged with TMS-1 or TMS-2. Protein concentrations ranged between 10 and 160 μM in 50 mM HEPES, pH 8.0. A red star marks the position of a minor species observed with the TMS-1 but not the TMS-2 tag, which is attributed to a *cis*-amide bond of the tag. (B) 1D $^{1}$H NMR of wild-type ZiPro and ZiPro V36U tagged with TMS-2. Protein concentrations were 50 and 14 μM, respectively, in 20 mM MES, pH 6.8, 150 mM NaCl and 1 mM DTT.

To demonstrate the use of the TMS signals for ligand binding studies, we titrated TMS-1-tagged PpiB H147U with the substrate analogue succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (suc-AAPF-pNA).[33,34] The Sec residue is near the substrate binding site (Figure 6A).[33] The chemical shift of the TMS signal changed gradually with increasing ligand concentration, indicating fast chemical exchange[35] with a dissociation constant $K_d$ of 12 mM (Figure 6). Based on 2D NMR spectra, a weak binding affinity of suc-AAPF-pNA has been reported previously,[34] but the $K_d$ value was not determined.

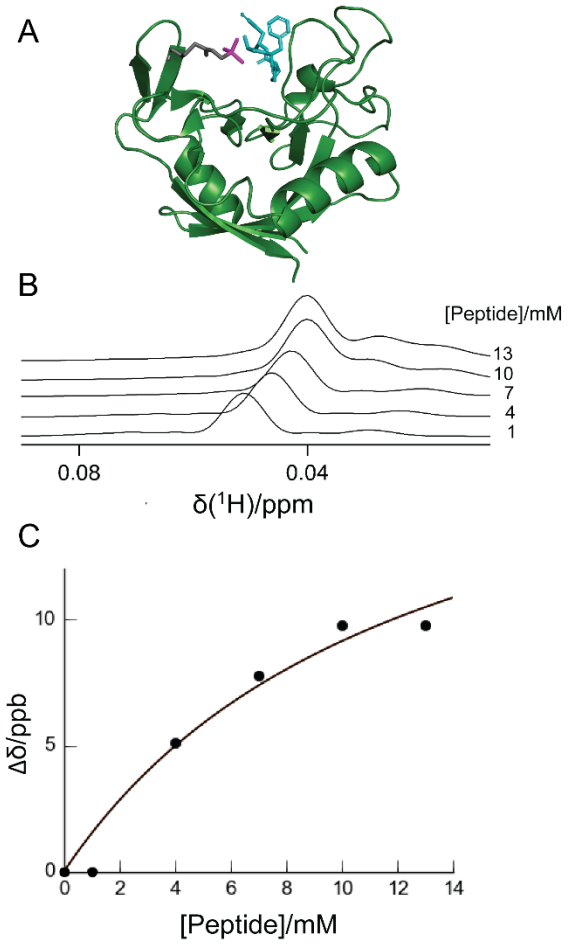



**Figure 6.** Peptide binding study with PpiB H147U tagged with TMS-1. (A) Substrate binding site of PpiB with a Sec-TMS-1 residue modeled at position 147. The protein is displayed in a ribbon representation (PDB ID: 1LOP),[33] showing the suc-AAPF-pNA peptide in cyan and the TMS group in magenta. (B) 1D $^{1}$H NMR spectra of the TMS signal of PpiB H147U tagged with TMS-1, illustrating the chemical shifts observed during titration with suc-AAPF-pNA peptide. The peptide concentrations are indicated. The spectra were recorded with a 10 µM protein solution in 50 mM HEPES buffer, pH 8.0, at 25 °C using a Bruker 800 MHz NMR spectrometer. (C) Change in chemical shift of the TMS group in response to increasing peptide concentration. The binding curve corresponds to the best fit using equation 1.

In ZiPro V36U, the TMS-2 tag is attached close to the catalytic center formed by Ser135, His51, and Asp75 (Figure 7A).[36] The inhibitor in Figure 7B was previously shown to bind to the triple mutant V36C/C80S/C143S of ZiPro tagged with the TMS-2 tag with a $K_d$ value of 64 ± 27 µM.[32] The exchange between bound and free inhibitor was found to be slow, as determined by the appearance of a new TMS peak with increasing inhibitor concentration.[32,35] Figure 7C shows the same spectral appearance for ZiPro V36U tagged with the TMS-2 tag, confirming that the presence of the wild-type residues Cys80 and Cys143 did not affect the tagging. For the concentration of ZiPro V36U protein used in the inhibitor

binding experiments (14 μM) and the $K_d$ value (64 μM), about 20% of the protein is expected to be complexed with inhibitor at 1.5-fold excess of inhibitor, in agreement with the experimental result (Figure 7C).

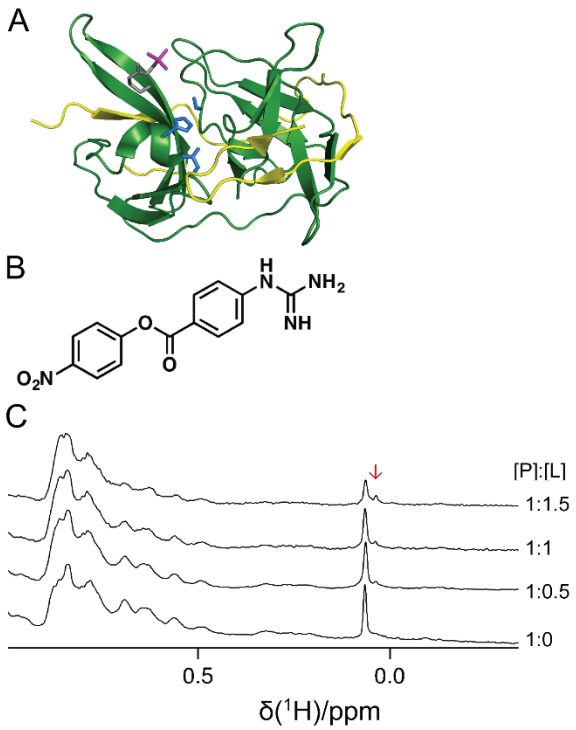



**Figure 7.** Inhibitor binding studies with ZiPro V36U-His6 tagged with TMS-2. (A) Ribbon representation of ZiPro (PDB ID: 5LC0),[36] showing the NS3 protease domain in green and the NS2B co-factor in yellow. The TMS group of the TMS-2 tag attached to the Sec residue at position 36 is shown in magenta and side chains of the catalytic triad in blue. (B) Chemical structure of the ZiPro inhibitor. (C) 1D $^1$H NMR spectra of a 14 μM solution of ZiPro V36U tagged with the TMS-2 tag in NMR buffer (20 mM MES, pH 6.8, 150 mM NaCl, 1 mM DTT). The TMS signal is at about 0.07 ppm. Protein-to-ligand titration ratios are indicated. An arrow marks the new peak attributed to the complex, which appears upon titration with inhibitor. All spectra were recorded at 25 °C with 14 μM protein solutions.

## CONCLUSION

The site-selective Sec incorporation strategy presented here overcomes a long-standing stumbling block for accessing pure selenoproteins in good yield in *E. coli* and, in principle, with free choice of the insertion sites of the Sec residues. Although the bulkiness of a PSc residue could interfere with protein folding, if it is placed in a buried site, such interference is less likely to occur for solvent exposed sites which are of prime interest for chemical protein modifications. *Mm* PylRS systems have been shown to be also orthogonal in yeast and mammalian expression systems.[37–40] The present approach thus establishes a

valuable platform for the myriad of FRET, EPR, and NMR tags that have been developed for tagging single cysteine residues[14–17] and invites many other applications,[11] including the use of Sec for expressed protein ligation.[17–19]

## EXPERIMENTAL SECTION

### Plasmid construction

The *Methanosarcina mazei* analogue of *Methanosarcina barkeri* PCC2RS[9] (*Mm* N346Q/C348A/V401M) was cloned into *Nco*I and *Sac*I restriction sites of the modular four-component vector pCDF Duet_4-components_lac(P1)[29] to obtain the plasmid pCDF-*Mm* PCC2RS/$tRNA_{CUA}$. In addition, the third nucleotide of the $tRNA_{CUA}$ (G) was replaced by A. (The vector also contained *Methanococcus jannaschi* $tRNA_{UCCU}$ which is not recognized by *Mm* PCC2RS.)[29] All PpiB and ZiPro mutants contained a C-terminal $His_6$ tag and were cloned into the *Nde*I and *Eco*RI sites of the T7 expression vector pETMSCI.[41] The wild-type ZiPro construct was described previously.[42] It contained residues 48*–95* of NS2B (where the star indicates residues in NS2B) linked to the NS3 protease domain (residues 1–170) by a $Gly_4SerGly_4$ linker. It also contained the mutations R95*A, K15N, and R29G to increase stability towards auto-cleavage. The ZiPro V36U mutant contained an amber stop codon at position 36. The ZiPro V36C mutant had two additional mutations, C80S and C143S, to make C36 the only cysteine residue in the protein.

### Expression and purification of PpiB

PpiB with PSc incorporated at position 147 was produced in *E. coli* BL21(DE3). The cells were transformed with the plasmids pCDF-*Mm* PCC2RS/$tRNA_{CUA}$ and pETMCSI-PpiB H147U-$His_6$ and starter cultures grown for 16 h at 37 °C in Luria-Bertani (LB) medium supplemented with 100 µg/mL ampicillin and 25 µg/mL spectinomycin. Overnight cultures were inoculated into fresh LB medium (1:100 dilution), supplemented with ampicillin and spectinomycin, and grown at 37 °C. After the cultures reached an $OD_{600}$ value of 0.5–0.8, isopropyl-*β*-D-thiogalactopyranoside (IPTG) and PSc were added to final concentrations of 1 mM. The cells were harvested following overnight expression in the dark at room temperature. Wild-type PpiB and PpiB H147C were produced in the same way, following transformation with the pETMCSI vectors of the respective proteins and growth in the absence of spectinomycin and PSc without exclusion of light. The cell pellet was resuspended in buffer A (50 mM phosphate, pH 7.5, 300 mM NaCl) and lysed by sonication at 4 °C. The lysate was clarified by centrifugation (17300 g, 40 min, 4 °C) and loaded onto a 5 mL Co-NTA column (GE Healthcare, USA) equilibrated with buffer A. The target proteins were eluted using a gradient buffer mixture of buffer A and buffer B (buffer A containing, in addition, 300 mM imidazole). For PpiB H147U, buffer B was exchanged for buffer C (50 mM phosphate, pH 7.5, 150 mM NaCl, 1 mM DTT, 10% glycerol) for storage at -80 °C or buffer D (100 mM

acetate, pH 5.1) for photolysis, using centrifugal filter units with a molecular weight cutoff of 10 kDa (Amicon Ultra, Millipore, Billerica, USA). Wild-type PpiB and PpiB H147C were stored in buffer C.

**Expression and purification of ZiPro**

ZiPro with PSc incorporated at position 36 was produced in the same way as described above for PpiB, except that the protein was expressed using the vector pETMCSI-ZiPro V36U-$His_6$, and the wild-type ZiPro and ZiPro V36C/C80S/C143S proteins were expressed similarly using pETMCSI vectors. Mass spectra showed that the protein loses its N-terminal methionine during *in vivo* expression (Figure S3C).

**Photolysis**

Decaging of the PSc residue in PpiB H147U and ZiPro V36U was performed in buffer D. Fresh dithiothreitol (DTT) was added to a final concentration of 1 mM, the samples were left exposed to nitrogen in a low-oxygen glove box for 10 minutes at room temperature, sealed and shaken for another 20 minutes. Then the samples were illuminated for 60 min at 4 °C, using a hand-held UV lamp (365 nm, 1200 μW/$cm^2$; UVGL-58 from UVP, Cambridge, UK).

**Selective Sec labeling**

A five-fold excess of TMS-1 or TMS-2 tag was added to the decaged protein samples and shaken for 5 min in the dark at room temperature. The reaction was quenched by the addition of 2% β-mercaptoethanol.[15] Excess tag was removed by dialysis into NMR buffer (50 mM HEPES-KOH, pH 8, for the PpiB samples and 20 mM MES-KOH, pH 6.8, 150 mM NaCl, 1 mM DTT for the ZiPro samples). Control experiments with wild-type ZiPro and ZiPro V36PSc were conducted under the same conditions, preparing the protein samples in buffer D, adding fresh DTT to a final concentration of 1 mM, leaving the samples in a low-oxygen glove box for 10 minutes at room temperature, sealing and shaking for another 20 minutes, and finally carrying out the same labeling procedure with TMS-1.

**Cysteine labeling**

Fresh DTT was added to a final concentration of 1 mM to solutions of wild-type PpiB and wild-type ZiPro in buffer E (50 mM phosphate, pH 7.5, 150 mM NaCl) and the samples were incubated at room temperature for 1 hour. Excess DTT was removed by passing the samples through PD10 columns (GE Healthcare, USA) before adding a five-fold excess of TMS-1. The reaction mixtures were incubated overnight in the dark at room temperature. PpiB H147C and ZiPro V36C/C80S/C143S were labeled with TMS-1 in buffer D, pH 5.1, following the same procedure. Reaction rates were quenched for aliquots taken in regular intervals over a period of 16 h by adding 2% β-mercaptoethanol.

**NMR spectroscopy**

All 1D $^1$H NMR spectra were recorded of 10-50 μM protein solutions in aqueous buffer containing 10% $D_2O$ at 25 ºC, using a Bruker 800 MHz NMR spectrometer equipped with a TCI cryoprobe. The solvent signal was suppressed using the double spin-echo pulse sequence.[43]

**Mass spectrometry**

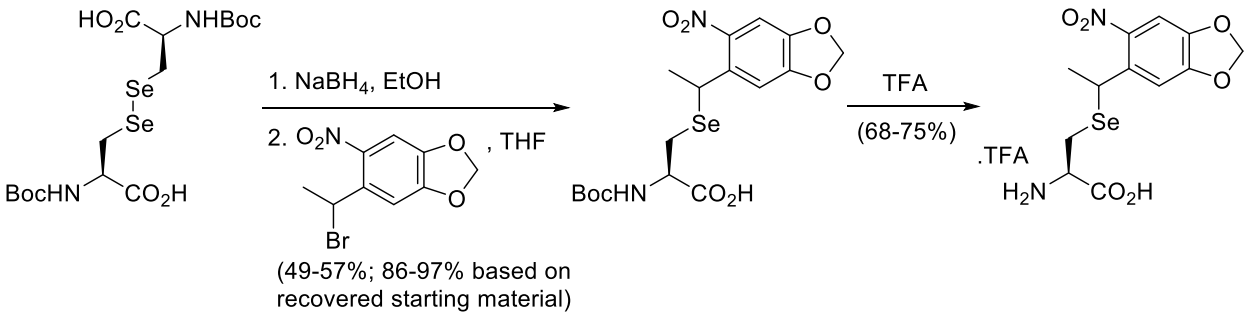
Mass spectrometry of proteins was carried out using an Orbitrap Elite Hybrid Ion Trap-Orbitrap mass spectrometer coupled with an UltiMate 3000 UHPLC (Thermo Scientific, USA). Samples were injected into the mass analyser via an Agilent ZORBAX SB-C3 Rapid Resolution HT Threaded Column.

**Determination of $K_d$ values**

The dissociation constant $K_d$ was determined by fitting the equation 1 to the data points of Figure 6C. In equation 1,

$$\Delta\delta = \Delta\delta_{max}\frac{([L] + [P] + K_d) - \sqrt{([L] + [P] + K_d)^2 - 4[L][P]}}{2[P]} \quad (1)$$

Δδ is the observed chemical shift difference, $\Delta\delta_{max}$ the maximal chemical shift difference at saturation, and [L] and [P] are the total concentrations of ligand and protein, respectively. The same equation applies to signal intensities measured in the limit of slow chemical exchange, with Δδ replaced by the relative signal intensity of the complex and $\Delta\delta_{max}$ replaced by the total signal intensity of free and bound protein.[32]




**Scheme 1.** Synthesis of photocaged selenocysteine (PSc).

**Synthetic procedures and methods**

Chemicals and solvents were purchased from standard suppliers and used without further purification. Analytical thin-layer chromatography analysis (TLC) was performed on pre-coated silica gel aluminium-backed plates (Merck Kieselgel 60 $F_{254}$). Visualisation was by examination under UV light at 254 nm, with subsequent staining by KMnO4. Dry-flash column chromatography[44] was run using Davisil® P60 silica gel (40–63µm). Optical rotation was measured in a JASCO P-2000 polarimeter at the sodium D line in a cell with 100 mm path length. NMR spectra for compound analysis were recorded on a Bruker NMR spectrometer at 400 MHz ($^1$H) and 101 MHz ($^{13}$C). Chemical shifts (δ) are recorded in parts per million (ppm) downfield in relation to tetramethylsilane using the residual solvent as internal standard ($CDCl_3$, $\delta_H$ = 7.26 and $\delta_C$ = 77.16; DMSO, $\delta_H$ = 2.50 and $\delta_C$ = 39.52).[45] High-resolution mass spectra (HRMS) were obtained from an Agilent 6224 TOF LC/MS Mass Spectrometer coupled to an Agilent 1290 Infinity. All data were acquired and reference mass corrected *via* a dual-spray electrospray ionisation (ESI) source. Liquid chromatography mass spectrometry (LCMS) was performed with an Agilent 6120 Series Single Quad coupled to an Agilent 1260 Series HPLC. The following buffers were used: buffer A: 0.1% formic acid in H2O; buffer B: 0.1% formic acid in MeCN. The following gradient was used with a Poroshell 120 EC-C18 50 x 3.0 mm 2.7 micron column, and a flow rate of 0.5 mL/min and total run time of 5 min: 0–1 min 95% buffer A and 5% buffer B, from 1–2.5 min up to 0% buffer A and 100% buffer B, held at this composition until 3.8 min, 3.8–4 min 95% buffer A and 5% buffer B, held until 5 min at this composition. Mass spectra were acquired in positive and negative ion mode with a scan range of 100-1000 m/z. UV detection was carried out at 214 and 254 nm.

**_N_-(_tert_-butoxycarbonyl)-[(_R_,_S_)-1-{4',5'-(methylenedioxy)-2'-nitrophenyl}ethyl]-L-selenocysteine.**

The synthesis of photocaged selenocysteine followed Scheme 1. (*R*,*S*)-1-Bromo-1-[4',5'-(methylenedioxy)-2'-nitrophenyl]ethane[9] and Boc-L-selenocysteine[46] were prepared as previously reported. The reaction was run in parallel two times on 1.81 g scale. Sodium borohydride (641 mg, 16.9 mmol) was added to a cooled (0 °C) solution of Boc-L-selenocysteine (1.81 g, 3.39 mmol) in EtOH (33 mL), under $N_2$. The mixture was stirred for 15 min at 0 °C and then allowed to warm to room temperature (RT) for 15 min. A solution of (*R*,*S*)-1-bromo-1-[4',5'-(methylenedioxy)-2'-nitrophenyl]ethane (2.18 g, 7.94 mmol) in anhydrous THF (13 mL) was then added. The two reactions were then stirred at RT for 18 h, before quenching with a solution of saturated aqueous $NH_4Cl/H_2O$ (1:1, 50 mL). The organic solvent was removed under reduced pressure and the resultant aqueous mixture from the two reactions were combined. The aqueous was acidified to pH 1 with concentrated HCl (~ 3mL) and then extracted with $CHCl_3$ (3 × 100 mL). The combined organics were washed with brine, dried ($Na_2SO_4$) and solvent was removed under reduced pressure. The crude material was purified by silica dry-flash chromatography, eluting with a gradient of 15–35% EtOAc in petroleum benzene, to give the title compound as a yellow

foam (3.56 g, 57%). $[\alpha]^{25}_{D}$ = +10.1 (*c* = 1.10, $CHCl_3$). $^1$H NMR (400 MHz, $CDCl_3$) δ (Mixture of epimers, 3:2) 7.31 (s, 1H), 7.15 and 7.13 (2 × s, 1H), 6.10 and 6.09 (2 × d, *J* = 0.9 Hz, 2H), 5.25 (br s, 1H), 5.01 (q, *J* = 6.8 Hz, 1H), 4.61–4.50 (br m, 1H), 3.08–2.97 (br m, 1H), 2.91 (dd, *J* = 12.5, 5.7 Hz, 1H) 1.70 and 1.68 (2 × d, *J* = 5.8 and 6.9 Hz, 3H), 1.44 (s, 9H). $^{13}$C NMR ($CDCl_3$) δ (Mixture of epimers) 175.0 (C), 155.5 (C), 152.0 and 152.0 (C), 146.8 (C), 142.5 and 142.5 (C), 136.2 (C), 108.5 and 108.5 (CH), 105.2 (CH), 103.1 ($CH_2$), 80.7 (C), 53.5 (CH), 32.5 and 32.4 (CH), 28.4 ($CH_3$), 26.6 and 26.4 ($CH_2$), 23.3 ($CH_3$). MS (ES) *m/z*: 461 [M-H]. LCMS: $R_t$ = 3.36 min; area 100%. HRMS: calcd for $C_{17}H_{22}N_2O_8Se$ + $Na^+$, 485.0435; found (ES, $[M+Na]^+$ obsd), 485.0440. Eluting the column with 50% EtOAc in petroleum benzene, followed by 15% MeOH in DCM gave Boc-L-selenocysteine as a pale brown foam (1.44 g, 40%). Spectroscopic data were in agreement with those reported.[16]

**[(*R*,*S*)-1-{4',5'-(methylenedioxy)-2'-nitrophenyl}ethyl]-L-selenocysteine trifluoroacetate.** TFA (27 mL) was added to *N*-(*tert*-butoxycarbonyl)-[(*R*,*S*)-1-{4',5'-(methylenedioxy)-2'-nitrophenyl}ethyl]-L-selenocysteine (3.72 g, 8.06 mmol) and the resultant solution was stirred at RT for 10 min. Volatile components were removed under reduced pressure at 20 °C. The obtained residue was dissolved in MeCN (8 mL) and precipitated into $Et_2O$ (200 mL) by slow addition while stirring vigorously, in a round bottom flask. The solvent was decanted and the solid was washed with $Et_2O$ (50 mL). After decanting, final traces of solvent were removed under reduced pressure to give the title compound as a yellow hydroscopic solid (2.87 g, 75%). $[\alpha]^{26}_{D}$ = -0.3 (*c* = 1.00, DMSO). $^1$H NMR (400 MHz, DMSO-$d_6$) δ (Mixture of epimers, 1:1) 7.49 and 7.48 (2 × s, 1H), 7.29 and 7.28 (2 × s, 1H), 6.21 and 6.20 (2 × s, 2H), 4.79 and 4.77 (2 × q, *J* = 7.0 and 7.0 Hz, 1H), 3.59–3.50 (br m, 1H), 2.95 and 2.87 (2 × dd, *J* = 12.5, 5.3 and 12.5, 5.5 Hz, 1H), 2.81–2.71 (br m, 1H), 1.67 and 1.65 (2 × d, *J* = 7.0 and 7.0 Hz, 3H). $^{13}$C NMR (DMSO-$d_6$) δ (Mixture of epimers) 169.2 and 169.2 (C), 151.6 and 151.6 (C), 146.4 and 146.4 (C), 142.0 and 141.8 (C), 135.7 and 135.5 (C), 108.1 and 108.0 (CH), 104.7 and 104.6 (CH), 103.3 ($CH_2$), 53.7 and 53.6 (CH), 32.2 and 31.8 (CH), 24.7 and 24.5 ($CH_2$), 22.5 and 22.5 ($CH_3$). MS (ES) *m/z*: 361 [M-H]. LCMS: $R_t$ = 2.82 min; area 98%. HRMS: calcd for $C_{12}H_{14}N_2O_6Se$ + $H^+$, 363.0090; found (ES, $[M+H]^+$ obsd), 363.0093.


## ACKNOWLEDGMENTS

We thank Mr Mithun Chamikara Mahawaththage for analyses by mass spectrometry and Dr Christoph Nitsche for the plasmid of ZiPro V36C/C80S/C143S. Financial support by the Australian Research Council, including a Laureate Fellowship to G.O., is gratefully acknowledged.


## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

$^1$H NMR spectra of the synthetic intermediates and PSc.

**REFERENCES**


(1) Chalker, J. M., Bernardes, G. J. L., Lin, Y. A., and Davis, B. G. (2009) Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem. Asian J. 4*, 630–640.

(2) Baslé, E., Joubert, N., and Pucheault, M. (2010) Protein chemical modification on endogenous amino acids. *Chem. Biol. 17*, 213–227.

(3) Su, X.-C., and Otting, G. (2010) Paramagnetic labeling of proteins and oligonucleotides for NMR. *J. Biomol. NMR 46*, 101–112.

(4) Spicer, C. D., and Davis, B. G. (2014) Selective chemical protein modification. *Nat. Commun. 5*, 4740.

(5) Boutureira, O., and Bernardes, G. J. L. (2015) Advances in chemical protein modification. *Chem. Rev. 115*, 2174–2195.

(6) Krall, N., da Cruz, F. P., Boutureira, O., and Bernardes, G. J. L. (2016) Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem. 8*, 103–113.

(7) Roser, P., Schmidt, M. J., Drescher, M., and Summerer, D. (2016) Site-directed spin labeling of proteins for distance measurements *in vitro* and in cells. *Org. Biomol. Chem. 14*, 5468–5476.

(8) Nitsche, C., and Otting, G. (2017) Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags. *Prog. Nucl. Magn. Reson. Spectrosc. 98–99*, 20–49.

(9) Nguyen, D. P., Mahesh, M., Elsässer, S. J., Hancock, S. M., Uttamapinant, C., and Chin, J. W. (2014) Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc. 136*, 2240–2243.

(10) Yang, K., Li, G., Gong, P., Gui, W., Yuan, L., and Zhuang, Z. (2016) Chemical protein ubiquitylation with preservation of the native cysteine residues. *ChemBioChem 17*, 995–998.

(11) Johansson, L., Gafvelin, G., and Arnér, E. S. (2005) Selenocysteine in proteins-properties and biotechnological use. *Biochim. Biophys. Acta 1726*, 1–13.

(12) Arnér, E. S. (2010) Selenoproteins-what unique properties can arise with selenocysteine in place of cysteine? *Exp. Cell Res. 316*, 1296–1303.

(13) Mousa, R., Dardashti, N. R., and Metanis, N. (2017) Selenium and selenocysteine in protein chemistry. *Angew. Chem. Int. Ed. 56*, 15818–15827.

(14) Cheng, Q., Johansson, L., Thorell, J. O., Fredriksson, A., Samén, E., Stone-Elander, S., and Arnér, E. S. J. (2006) Selenolthiol and dithiol C-terminal tetrapeptide motifs for one-step purification and labeling of recombinant proteins produced in *E. coli*. *ChemBioChem 7*, 1976–1981.

(15) Cheng, Q., Stone-Elander, S., and Arnér, E. S. (2006) Tagging recombinant proteins with a Sel-tag for purification, labeling with electrophilic compounds or radiolabeling with 11C. *Nat. Protoc. 1*, 604–613.

(16) Rakauskaitė, R., Urbanavičiūtė, G., Rukšėnaitė, A., Liutkevičiūtė, Z., Juškėnas, R., Masevičius, V., and Klimašauskas, S. (2015) Biosynthetic selenoproteins with genetically-encoded photocaged selenocysteines. *Chem. Commun. 51*, 8245–8248.

(17) Liu, J., Chen, Q., and Rozovsky, S. (2017) Utilizing selenocysteine for expressed protein ligation and bioconjugations. *J. Am. Chem. Soc. 139*, 3430–3437.

(18) Mitchell, N. J., Kulkarni, S. S., Malins, L. R., Wang, S., and Payne, R. J. (2017) One-pot ligation–oxidative deselenization at selenocysteine and selenocystine. *Chem. Eur. J*., *23*, 946–952.

(19) Malins, L. R., Mitchell, N. J., McGowan, S., and Payne, R. J. (2015) Oxidative deselenization of selenocysteine: applications for programmed ligation at serine. *Angew. Chem. Int. Ed*. *54*, 12716–12721.

(20) Cheng, Q., and Arnér, E. S. J. (2017) Selenocysteine at a predefined UAG codon in a release factor 1 (RF1)-depleted *Escherichia coli* host strain bypasses species barriers in recombinant selenoprotein translation. *J. Biol. Chem. 292*, 5476–5487.

(21) Aldag, C., Bröcker, M. J., Hohn, M. J., Prat, L., Hammond, G., Plummer, A., and Söll, D. (2013) Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. *Angew. Chem. Int. Ed. 52*, 1441–1445.

(22) Thyer, R., Robotham, S. A., Brodbelt, J. S., and Ellington, A. D. (2015) Evolving tRNA(Sec) for efficient canonical incorporation of selenocysteine. *J. Am. Chem. Soc. 137*, 46–49.

(23) Miller, C., Bröcker, M. J., Prat, L., Ip, K., Chirathivat, N., Feiock, A., Veszprémi, M., and Söll, D. (2015) A synthetic tRNA for EF-Tu mediated selenocysteine incorporation *in vivo* and *in vitro*. *FEBS Lett. 589*, 2194–2199.

(24) Fan, Z., Song, J., Guan, T., Lv, X., and Wei, J. (2018) Efficient expression of glutathione peroxidase with chimeric tRNA in amber-less *Escherichia coli*. *ACS Synth. Biol. 7*, 249–257.

(25) Fu, X., Söll, D., and Sevostyanova, A. (2018) Challenges of site-specific selenocysteine incorporation into proteins by *Escherichia coli*. *RNA Biol.* doi: 10.1080/15476286.2018.1440876.

(26) Mukai, T., Sevostyanova, A., Suzuki, T., Fu, X., and Söll, D. (2018) A facile method for producing selenocysteine-containing proteins. *Angew. Chem. Int. Ed. 57,* doi: 10.1002/anie.201713215.

(27) Müller, S., Senn, H., Gsell, B., Vetter, W., Baron, C., and Böck, A. (1994) The formation of diselenide bridges in proteins by incorporation of selenocysteine residues: biosynthesis and characterization of $(Se)_2$-thioredoxin. *Biochemistry 33*, 3404–3412.

(28) Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M., and Schultz, P. G. (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol. 3*, 769–772.

(29) Lammers, C., Hahn, L. E., and Neumann, H. (2014) Optimized plasmid systems for the incorporation of multiple different unnatural amino acids by evolved orthogonal ribosomes. *ChemBioChem 12,* 1800–1804.

(30) Ma, S., Caprioli, R. M., Hill, K. E., and Burk, R. F. (2003) Loss of selenium from selenoproteins: conversion of selenocysteine to dehydroalanine *in vitro*. *J. Am. Soc. Mass Spectr. 14*, 593–600.

(31) Jabar, S., Adams, L. A., Wang, Y., Aurelio, L., Graham, B., and Otting, G. (2017) Chemical tagging with tert-butyl and trimethylsilyl groups for measuring intermolecular nuclear overhauser effects in a large protein-ligand complex. *Chem. Eur. J. 23*, 13033–13036.

(32) Becker, W., Adams, L. A., Graham, B., Wagner, G. E., Zangger, K., Otting, G., and Nitsche, C. (2018) Trimethylsilyl tag for probing protein-ligand interactions by NMR. *J. Biomol. NMR*, doi 10.1007/s10858-018-0173-6.

(33) Konno, M., Ito, M., Hayano, T., and Takahashi, N. (1996) The substrate-binding site in *Escherichia coli* cyclophilin A preferably recognizes a *cis*-proline isomer or a highly distorted form of the *trans* isomer. *J. Mol. Biol. 256*, 897−908.

(34) Guignard, L., Ozawa, K., Pursglove, S. E., Otting, G., and Dixon, N. E. (2002) NMR analysis of *in vitro*-synthesized proteins without purification: a high-throughput approach. *FEBS Lett. 524*, 159−162.

(35) Williamson, M. P. (2013) Using chemical shift perturbation to characterize ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc. 73*, 1–16.

(36) Lei, J., Hansen, G., Nitsche, C., Klein, C. D., Zhang, L., and Hilgenfeld, R. (2016) Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. *Science 353*, 503–505.

(37) Srinivasan, G., James, C. M., and Krzycki, J. A. (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science 296*, 1459–1462.

(38) Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008) Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun. 371*, 818–822.

(39) Bianco, A., Townsley, F. M., Greiss, S., Lang, K., and Chin, J. W. (2012) Expanding the genetic code of *Drosophila melanogaster*. *Nat. Chem. Biol. 8*, 748–750.

(40) Hancock, S. M., Uprety, R., Deiters, A., and Chin, J. W. (2010) Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J. Am. Chem. Soc. 132*, 14819–14824.

(41) Neylon, C., Brown, S. E., Kralicek, A. V., Miles, C. S., Love, C. A., and Dixon, N. E. (2000) Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry 39,* 11989–11999.

(42) Mahawaththa, M. C., Pearce, B. J. G., Szabo, M., Graham, B., Klein, C. D., Nitsche, C., and Otting,

G. (2017) Solution conformations of a linked construct of the Zika virus NS2B-NS3 protease. *Antiviral Res. 142,* 141–147.

(43) Hwang, T. L., and Shaka, A. J. (1995) Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson.* A *112,* 275–279.

(44) Harwood, L. M. (1985) *Aldrichimica Acta 18*, 25.

(45) Gottlieb, H. E., Kotlyar, V., and Nudelman, A. (1997) NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem. 62*, 7512–7515.

(46) Agnihotri, G., Crall, B. M., Lewis, T. C., Day, T. P., Balakrishna, R., Warshakoon, H. J., Malladi, S. S., and David, S. A. (2011) Structure-activity relationships in Toll-like receptor 2-agonists leading to simplified monoacyl lipopeptides. *J. Med. Chem. 54*, 8148−8160.