

Cell-free synthesis of selenoproteins in high yield and purity for selective protein tagging

Adarshi P. Welegedara,^{†§} Ansis Maleckis[‡], Ruchira Bandara,[†] Mithun C. Mahawaththa,[†]
Iresha Dilhani Herath,[†] Yi Jiun Tan,[†] Angeliki Giannoulis,[¶] Daniella Goldfarb,[¶] Gottfried
Otting,^{*†} Thomas Huber^{*†}

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

[‡] Latvian Institute of Organic Synthesis, LV-1006, Riga, Latvia

[¶] Department of Chemical and Biological Physics, Weizmann Institute of Science, Rehovot, 76100, Israel

Corresponding Authors

* t.huber@anu.edu.au, gottfried.otting@anu.edu.au

Present Addresses

§ Department of Chemistry, University of Peradeniya, Peradeniya, 20400, Sri Lanka.

Abstract

The selenol group of selenocysteine is much more nucleophilic than the thiol group of cysteine. Selenocysteine residues in proteins thus offer reactive points for rapid post-translational modification. Here we show that selenoproteins can be expressed in high yield and purity by cell-free protein synthesis by global substitution of cysteine by selenocysteine. Complete alkylation of solvent-exposed selenocysteine residues was achieved in 10 minutes with 4-chloro-methylene-dipicolinic acid (4Cl-MDPA) under conditions that left cysteine residues unchanged even after overnight incubation. Gd(III)–Gd(III) distances measured by double electron–electron resonance (DEER) experiments of maltose binding protein (MBP) containing two selenocysteine residues tagged with 4Cl-MDPA-Gd(III) were indistinguishable from Gd(III)–Gd(III) distances measured of MBP containing cysteine reacted with 4Br-MDPA tags.

Keywords

Selenoproteins, selenocysteine, protein labelling, paramagnetic tags, electron paramagnetic resonance

Introduction

Site-specific chemical modifications of proteins are fundamental to protein research and required in many applications, such as immobilization of proteins on chips or the posttranslational attachment of solubility tags or spectroscopic probes for studies by fluorescence, nuclear magnetic resonance (NMR) spectroscopy, or electron paramagnetic resonance (EPR).¹⁻⁸ EPR spectroscopy with paramagnetic tags has recently emerged as an important tool in structural biology, where double electron–electron resonance (DEER) experiments yield the distance distribution between two site-specifically attached paramagnetic tags.⁹ DEER experiments are attractive, because they give access to accurate distances ranging between about 2–8 nm, which are difficult to measure by any other spectroscopic technique and particularly suited to study large conformational changes of proteins.¹⁰ Furthermore, DEER experiments yield not only distances, but entire distance distributions.

Most often, the site-specific attachment of the paramagnetic tags is achieved by reaction with the thiol group of a cysteine residue in the target protein, with formation of either a covalent disulfide or a thioether bond. For example, the most commonly used EPR tag, *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL), spontaneously reacts with cysteine to form a disulfide bond.¹² Cysteine stands out as the most nucleophilic and oxidation sensitive amino acid among the 20 canonical amino acids, enabling site-specific tagging. Tagging of site-specifically introduced cysteine residues, however, is incompatible with proteins that depend on cysteines in their wild-type amino acid sequence to maintain structural and functional integrity. It is therefore of interest to develop chemistries that allow site-specific tagging in a manner that leaves native cysteine residues intact.

An ideal scenario would involve an unnatural amino acid that contains a paramagnetic center and can be incorporated into proteins in response to a stop codon with the help of a suppressor tRNA and a suitable aminoacyl-tRNA synthetase. Such a system has been established, but the amino acid carries the paramagnetic center at the end of a long side chain, limiting its use for accurate distance measurements.¹³ Alternatively, *p*-acetylphenylalanine or *p*-azido-phenylalanine can be incorporated into proteins by similar systems and their terminal groups reacted with EPR tags.¹⁴⁻¹⁶ This also allows the attachment of tags containing Gd(III) ions, which are particularly suited for sensitive DEER measurements with W-band EPR instruments.^{11, 14-17} Again, these schemes generate tethers between protein and EPR-active paramagnetic center that are long compared with tags attached to cysteine residues. Even though the tethers can be relatively rigid,¹⁴⁻¹⁶ the ultimate tether length limits the utility of these amino acids in structural biology. In the context of DEER experiments, the paramagnetic centers of the tags are best positioned close to the protein backbone, so that the width of the DEER distance distribution reflects the structure and conformational variability of the protein rather than the tags.

A possible solution of these limitations could be provided by a scheme, where photocaged selenocysteine is incorporated into the target protein as an unnatural amino acid and subsequently deprotected to yield selenocysteine (Sec), which can be tagged like cysteine (Cys) while showing greater reactivity at neutral pH owing to the greater acidity of the selenol group ($pK_a = 5.2$) versus the thiol group ($pK_a = 8.5$).¹⁸⁻²⁰ This scheme would produce a short tether, while the increased reactivity of selenocysteine versus cysteine provides selectivity. The present work explores the suitability of selenocysteine versus cysteine for the selective attachment of Gd(III) tags.

A unique translation mechanism exists in nature, by which selenocysteine can be incorporated in response to a UGA stop codon. It involves a series of specific enzymes, a suppressor tRNA (tRNA^{Sec}), and a conserved mRNA sequence (selenocysteine insertion sequence or SECIS) in the vicinity of the target UGA codon.^{18, 20} The efficiency and applicability of this mechanism has been greatly broadened by enzyme and RNA engineering, but it is still not as efficient as the incorporation of cysteine.²¹⁻³³ In a much simpler approach, proteins with selenocysteine can be made in bacteria simply by replacing Cys in the feedstock by Sec, substituting all Cys residues by Sec by misloading cysteinyl-tRNA with selenocysteine.^{22, 30} Notably, however, Sec is highly toxic and Cys released by protein turnover will compete with Sec incorporation. A 90% replacement yield of single Cys residues by Sec has been reported for a Cys-auxotroph *E. coli* strain in a defined growth medium containing Sec instead of Cys.³⁰

Here we introduce a cell-free protein synthesis (CFPS) platform to synthesise proteins with selenocysteine in high yield and purity. CFPS is ideally suited for synthesis under cytotoxic conditions and, as many biosynthetic enzymes are inactivated in cell-free extracts, we hypothesized that global substitution of Cys for Sec would be straightforward. We show for three proteins, the NS2B-NS3 protease from Zika virus (ZiPro), GB1, and *E. coli* maltose binding protein (MBP), that milligram quantities of protein containing Sec can be produced from a single milliliter of reaction mixture. Furthermore, we demonstrate that Sec can be tagged with alkylating reagents under conditions that barely affect Cys residues. The specificity of the tagging reaction is illustrated by DEER distance distributions that are indistinguishable between tags attached to the double-Sec mutant of MBP and the corresponding double-Cys mutant.

Results

Production of selenoproteins by cell-free protein synthesis. Selenocysteine versions of four protein mutants, ZiPro C80U, ZiPro V36U, GB1 Q32U and MBP T237U/T345U, where the character U stands for Sec, were produced in good yield by misloading cysteinyl tRNA with selenocysteine during CFPS. As selenocysteine is highly sensitive to oxidation and not commercially available, our experiments started from selenocystine, which had to be reduced in the reaction mixture. We used DTT to maintain reducing conditions, as DTT has been reported to be more effective than sodium borohydride or β -mercaptoethanol for the reduction of diselenide bonds.³⁴ Initial CFPS experiments using our standard amount of DTT (1 mM), however, failed to produce significant amounts of selenoproteins. A series of CFPS experiments with increasing concentrations of DTT showed that ZiPro C80U was produced in much greater yields simply by raising the concentration of DTT to 10 mM (Figure 1a). Therefore, we conducted all subsequent CFPS reactions of selenoproteins in the presence of 10 mM DTT. At this DTT concentration, 1 mL of CFPS reaction mixture yielded about 1.2 mg of ZiPro C80U with selenocysteine. LC-MS/MS of a tryptic digest of the ZiPro C80U isolated by SDS-PAGE confirmed the presence of a Sec residue in position 80 as expected (Figure 1b). The same CFPS reaction conditions produced 1.5 mg of the double-Sec mutant of MBP (MBP T237U/T345U), 1.8 mg of ZiPro V36U, and 5.0 mg of GB1 Q32U. Red precipitate formed in the course of the CFPS reaction, indicating the generation of a limited amount of elemental selenium. Attempts to purify ZiPro C80U by chromatography over a Ni-NTA column resulted in a product of molecular weight 296 Dalton greater than expected, suggesting reaction of the Sec residue with an unknown compound in the CFPS reaction.

ESI mass spectrometry of MBP T237U/T345U and of ZiPro V36U indicated that the proteins were monomers, and there was no indication of misincorporation of any of the canonical amino acids in response to the cysteine codon (Figure 2a and c). In contrast, GB1 Q32U was a dimer (Figure 2d). Selenocysteine is prone to oxidation to selenocystine³⁵ and ready formation of a selenium–selenium bond would be expected for the solvent-exposed selenocysteine residue of GB1 Q32U.

Relative reactivities of 4F-MDPA, 4Cl-MDPA, and 4Br-MDPA with cysteine and selenocysteine. To compare the reactivity of selenocysteine and cysteine in nucleophilic tagging reactions, we tested reactions with the three 4-halogen-methylene-dipicolinic acid (4halogen-MDPA) tags 4Br-MDPA, 4Cl-MDPA and 4F-MDPA (Figure 3). In particular, high tagging yields were sought with the aim of performing DEER experiments with Gd(III) ions on tagged double-Sec and double-Cys mutants of MBP (MBP T237U/T345U and MBP T237C/T345C, respectively). 4Cl-MDPA did not react with the double-Cys mutant even after 16 hours at room temperature and pH 7.5. In contrast, tagging of the double-Sec mutant with 4Cl-MDPA was complete within 10 minutes at pH 7.5 and room temperature (Figure 2a). To prevent oxidation of the selenol group, the tagging reaction was conducted in the presence of 5 mM DTT. In contrast, tagging of the double-Cys mutant required 4Br-MDPA and was conducted in the absence of DTT to avoid reaction of the tag with DTT. Complete ligation of the double-Cys mutant was achieved by incubation with 4Br-MDPA for 16 hours at room temperature and pH 7.5 (Figure 2b). These results highlight the much greater reactivity of selenol groups in nucleophilic substitution reactions compared with thiol groups. 4F-MDPA proved completely inert, failing to react with the Sec mutant ZiPro V36U to any significant extent even after 32 h incubation at pH 7.5 in the presence of TCEP, whereas 4Cl-MDPA reacted within 10 minutes (Figure 2c).

Presence of a Se–Se bond in the dimer of GB1 Q32U did not prevent tagging, as the addition of 5 mM DTT proved sufficient to tag GB1 Q32U with 4Cl-MDPA within 10 minutes (Figure 2d).

DEER experiment of double-tagged MBP. To assess the utility of the 4Cl-MDPA and 4Br-MDPA tags for distance measurements by DEER experiments, the double-Cys and double-Sec mutants of MBP tagged with MDPA were titrated with Gd(III) in 1:1 ratio of Gd(III):tag and measured on a W-band EPR spectrometer. The echo-detected EPR spectra (Figure S1) revealed full widths at half height of the central transition of, respectively, 5.2 mT and 6.6 mT for the double-Sec and double-Cys mutants of MBP. As the selenium or sulfur atoms are too far from the Gd(III) ion to influence its zero-field splitting, we attribute the difference to the presence of some free Gd(III) in the sample of the double-Sec mutant, which would also explain the longer Gd(III) phase memory time observed for the double-Sec mutant. The DEER data yielded very similar distance distributions, indicating equivalent performance of the double-Sec and double-Cys mutants (Figure 4). A somewhat lesser modulation depth observed for the double-Sec mutant again is consistent with the presence of some free Gd(III) in this sample. The maxima of the experimental distance distributions, observed at about 4.2 and 4.3 nm, respectively, were in reasonable agreement with the distance distribution modelled on the crystal structure 1OMP,³⁶ which suggested 3.9 and 4.0 nm as the most frequent Gd–Gd distance (Figure 4).

Discussion

Cell-free production of selenoproteins. The present work shows that, with modified conditions, cell-free protein synthesis provides exceptionally facile access to pure selenoproteins by global substitution of cysteine. Compared to *in vivo* expression systems, there is no need to take recourse to cysteine auxotrophs or to guard for the presence of cysteine produced by biosynthesis or released by protein turnover. Furthermore, CFPS is ideally suited for the production of proteins with amino acids that are toxic *in vivo*, such as selenocysteine, and the proteins produced in this work were obtained in yields comparable to proteins produced from the 20 canonical amino acids. While *in vivo* protocols recommend washing of the cells to flush out cysteine prior to the addition of selenocysteine,³⁰ the CFPS protocol is much more straightforward. Not having to lyse cells in order to release the selenoproteins also minimizes the chances for selenium oxidation and elimination that have been reported previously.

To misload cysteinyl-tRNA with selenocysteine requires the provision of selenocysteine rather than its oxidized form selenocystine, in which it is commercially available. The diselenide bond in selenocystine can be reduced with water soluble reducing agents such as DTT, TCEP, or NaBH₄, but too strong reducing conditions are detrimental as the selenium of selenol groups is easily reduced to elemental selenium. On the other hand, oxygen readily re-oxidizes selenocysteine to selenocystine. In our hands, increased concentrations of DTT proved to provide the right balance and only a small amount of elemental selenium precipitation was observed. Furthermore, analysis by mass spectrometry provided no evidence of selenium elimination to dehydroalanine or serine, which can arise as a consequence of selenoprotein oxidation followed by *syn*- β -elimination of selenenic acid.^{21, 37}

Reactivities of 4F-MDPA, 4Cl-MDPA, and 4Br-MDPA towards Sec and Cys. Sec residues in proteins present sites of greater chemical reactivity than Cys residues, allowing enhanced rates of chemical labeling. In the case of 4Cl-MDPA and at pH 7.5, labeling of Sec residues was complete within 10 minutes, whereas Cys residues were not noticeably modified even after 16 h incubation at room temperature. Tagging of Cys residues was achieved with 4Br-MDPA but, in the example of MBP T237C/T345C, still required overnight incubation to go to completion, as 60 % was still untagged after 3 hours incubation. 4F-MDPA was expected to be the least reactive tag, as the relative performance of halogen ions as leaving group in nucleophilic substitution reactions decreases in the sequence Br⁻ > Cl⁻ > F⁻. Our failure to tag selenocysteine with 4F-MDPA thus was not unexpected.

The 4Cl-MDPA tag thus emerged as ideally tuned to discriminate between Sec and Cys. The very large difference in reactivity of 4Cl-MDPA towards Sec as opposed to Cys at near-neutral pH revealed in the present work suggests that, quite generally, Sec residues can be selectively and quantitatively tagged in the presence of solvent-exposed cysteine residues without any significant conversion of the cysteine thiol groups. Furthermore, the reaction rate of 4Cl-MDPA with Sec was fast, which bodes well for labeling of proteins that are sensitive to precipitation or other pathways of degradation. This opens the door to a bio-orthogonal labeling scheme, whereby the installation of a single chemically caged Sec residue in the target protein provides a selective site for tag attachment. Systems for genetic encoding of photocaged selenocysteine have been described, but unfortunately are not very efficient.^{21, 31} Enhancing the yields of proteins containing masked selenocysteine will be a worthwhile project.

Tagging of Sec and Cys residues for DEER experiments. Dipicolinic acid is a small and rigid chemical moiety that binds lanthanide ions with nanomolar affinities.³⁸ Lanthanide tags based on DPA thus offer straightforward ways to attach paramagnetic Gd(III) ions to proteins for

DEER measurements.³⁹⁻⁴¹ The first tag was 4-mercaptomethyl-dipicolinic acid (4MMDPA), which required activation of the cysteine residues prior to tagging and produced a relatively flexible tether with a disulfide bond.³⁹⁻⁴⁰ In subsequent tags, the methylene group in the tether was eliminated by the use of 3-mercaptopicolinic acid (3MDPA) or 4-mercaptopicolinic acid (4MDPA), which reduces the flexibility of the tether.⁴¹⁻⁴² To obtain a ligation product that is stable towards reducing agents, 4-vinyl-DPA was introduced, which leads to a thioether but also two methylene groups between the sulfur and DPA moiety.⁴³ The same number of heavy atoms between the sulfur atom (or selenium atom in the case of Sec) and the DPA moiety are obtained with the new 4Cl-MDPA tag. 4-(phenylsulfonyl)-pyridine-2,6-dicarboxylic acid (4PS-DPA) is the only DPA tag that features fewer atoms between the DPA moiety and the protein backbone by direct attachment of the pyridine ring to the sulfur of the cysteine side chain.⁴⁴

The width of the Gd(III)–Gd(III) distance distribution observed with the 4Cl-MDPA tag was 0.3 nm wider at half height than the width obtained previously with C9-Gd(III) tags attached to the same double-Cys mutant of MBP.⁴⁵ In contrast to the DPA tags, the C9 tag is a cyclen tag with bulky phenylethyl-amide pendants that reacts with cysteine residues by formation of a disulfide bridge, and at least one of the methylene groups in the tether is flexible.⁴⁶ The narrow width of the distance distributions obtained with the C9-Gd(III) tag may be attributed to the greater steric restrictions associated with a bulky tag compared to those associated with a smaller tag such as DPA. The maxima of the DEER distance distributions indicated distances that were about 0.3 nm longer than modeled by a rotamer library that ignores potential preferential interactions with other residues (Figure S4). In the case of the T237/T345 mutants of MBP, the site of T237 is highly solvent exposed, whereas the site of T345 is in the middle of a helix with other amino acid side chains nearby. As the modeled distances ranged from 2.4 to 5.1 nm, preferential interactions could well cause a bias in the distance distribution. Nonetheless, the maximum of the DEER distance distribution obtained with the 4Br-MDPA tag on MBP T237C/T345C was within about 0.2 nm of that obtained with the C9-Gd(III) tag.⁴⁷ Narrower and more defined distance distributions may be obtained for sites, where the lanthanide ion can also coordinate to a negatively charged carboxyl group.⁴⁷

Conclusion

Selenocysteine has long been recognized to offer an exquisitely reactive chemical handle for chemical modifications of proteins, but the complexity of selenoprotein synthesis restricted its use in protein chemistry. The cell-free protein synthesis protocol established in the present work provides easy access to selenoproteins in high yield and exceptional purity by global substitution of cysteine. The very fast and selective alkylation of selenocysteine versus cysteine is highly attractive for bioconjugations. Our results encourage the search for methods, by which selenocysteine can be incorporated into proteins site-selectively in high yield and in the presence of cysteine.

Experimental procedures

Synthesis of 4Cl-MDPA, 4Br-MDPA, and 4F-MDPA. The 4Cl-MDPA and 4Br-MDPA tags were synthesized from 4-(hydroxymethyl)pyridine-2,6-dicarboxylate. The protocols and analytical data are provided in the Supporting information.

Plasmid constructs. In the following, we denote selenocysteine as “U” for brevity in mutant names, even though selenocysteine was incorporated in response to a cysteine codon. The ZiPro C80U and ZiPro V36U constructs contained a C-terminal His₆ tag and were cloned into the *Nde*I and *Eco*RI sites of the T7 expression vector pETMSCI.⁴⁸ The wild-type ZiPro construct was described previously.⁴⁹ 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 Gly₄SerGly₄ linker. It also contained the mutations R95*A, K15N, and R29G to increase stability toward autocleavage and the mutations C80S and C143S to eliminate all cysteine residues from the wild-type protein. In the ZiPro V36U mutant, the codon of Val36 was replaced by a cysteine codon, while in the ZiPro C80U mutant, the natural cysteine codon was retained for residue 80.

The double-mutant T237C/T345C of MBP was cloned into the *Nde*I and *Eco*RI sites of the T7 expression pETMCSIII vector⁴⁸ with a N-terminal His₆-tag followed by a tobacco etch virus (TEV) protease recognition site.

The GB1 Q32U construct was cloned into the *Nde*I and *Eco*RI sites of the T7 expression vector pETMSCI.⁴⁸ It contained a MASMTG expression tag at the N-terminus and a TEV protease site followed by a His₆ tag at the C-terminus.

Protein production. All protein samples were produced by continuous exchange cell-free protein synthesis conducted at 30 °C for 16 h, using 1 mL inner reaction mixture and 10 mL outer buffer. The genes of the ZiPro, MBP and GB1 mutants were PCR-amplified with eight-nucleotide single-stranded overhangs to generate DNA suitable for use in CFPS.⁵⁰ The cell-free reactions were set up according to an established protocol⁵¹⁻⁵² modified to promote the incorporation of Sec. Cysteine was omitted and L-selenocystine was added in the inner and outer buffers at a final concentration of 1 mM. DTT was used to reduce selenocystine to selenocysteine. No further special precautions were taken to prevent oxidation. MBP T237U/T345U was produced using the same vector as for the production of MBP T237C/T345C. The DTT concentration was optimized in CFPS experiments of ZiPro C80U, using DTT concentrations in the inner and outer buffers ranging from 2 to 20 mM. 10 mM DTT was used for all subsequent protein expressions.

Protein purification. ZiPro V36U and C80U were purified using a 1 mL His GraviTrap column (GE Healthcare Life Sciences, USA) equilibrated with buffer A (50 mM Tris-HCl, 300 mM NaCl, pH 7.5), washed with buffer B (same as buffer A but with 10 mM imidazole), and eluted with buffer C (same as buffer A but with 300 mM imidazole, 100 mM EDTA). The purified protein samples were dialysed overnight against storage buffer D (same as buffer A but with 5 mM DTT and 10 % glycerol) and concentrated using an Amicon centrifugal ultrafiltration tube with a molecular weight cutoff (MWCO) of 10 kDa.

MBP T237U/T345U and MBP T237C/T345C were purified using a 1 mL mL His GraviTrap column (GE Healthcare Life Sciences, USA) equilibrated with buffer A, washed with buffer E (same as buffer A but with 15 mM imidazole), and eluted with buffer F (same as buffer A but with 300 mM imidazole). Finally, the protein samples were dialyzed against buffer A and concentrated using an Amicon centrifugal ultrafiltration tube (MWCO 10 kDa). GB1 Q32U was purified in the same way, except that the final concentration step used an ultrafiltration tube with MWCO 3 kDa.

Protein tagging with 4Cl-MDPA and 4Br-MDPA. To maintain reducing conditions, the double-Sec and double-Cys mutants of MBP (MBP T237U/T345U and MBP T237C/T345C) were kept in buffer A with 5 mM DTT. MBP T237U/T345U was incubated with the tags at room temperature for 1 hour. 4Cl-MDPA dissolved in DMSO was added in 5-fold molar excess. The samples were shaken at room temperature for 10 minutes and the reactions were quenched by putting the samples into liquid nitrogen. The completeness of the ligation reactions was assessed by mass spectrometry. ZiPro V36U and GB1 Q32U were tagged in the same way. The same tagging protocol was also applied to MBP T237C/T345C, except that, prior to the addition of tag, the protein was incubated with 5 mM DTT for 1 hour followed by removal of the DTT using an Amicon centrifugal ultrafiltration tube (MWCO 10 kDa). For complete tagging of MBP T237C/T345C with 4Br-MDPA, a 50 μ M solution of the protein was incubated overnight with 10-fold molar excess of tag at room temperature.

Mass spectrometry. The proteins were analyzed 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 analyzer via an Agilent ZORBAX SB-C3 Rapid Resolution HT Threaded Column, using an acetonitrile gradient (10–85 %) and 0.1 % formic acid. In-gel digestion for mass spectrometric characterization of ZiPro C80U was performed according to a published protocol.⁵³ In-gel tryptic digestion was performed as described⁵⁴ and the peptide containing the selenocysteine identified by LC-MS/MS.

Sample preparation for EPR spectroscopy. The tagged MBP mutants were concentrated and exchanged into EPR buffer (50 mM MES-KOH in D₂O, pH 6.7, uncorrected pH meter reading) to a final concentration of 100 μ M using an Amicon ultracentrifugation tube (MWCO 10 kDa). Perdeuterated glycerol was added to reach a 20 % (v/v) final composition. GdCl₃ dissolved in D₂O was added in 1:1 molar ratio of GdCl₃ to conjugated Cl/Br-MDPA tag.

DEER experiments. All data were recorded on a home-built W-band (94.9 GHz) spectrometer.⁵³ The data were recorded at 10 K and the temperature was stabilized with a cryo-free cooling system from ColdEdge Technologies. Echo-detected EPR (ED-EPR) spectra (Figure S1) were recorded with a Hahn echo sequence ($\pi/2 - \tau - \pi - \text{echo}$) and sweeping the magnetic field. The microwave pulse lengths were $\pi/2 = 15$ ns, $\pi = 30$ ns, the inter-pulse delay $\tau = 550$ ns and the repetition rate 1 ms. Echo decay traces (Figure S2) were recorded at the maximum of the Gd(III) EPR spectrum monitoring the echo intensity of a Hahn echo sequence by increasing the τ interval in steps of 50 ns and using pulse lengths and repetition times as above.

Gd(III)–Gd(III) DEER data were recorded with the four-pulse ‘reversed’ ($\pi/2_{\text{vobs}} - \tau_1 - \pi_{\text{vobs}} - (\tau_1 - t) - \pi_{\text{vpump}} - (\tau_2 + t) - \pi_{\text{vobs}} - \tau_2 - \text{echo}$) DEER sequence (rDEER)⁵⁵ using chirp pump pulses⁵⁶⁻⁵⁷ and monitoring the echo intensity with increasing time t . An eight-step phase cycle was applied. The microwave pulse lengths at the observe frequency were $\pi/2_{\text{vobs}} = 15$ ns, $\pi_{\text{vobs}} = 30$ ns, the inter-pulse delays $\tau_1 = 5.7$ μ s and $\tau_2 = 5$ μ s, and the repetition rate 0.2 ms. The DEER data were recorded by observing at the maximum of the EPR line at 94.9 GHz and by pumping spins with two chirp pulses set at +100 MHz and –100 MHz from the observed position with a width of 300 MHz (*i.e.* at 95.0–95.3 and 94.5–94.8 GHz, respectively) using pulse lengths of 96 ns. Due to limits set by the cavity bandwidth, the actual excitation bandwidth of the chirp pulses is estimated to be about 100 MHz.

The primary DEER data (Figure S3) were transformed into distance distributions using the DeerAnalysis2018⁵⁸ software and Tikhonov regularization. The background contributions to

the primary DEER data were removed by fitting the dimensionality of the background function to obtain reasonable fits in the frequency and time domain data, as judged by visual inspection. The regularization parameter was 100. The removal of the background contributions was evaluated within the validation tool of the DeerAnalysis2018 program performed from 5 % to 80 % of the DEER time trace in 16 trials, 50% random noise was added in 10 trials and datasets within 15 % of the best root mean square deviation were retained (*i.e.* default prune level 1.15).

ACKNOWLEDGMENT

Financial support by the Australian Research Council (DP200100348 and FL170100019) is gratefully acknowledged.

References

1. Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G., Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chemistry – An Asian Journal* **2009**, *4* (5), 630-640.
2. Baslé, E.; Joubert, N.; Pucheault, M., Protein Chemical Modification on Endogenous Amino Acids. *Chemistry & Biology* **2010**, *17* (3), 213-227.
3. Su, X.-C.; Otting, G., Paramagnetic labelling of proteins and oligonucleotides for NMR. *Journal of Biomolecular NMR* **2010**, *46* (1), 101-112.
4. Spicer, C. D.; Davis, B. G., Selective chemical protein modification. *Nature Communications* **2014**, *5* (1), 4740.
5. Boutureira, O.; Bernardes, G. J. L., Advances in Chemical Protein Modification. *Chemical Reviews* **2015**, *115* (5), 2174-2195.
6. Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L., Site-selective protein-modification chemistry for basic biology and drug development. *Nature Chemistry* **2016**, *8* (2), 103-113.
7. Roser, P.; Schmidt, M. J.; Drescher, M.; Summerer, D., Site-directed spin labeling of proteins for distance measurements in vitro and in cells. *Organic & Biomolecular Chemistry* **2016**, *14* (24), 5468-5476.
8. Nitsche, C.; Otting, G., Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2017**, *98-99*, 20-49.
9. Haugland, M. M.; Lovett, J. E.; Anderson, E. A., Advances in the synthesis of nitroxide radicals for use in biomolecule spin labelling. *Chemical Society Reviews* **2018**, *47* (3), 668-680.
10. Jeschke, G., DEER Distance Measurements on Proteins. *Annual Review of Physical Chemistry* **2012**, *63* (1), 419-446.
11. Feintuch, A.; Otting, G.; Goldfarb, D., Chapter Sixteen - Gd³⁺ Spin Labeling for Measuring Distances in Biomacromolecules: Why and How? In *Methods in Enzymology*, Qin, P. Z.; Warncke, K., Eds. Academic Press: 2015; Vol. 563, pp 415-457.
12. Hubbell, W. L.; Gross, A.; Lengen, R.; Lietzow, M. A., Recent advances in site-directed spin labeling of proteins. *Current Opinion in Structural Biology* **1998**, *8* (5), 649-656.
13. Schmidt, M. J.; Borbas, J.; Drescher, M.; Summerer, D., A Genetically Encoded Spin Label for Electron Paramagnetic Resonance Distance Measurements. *Journal of the American Chemical Society* **2014**, *136* (4), 1238-1241.
14. Abdelkader, E. H.; Feintuch, A.; Yao, X.; Adams, L. A.; Aurelio, L.; Graham, B.; Goldfarb, D.; Otting, G., Protein conformation by EPR spectroscopy using gadolinium tags

- clicked to genetically encoded p-azido-l-phenylalanine. *Chemical Communications* **2015**, 51 (88), 15898-15901.
15. Mahawaththa, M. C.; Lee, M. D.; Giannoulis, A.; Adams, L. A.; Feintuch, A.; Swarbrick, J. D.; Graham, B.; Nitsche, C.; Goldfarb, D.; Otting, G., Small neutral Gd(III) tags for distance measurements in proteins by double electron-electron resonance experiments. *Physical Chemistry Chemical Physics* **2018**, 20 (36), 23535-23545.
 16. Evans, E. G. B.; Millhauser, G. L., Chapter Nineteen - Genetic Incorporation of the Unnatural Amino Acid p-Acetyl Phenylalanine into Proteins for Site-Directed Spin Labeling. In *Methods in Enzymology*, Qin, P. Z.; Warncke, K., Eds. Academic Press: 2015; Vol. 563, pp 503-527.
 17. Fleissner, M. R.; Brustad, E. M.; Kálai, T.; Altenbach, C.; Cascio, D.; Peters, F. B.; Hideg, K.; Peucker, S.; Schultz, P. G.; Hubbell, W. L., Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proceedings of the National Academy of Sciences* **2009**, 106 (51), 21637-21642.
 18. Johansson, L.; Gafvelin, G.; Arnér, E. S. J., Selenocysteine in proteins—properties and biotechnological use. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2005**, 1726 (1), 1-13.
 19. Arnér, E. S. J., Selenoproteins—What unique properties can arise with selenocysteine in place of cysteine? *Experimental Cell Research* **2010**, 316 (8), 1296-1303.
 20. Mousa, R.; Notis Dardashti, R.; Metanis, N., Selenium and Selenocysteine in Protein Chemistry. *Angewandte Chemie International Edition* **2017**, 56 (50), 15818-15827.
 21. Rakauskaitė, R.; Urbanavičiūtė, G.; Rukšėnaitė, A.; Liutkevičiūtė, Z.; Juškėnas, R.; Masevičius, V.; Klimašauskas, S., Biosynthetic selenoproteins with genetically-encoded photocaged selenocysteines. *Chemical Communications* **2015**, 51 (39), 8245-8248.
 22. Liu, J.; Chen, Q.; Rozovsky, S., Utilizing Selenocysteine for Expressed Protein Ligation and Bioconjugations. *Journal of the American Chemical Society* **2017**, 139 (9), 3430-3437.
 23. Cheng, Q.; Arnér, E. S. J., Selenocysteine Insertion at a Predefined UAG Codon in a Release Factor 1 (RF1)-depleted Escherichia coli Host Strain Bypasses Species Barriers in Recombinant Selenoprotein Translation. *Journal of Biological Chemistry* **2017**, 292 (13), 5476-5487.
 24. Aldag, C.; Bröcker, M. J.; Hohn, M. J.; Prat, L.; Hammond, G.; Plummer, A.; Söll, D., Rewiring Translation for Elongation Factor Tu-Dependent Selenocysteine Incorporation. *Angewandte Chemie International Edition* **2013**, 52 (5), 1441-1445.
 25. Thyer, R.; Robotham, S. A.; Brodbelt, J. S.; Ellington, A. D., Evolving tRNA^{Sec} for Efficient Canonical Incorporation of Selenocysteine. *Journal of the American Chemical Society* **2015**, 137 (1), 46-49.
 26. Miller, C.; Bröcker, M. J.; Prat, L.; Ip, K.; Chirathivat, N.; Feiock, A.; Veszprémi, M.; Söll, D., A synthetic tRNA for EF-Tu mediated selenocysteine incorporation in vivo and in vitro. *FEBS Letters* **2015**, 589 (17), 2194-2199.
 27. Fan, Z.; Song, J.; Guan, T.; Lv, X.; Wei, J., Efficient Expression of Glutathione Peroxidase with Chimeric tRNA in Amber-less Escherichia coli. *ACS Synthetic Biology* **2018**, 7 (1), 249-257.
 28. Fu, X.; Söll, D.; Sevostyanova, A., Challenges of site-specific selenocysteine incorporation into proteins by Escherichia coli. *RNA Biology* **2018**, 15 (4-5), 461-470.
 29. Mukai, T.; Sevostyanova, A.; Suzuki, T.; Fu, X.; Söll, D., A Facile Method for Producing Selenocysteine-Containing Proteins. *Angewandte Chemie International Edition* **2018**, 57 (24), 7215-7219.

30. Mueller, S.; Senn, H.; Gsell, B.; Vetter, W.; Baron, C.; Boeck, A., The Formation of Diselenide Bridges in Proteins by Incorporation of Selenocysteine Residues: Biosynthesis and Characterization of (Se)₂-Thioredoxin. *Biochemistry* **1994**, *33* (11), 3404-3412.
31. Welegedara, A. P.; Adams, L. A.; Huber, T.; Graham, B.; Otting, G., Site-Specific Incorporation of Selenocysteine by Genetic Encoding as a Photocaged Unnatural Amino Acid. *Bioconjugate Chemistry* **2018**, *29* (7), 2257-2264.
32. Liu, J.; Zheng, F.; Cheng, R.; Li, S.; Rozovsky, S.; Wang, Q.; Wang, L., Site-Specific Incorporation of Selenocysteine Using an Expanded Genetic Code and Palladium-Mediated Chemical Deprotection. *Journal of the American Chemical Society* **2018**, *140* (28), 8807-8816.
33. Cheng, Q.; Stone-Elander, S.; Arnér, E. S. J., Tagging recombinant proteins with a Sel-tag for purification, labeling with electrophilic compounds or radiolabeling with ¹¹C. *Nature Protocols* **2006**, *1* (2), 604-613.
34. Burnell, J. N.; Karle, J. A.; Shrift, A., Reduction of dl-selenocystine and isolation of l-selenocysteine. *Journal of Inorganic Biochemistry* **1980**, *12* (4), 343-351.
35. Moroder, L., Isosteric replacement of sulfur with other chalcogens in peptides and proteins. *Journal of Peptide Science* **2005**, *11* (4), 187-214.
36. Sharff, A. J.; Rodseth, L. E.; Spurlino, J. C.; Quioco, F. A., Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **1992**, *31* (44), 10657-10663.
37. Ma, S.; Caprioli, R. M.; Hill, K. E.; Burk, R. F., Loss of selenium from selenoproteins: Conversion of selenocysteine to dehydroalanine in vitro. *Journal of the American Society for Mass Spectrometry* **2003**, *14* (6), 593-600.
38. Grenthe, I., Stability Relationships Among the Rare Earth Dipicolinates. *Journal of the American Chemical Society* **1961**, *83* (2), 360-364.
39. Su, X.-C.; Man, B.; Beeren, S.; Liang, H.; Simonsen, S.; Schmitz, C.; Huber, T.; Messerle, B. A.; Otting, G., A Dipicolinic Acid Tag for Rigid Lanthanide Tagging of Proteins and Paramagnetic NMR Spectroscopy. *Journal of the American Chemical Society* **2008**, *130* (32), 10486-10487.
40. Potapov, A.; Yagi, H.; Huber, T.; Jergic, S.; Dixon, N. E.; Otting, G.; Goldfarb, D., Nanometer-Scale Distance Measurements in Proteins Using Gd³⁺ Spin Labeling. *Journal of the American Chemical Society* **2010**, *132* (26), 9040-9048.
41. Man, B.; Su, X.-C.; Liang, H.; Simonsen, S.; Huber, T.; Messerle, B. A.; Otting, G., 3-Mercapto-2,6-Pyridinedicarboxylic Acid: A Small Lanthanide-Binding Tag for Protein Studies by NMR Spectroscopy. *Chemistry – A European Journal* **2010**, *16* (12), 3827-3832.
42. Jia, X.; Maleckis, A.; Huber, T.; Otting, G., 4,4'-Dithiobisdipicolinic Acid: A Small and Convenient Lanthanide Binding Tag for Protein NMR Spectroscopy. *Chemistry – A European Journal* **2011**, *17* (24), 6830-6836.
43. Li, Q.-F.; Yang, Y.; Maleckis, A.; Otting, G.; Su, X.-C., Thiol–ene reaction: a versatile tool in site-specific labelling of proteins with chemically inert tags for paramagnetic NMR. *Chemical Communications* **2012**, *48* (21), 2704-2706.
44. Yang, Y.; Wang, J.-T.; Pei, Y.-Y.; Su, X.-C., Site-specific tagging proteins via a rigid, stable and short thioether tether for paramagnetic spectroscopic analysis. *Chemical Communications* **2015**, *51* (14), 2824-2827.
45. Welegedara, A. P.; Yang, Y.; Lee, M. D.; Swarbrick, J. D.; Huber, T.; Graham, B.; Goldfarb, D.; Otting, G., Double-Arm Lanthanide Tags Deliver Narrow Gd³⁺–Gd³⁺ Distance Distributions in Double Electron–Electron Resonance (DEER) Measurements. *Chemistry – A European Journal* **2017**, *23* (48), 11694-11702.
46. Abdelkader, E. H.; Lee, M. D.; Feintuch, A.; Cohen, M. R.; Swarbrick, J. D.; Otting, G.; Graham, B.; Goldfarb, D., A New Gd³⁺ Spin Label for Gd³⁺–Gd³⁺ Distance

Measurements in Proteins Produces Narrow Distance Distributions. *The Journal of Physical Chemistry Letters* **2015**, 6 (24), 5016-5021.

47. Nitsche, C.; Mahawaththa, M. C.; Becker, W.; Huber, T.; Otting, G., Site-selective tagging of proteins by pnictogen-mediated self-assembly. *Chemical Communications* **2017**, 53 (79), 10894-10897.

48. Neylon, C.; Brown, S. E.; Kralicek, A. V.; Miles, C. S.; Love, C. A.; Dixon, N. E., 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* **2000**, 39 (39), 11989-11999.

49. Mahawaththa, M. C.; Pearce, B. J. G.; Szabo, M.; Graham, B.; Klein, C. D.; Nitsche, C.; Otting, G., Solution conformations of a linked construct of the Zika virus NS2B-NS3 protease. *Antiviral Research* **2017**, 142, 141-147.

50. Wu, P. S. C.; Ozawa, K.; Lim, S. P.; Vasudevan, S. G.; Dixon, N. E.; Otting, G., Cell-Free Transcription/Translation from PCR-Amplified DNA for High-Throughput NMR Studies. *Angewandte Chemie International Edition* **2007**, 46 (18), 3356-3358.

51. Apponyi, M. A.; Ozawa, K.; Dixon, N. E.; Otting, G., Cell-free protein synthesis for analysis by NMR spectroscopy. *Methods Mol Biol* **2008**, 426, 257-68.

52. Ozawa, K.; Loscha, K. V.; Kuppan, K. V.; Loh, C. T.; Dixon, N. E.; Otting, G., High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. *Biochemical and Biophysical Research Communications* **2012**, 418 (4), 652-656.

53. Goldfarb, D.; Lipkin, Y.; Potapov, A.; Gorodetsky, Y.; Epel, B.; Raitsimring, A. M.; Radoul, M.; Kaminker, I., HYSCORE and DEER with an upgraded 95GHz pulse EPR spectrometer. *Journal of Magnetic Resonance* **2008**, 194 (1), 8-15.

54. Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **2006**, 1 (6), 2856-2860.

55. Bahrenberg, T.; Yang, Y.; Goldfarb, D.; Feintuch, A., rDEER: A Modified DEER Sequence for Distance Measurements Using Shaped Pulses. *Magnetochemistry* **2019**, 5 (1), 20.

56. Bahrenberg, T.; Rosenski, Y.; Carmieli, R.; Zibzener, K.; Qi, M.; Frydman, V.; Godt, A.; Goldfarb, D.; Feintuch, A., Improved sensitivity for W-band Gd(III)-Gd(III) and nitroxide-nitroxide DEER measurements with shaped pulses. *Journal of Magnetic Resonance* **2017**, 283, 1-13.

57. Doll, A.; Qi, M.; Wili, N.; Pribitzer, S.; Godt, A.; Jeschke, G., Gd(III)-Gd(III) distance measurements with chirp pump pulses. *Journal of Magnetic Resonance* **2015**, 259, 153-162.

58. Jeschke, G.; Chechik, V.; Ionita, P.; Godt, A.; Zimmermann, H.; Banham, J.; Timmel, C. R.; Hilger, D.; Jung, H., DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **2006**, 30 (3), 473-498.

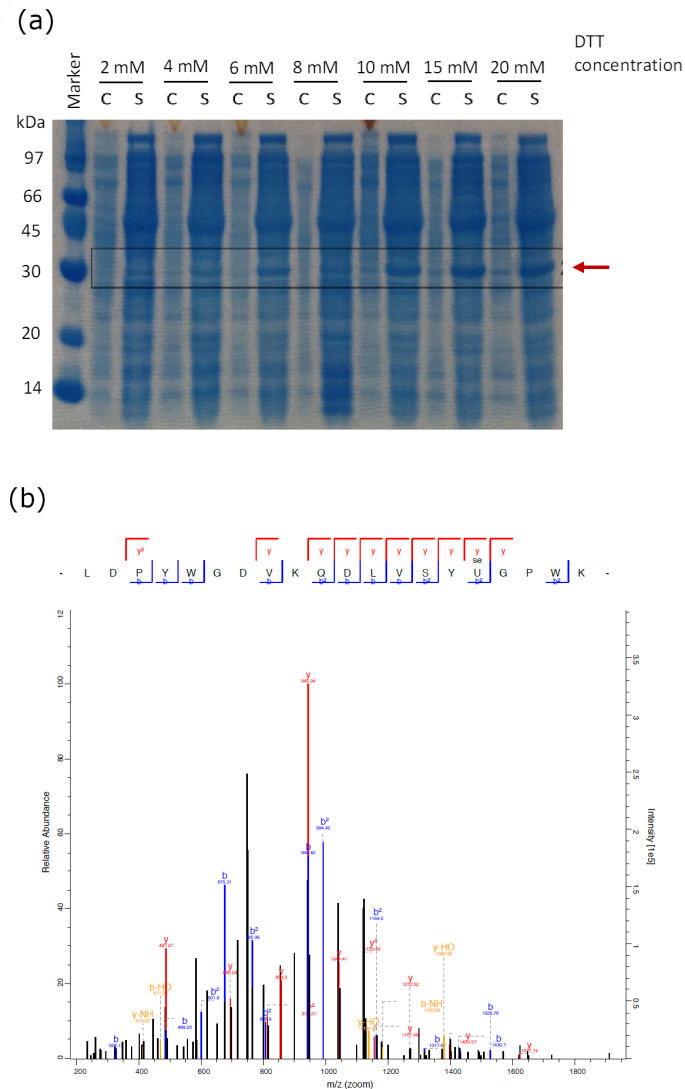


Figure 1. Optimization of DTT concentrations in CFPS of the C80U mutant of the ZiPro and confirmation of the incorporation of selenocysteine at position 80. (a) SDS-PAGE of seven CFPS reactions carried out in the presence of different concentrations of DTT (2 mM to 20 mM). The arrow identifies the band of ZiPro C80U. Crude samples (C) are from the inner reaction mixture after overnight CFPS and the corresponding supernatants (S) were obtained after spinning down the precipitates of the inner reaction mixtures. (b) LC-MS/MS analysis of ZiPro C80U, confirming selenocysteine incorporation in response to the cysteine codon at position 80. Annotated peaks in the LC-MS/MS spectrum correspond to the series of b ions (blue) and y ions (red) that resulted from fragmentation of the carbamidomethylated peptide

from Zipro C80U shown at the top. b- and y-ion fragmentation is also indicated in the amino acid sequence, where the position of the selenocysteine residue is marked by a star. The 3-charged (+3) ion with m/z of 825.706 was fragmented to produce the MS/MS spectrum.

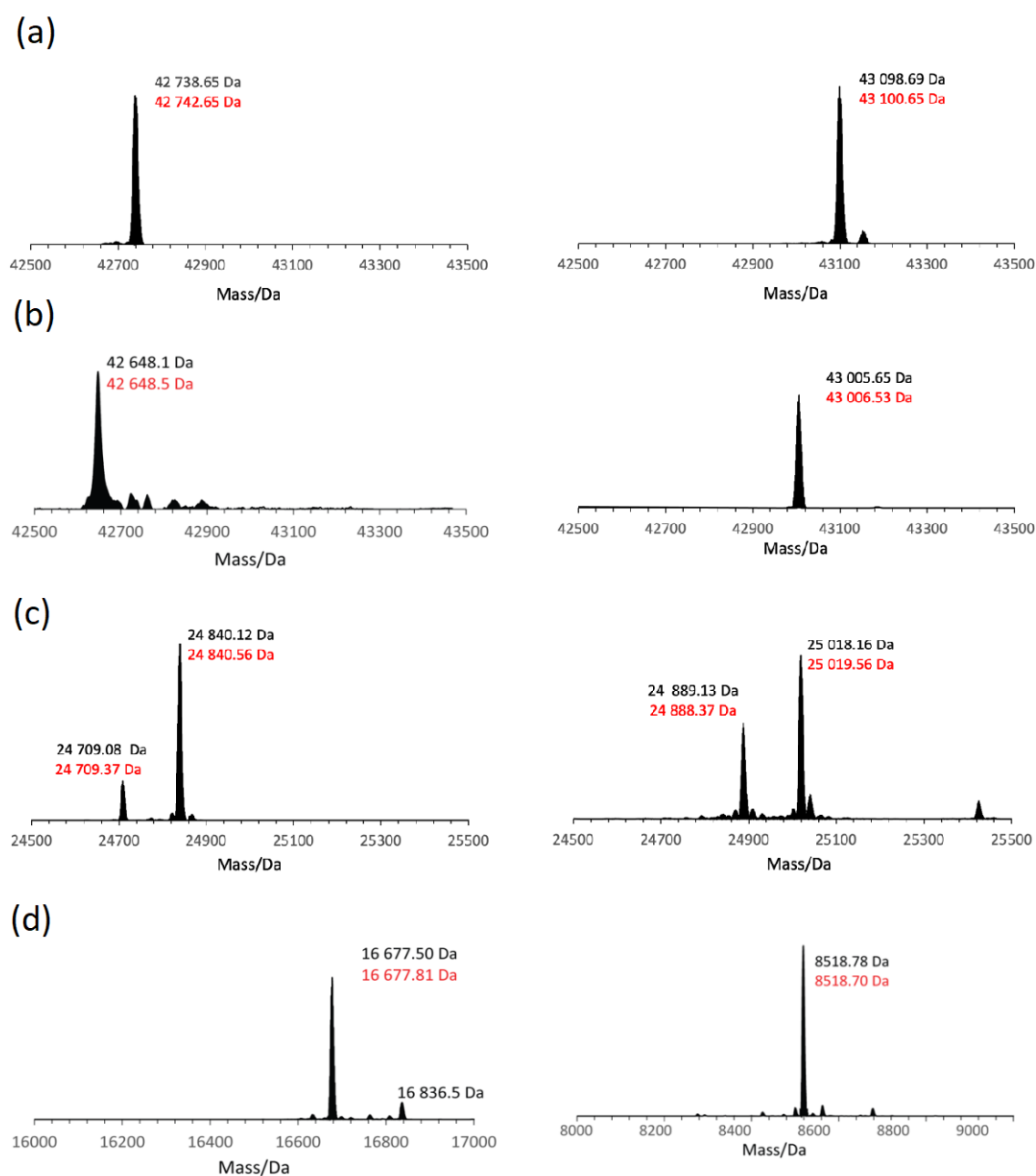


Figure 2. Labelling selenocysteine and cysteine with alkylating DPA tags (as shown in Figure 3). Observed and expected masses are indicated in black and red, respectively. The expected mass increase after tagging with two MDPA tags is 358 Da. The left and right panels show the mass spectra before and after tagging, respectively. (a) Mass spectra of MBP T237U/T345U before and after tagging with Cl-MDPA. The absence of untagged protein suggests complete tagging yields. (b) Mass spectra of MBP T237C/T345C before and after tagging with Br-MDPA. (c) Mass spectra of ZiPro V36U before and after tagging with Cl-MDPA. (d) Mass spectra of unlabeled GB1 Q32U and after labeling with Cl-MDPA.

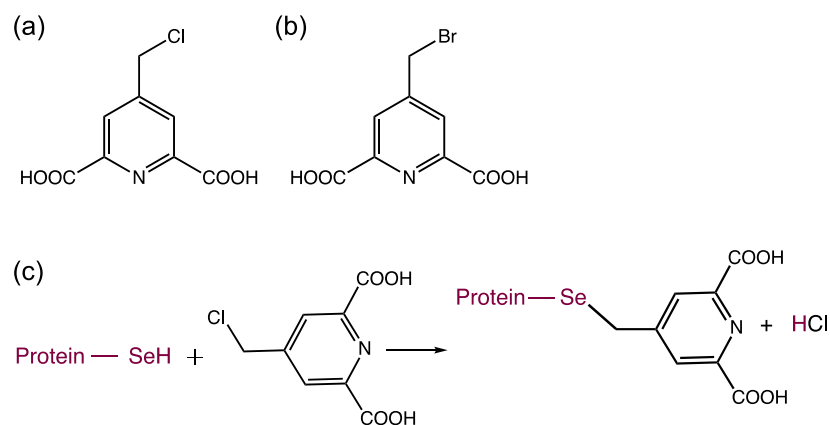


Figure 3. Tags and tagging reaction used. Chemical structure of (a) 4-chloromethylene-dipicolinic acid (4Cl-MDPA) and (b) 4-bromomethylene-dipicolinic acid (4Br-MDPA). (c) Reaction scheme between 4Cl-MDPA tag and a selenocysteine residue of a protein.

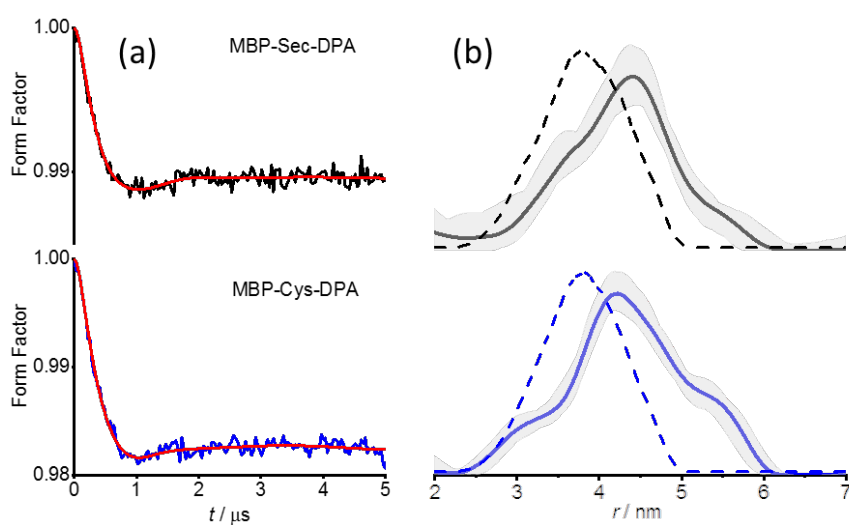


Figure 4. DEER results obtained with different tags on the MBP selenocysteine mutant (black) and the MBP cysteine mutant (blue). (a) Background-corrected DEER traces along with the fit obtained with the distance distributions shown on the right. The primary DEER traces are shown in Figure S3 of the Supporting Information. (b) Analyzed distance distributions including confidence ranges (in shaded gray) and comparison with modelled distances (in dashed lines).

