Site-directed labelling of proteins for NMR and EPR studies

Adarshi Priyanga Welegedara

Research School of Chemistry

August 2018

A thesis submitted for the degree of Doctor of Philosophy at

the Australian National University

Australian National University

DECLARATION OF ORIGINALITY

This thesis is composed of my original work and contains no material previously published or written by another person. Previously published data used in comaprisons and established methods have been acknowledged by citation of the original publications from which they derive. The content of my thesis is the result of work I have carried out since the commencement of my research degree candidature (Feb. 2015-Aug. 2018) and has not been previously submitted for another degree or diploma in any university or tertiary institution.

Adarshi Priyanga Welegedara

August 2018

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Professor Gottfried Otting for giving me the opportunity to work in his group and for the unwavering guidance and encouragement he has provided throughout my time as his PhD student. I have been lucky to have a supervisor who responded to my requests and questions so promptly and perfectly. I would also like to thank my co-supervisor Professor Thomas Huber for his inventive ideas and suggestions which were determinant for the success of the work presented in this thesis.

I also acknowledge our collaborators, Dr Bim Graham and his group at Monash University, Australia, Professor Xun-Cheng Su and his group at Nankai University, China for synthesizing tags and unnatural amino acids, Professor Daniella Goldfarb and her group at Weizmann Institute of Science, Israel for EPR measurements and Professor Richard Payne and his group at the University of Sydney, Australia for sharing their knowledge and experience on selenocysteine chemistry.

I would like to express my gratitude to past and present members of the Otting and Huber groups for their support and motivation. In particular, I would like to thank Dr Elwy Abdelkader for teaching me molecular biology techniques.

I would like to thank the friends I met in Canberra for making my stay in Canberra memorable, all my other friends for providing unwavering friendship and encouragement that I needed.

Finally, I would like to thank and extend my love to my parents, my sisters and their families for their unconditional love, support and continuous encouragement, without which I would have not come this far.

ABSTRACT

Site-specific protein labelling presents an important tool for protein structural biology by spectroscopic techniques. This thesis focuses on the development of new spectroscopic labels and labelling strategies to improve the sensitivity, accuracy and scope of NMR and EPR spectroscopy experiments of proteins.

Double electron–electron resonance (DEER) spectroscopy measures the distance between two paramagnetic metal centres introduced by site-specific attachment of suitable tag molecules. Good DEER tags possess rigid tethers to position their paramagnetic centres at a well-defined location relative to the protein and deliver narrow DEER distance distributions with high sensitivity, which can provide accurate information about protein flexibility. This thesis introduces new, cyclen-based Gd^{3+} tags and small, Gd^{3+} chelating tags designed to deliver narrow DEER distance distributions.

Chapter 2 describes the development of two cyclen-based double-arm Gd^{3+} tags designed for binding to the target protein at two and three points to obtain the narrowest Gd^{3+}–Gd^{3+} DEER distance distributions ever recorded with proteins. It also describes DEER distance measurements with the iminodiacetic acid tag attached to cysteine (Cys), where tags attached to two neighbouring Cys residues combine to chelate a single Gd^{3+} ion. These results have been published in a journal article (Welegedara et al., Chem. Eur. J. 2017, 23, 11694−11702).

Chapter 3 discusses two single-armed Gd^{3+} tags, a cyclen-based Gd^{3+} tag and a PyMTA tag that forms a heptadentate Gd^{3+} binding motif. Both tags deliver the shortest possible tethers to cysteine residues and are shown to produce narrow DEER distance distributions in proteins.

For applications in NMR spectroscopy, proteins can be labelled site-specifically with NMR probes such as trimethylsilyl (TMS) probes, which deliver readily detectable 1D ^{1}H-NMR signals. Introduction of paramagnetic tags and NMR probes by attachment to Cys residues requires mutations of native Cys residues to achieve site-selectivity, which is not possible with structurally and functionally important Cys residues. Chapter 4 demonstrates a solution to this limitation by introducing a selenocysteine (Sec) residue, which can be selectively reacted with probe molecules at slightly acidic pH without

affecting naturally occuring Cys residues. To achieve this, a Sec residue was introduced as a photocaged unnatural amino acid (UAA), PSc, to bypass the otherwise unavoidable challenges associated with the natural Sec incorporation mechanism. UV illumination of PSc yielded Sec with no evidence for the formation of undesired dehydroalanine byproducts. Selective tagging of Sec residues with TMS tags was shown to deliver a useful tool for studies of ligand binding to proteins. These results have also been published in a journal article (Welegedara et al., Bioconjugate Chem. 2018, 19, 2257−2264).

Site-selective incorporation of isotope-labelled PSc, photolysis and anaerobic deselenization opens an indirect route to labelling a single specific alanine residue in a protein with stable isotopes. Such samples would have important applications in heteronuclear NMR, as they allow the selective detection of the labelled alanine residue with maximal spectral resolution. As shown in Chapter 5, deselenization of selenoproteins into alanine is possible but requires extremely anaerobic conditions to eliminate serine as the main unwanted byproduct.

A range of UAAs has been developed to serve as spectroscopic probes or facilitate the introduction of spectroscopic probes via biorthogonal reactions. The increasing demand for proteins with different UAAs and the significant cost of some of these UAAs has led to increasing popularity of cell-free protein synthesis (CFPS) systems, which use amino acids more sparingly than *in vivo* expression systems. Mutants of pyrrolysyl-tRNA synthetase (PylRS)/$tRNA_{CUA}$ pairs have been developed into a particularly versatile tool for the incorporation of many structurally different UAAs, but most produce disappointedly poor protein yields *in vivo*. Chapter 6 describes attempts to develop an in-house CFPS system with a PylRS/$tRNA_{CUA}$ pair. In addition, the polyspecific G2 synthetase has been reported to facilitate the incorporation of sterically demanding UAAs and Chapter 7 of this thesis describes attempts to develop a CFPS system with the G2 synthetase.

LIST OF ABBREVIATIONS

3BrPy-MTA	3-bromo-4-(phenylsulfonyl)pyridin-2,6-diyl) bismethylenenitrilo tetrakis(acetic acid)
3MDPA	3-mercaptodipicolinic acid
4MMDPA	4-mercaptomethyldipicolinic acid
4PS-PyMTA	4-phenylsulfonyl-(pyridin-2,6-diyl)bismethylenenitrilo tetrakis(acetic acid)
aaRS	Aminoacyl-tRNA synthetase
Acd	Acridon-2-ylalanine
BibaF	4-(2′-bromoisobutyramido)phenylalanine
BocK	N^{ε} -(tert-butyloxycarbonyl)- lysine
CF_3-AcK	N^{ε}-trifluoroacetyl-lysine
CFPS	Cell-free protein synthesis
Ch-PylRS	Chimeric pyrrolysyl-tRNA synthetase
Cryo-EM	Cryo-electron microscopy
CSA	Chemical shift anisotropy
CTD	C-terminal domain
DEER	Double electron–electron resonance
Dha	*Desulfitobacterium hafniense*
Dha PylRSc	*Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase C-terminal domain
Dha PylRSn	*Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase N-terminal domain

DMNB	4,5-dimethoxy-2-nitrobenzyl
DNA	Deoxyribonucleic acid
DNP	Dynamic nuclear polarization
DO3A	1,4,7-tris(carboxymethylaza)cyclododecane-10-azaacetylamide
DO3MA	(1S,4S,7S)-α,α′,α″-trimethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid
DO3MA-3BrPy	(2R,2′R,2″R)-2,2′,2″-(10-(5-bromo-4-(phenylsulfonyl)pyridin-2-yl)methyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)tripropanoic acid
DOTA	1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid
DPA	Dipicolinic acid
DTNB	5,5-dithiobis-(2-nitrobenzoic acid)
DTT	1,4-dithiothreitol
E. coli	*Escherichia coli*
ECL	Expressed protein ligation
ED-EPR	Echo-detected electron paramagnetic resonance
EPR	Electron paramagnetic resonance
ESI-MS	Electrospray ionization mass spectrometry
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
HH ribozyme	Hammerhead ribozyme
HSQC	Heteronuclear single-quantum coherence
IDA	Iminodiacetic acid
IPTG	Isopropyl-β-D-thiogalactopyranoside
K_d	Dissociation constant

LB	Luria-Bertani
M. barkeri	*Methanosarcina barkeri*
M. mazei	*Methanosarcina mazei*
MAS ssNMR	Magic-angle spinning solid state nuclear magnetic resonance
MBP	Maltose binding protein
Mj	*Methanocaldococcus jannaschii*
MTS-ADO3A	(2,2′,2″-(10-{2-[2-(methylsulfonylthio)ethylamino]-2-oxoethyl}-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)triacetic acid)
MWCO	Molecular weight cut-off
NCL	Native chemical ligation
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effects
NTA	Nitrilotriacetic acid
NTD	N-terminal domain
o-tRNA	Orthogonal-tRNA
PCC	Photocaged cysteine
p-CNFRS	*p*-cyanophenylalanyl-tRNA synthetase
PDB	Protein data bank
PELDOR	Pulse electron–electron double-resonance
PpiB	Peptidyl-prolyl *cis-trans* isomerase
PRE	Paramagnetic relaxation enhancement
PCS	Pseudocontact shift
PSc	Photocaged selenocysteine
PylRS	Pyrrolysyl-tRNA synthetase
PyMTA	(Pyridin-2,6-diyl)bismethylenenitrilo tetrakis(acetic acid)

RIDME	Relaxation induced dipolar modulation enhancement
RNA	Ribonucleic acid
SDSL	Site-directed spin labelling
Sec	Selenocysteine
Cys	Cysteine
SECIS	Selenocysteine insertion sequence
SfGFP	Super-folded green fluorescent protein
SL	Spin label
T4L	T4 lysozyme
TAM	Triarylmethyl
TCEP	Tris(2-carboxyethyl)-phosphine
Tet-F	4-(6-methyl-S-tetrazine-3-yl)aminophenylalanine
TEV	Tobacco etch virus
UV	Ultra violet
TMS	Trimethylsilyl
TMSf	Trimethylsilyl phenylalanine
TyrRS	Tyrosyl-tRNA synthetae
PylRS	Pyrrolysyl-tRNA synthetase
UAA	Unnatural amino acid
XRD	X-ray diffraction
ZFS	Zero field splitting
ZiPro	Zika virus NS2B-NS3 protease

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1. General introduction

The function of a protein depends on its three-dimensional (3D) structure and conformational variability. Understanding 3D structures of proteins, their dynamics as well as how proteins and their complexes interact with ligands and other macromolecules is therefore important in elucidating their function and finds numerous applications in medicinal chemistry and in bioengineering, for instance in drug discovery and improving the activity of enzymes. This has triggered the development of many sophisticated techniques to study 3D structures of proteins, their dynamics and interactions. These techniques differ from each other mainly in terms of their applicability and accuracy of the results they deliver.

1.1 High-resolution protein characterization techniques in structural biology

Protein crystallography was the first technique developed and is by far the most widely used technique for 3D structure determination of proteins at atomic resolution. As X-ray diffraction (XRD) methods are performed on protein single crystals, structures determined by protein crystallography may not represent the structures under physiological conditions and provide only limited information on protein dynamics (Shi, 2014). Very different conformations of the protein can only be accessed by altering crystallization conditions or inclusion of ligands during crystallization (Shi, 2014). In general, the technique is not limited by the size of the protein, but by the requirement of protein single crystals, which are laborious to obtain or cannot be obtained at all.

Later, nuclear magnetic resonance (NMR) techniques emerged, which enabled structure determinations of proteins in solution (Wüthrich, 1989; Wüthrich, 1990). In contrast to the XRD methods, protein NMR can be performed in solution under near-physiological conditions (Wüthrich, 1989) and enables structure determination of proteins that are difficult to crystallize. Unfortunately, both the number of NMR resonances and increasing line widths affect the resolution of NMR spectra of larger proteins, making 3D structure determinations of proteins with a molecular mass greater

than about 40 kDa very difficult. Nevertheless, over the years, many examples have been reported, where solution NMR has successfully been applied not only to study 3D structures but also dynamic interactions. Recently, in-cell and on-cell NMR techniques have also been developed for protein characterization inside and on the surface of living cells, respectively (Freedberg and Selenko, 2014).

When poor solubility, unavailability of crystals or protein size limit the protein structure determination by XRD or solution NMR, magic-angle spinning solid state NMR (MAS ssNMR) offers an alternative platform for 3D structure determinations of proteins at atomic resolution and analysis of their dynamics both *in vitro* and in the in-cell environment (Freedberg and Selenko, 2014; McDermott, 2004; Thompson, 2002). Uniquely, ssNMR experiments can be conducted with sensitivity enhancements by dynamic nuclear polarization (DNP), where the large polarization of unpaired electrons is transferred to nuclear spins through microwave irradiation at the electron paramagnetic resonance (EPR) frequency. DNP has emerged as a powerful tool to enhance the sensitivity in MAS ssNMR spectroscopy (Sze et al., 2012).

The most prominent advance recently has been in cryo-electron microscopy (cryo-EM), which can nowadays determine high-resolution 3D structures of proteins and, more interestingly, large macromolecular complexes of proteins (Bai et al., 2015; Nogales and Scheres, 2015). The technique does not require crystalline samples and can be performed with significantly smaller amounts of sample (0.1 mg) and fewer restrictions on protein purity (Bai et al., 2015). Cryo-EM structures are more informative than conventional XRD structures as they also carry information on different conformations that the protein can adopt.

1.2 Studying protein-ligand interactions

Proteins in the cellular milieu interact with a variety of ligands varying from metal ions and small molecules to biological macromolecules such as DNA, RNA, peptides and other proteins. Characterizing the ligand binding sites on the protein, conformation of the bound ligand, covalent and non-covalent interactions between the bound ligand and the protein and binding affinities of the ligand is of significant importance in drug discovery.

Even though much of this information can be obtained by XRD data, co-crystallization of ligands with proteins is not always successful and crystal structures may not resemble physiological structures. Alternatively, as summarized in many reviews, sophisticated solution NMR techniques have been developed to study these dynamic

interactions of proteins under physiological conditions. Particularly, chemical shift perturbations, nuclear Overhauser effects (NOE) and paramagnetic effects such as paramagnetic relaxation enhancements (PRE) or pseudocontact shifts (PCS) have been used extensively to obtain structural restraints of ligand-protein complexes (Becker et al., 2018b; McCoy and Wyss, 2002; Nitsche and Otting, 2018; Pintacuda et al. 2007; Williamson, 2013; Zuiderweg, 2002) and measure ligand affinities to proteins (Becker et al., 2018b; Gossert and Jahnke, 2016; Williamson, 2013).

Proteins undergo conformational changes upon ligand binding. Studying these conformational changes is as important as studying the 3D structures of the free proteins as they reveal important information about function. In addition to NMR spectroscopy, several other spectroscopic techniques have been established to study large conformational changes in proteins. In a straightforward approach, these conformational changes can be followed by measuring the distance between two sites of the protein. Förster resonance energy transfer (FRET) and double electron-electron resonance (DEER) allow distance measurements between two spectroscopic probes installed on the protein and are two widely used techniques to study conformational changes of proteins. Importantly, both FRET and DEER experiments can be carried out *in vivo* and *in vitro* under physiological conditions.

FRET measures the distance between a pair of non-identical fluorescence probes that are positioned on the protein at a distance near the Förster radius (Piston and Kremers, 2007). The technique has been extensively used for measuring conformational dynamics of proteins (Weiss, 2000). Although it allows probing large conformational changes of proteins with single-molecule sensitivity, it is difficult to measure distances and distance changes accurately (Chao et al., 2008; Piston and Kremers, 2007). In addition, the fluorophores need to be carefully chosen based on the expected distance (Chao et al., 2008; Piston and Kremers, 2007) and simultaneous introduction of two types of fluorophores into the same protein can be challenging. In addition, the large size of fluorescent probes compromises the distance measurements.

In contrast, DEER measurements are based on EPR. DEER experiments measure the distance between two identical paramagnetic centres installed on a biological macromolecule. Compared with FRET, the result is more accurate and suitable for probing conformational variations (Grohmann et al., 2010; Jeschke, 2012).

Many important biological systems have been studied with the help of these techniques, not only to probe conformational changes of proteins upon ligand binding but also to assess subunit assembly and flexibility of parts of the biological macromolecules.

DEER and FRET measurements were compared in a study of conformational flexibility of archaeal RNA polymerase subunit F/E, which is a heterodimer that binds nascent RNA stimulating the processivity of elongation complexes (Grohmann et al., 2010). Notably, the DEER distance measurements showed an excellent agreement with the distances calculated based on the crystal structure of the protein, while the FRET distance measurements showed significant discrepancies between the measured and expected values. The inaccuracy of FRET measurements was attributed to the choice of fluorescent probes, illustrating the need to carefully select the fluorescent probes depending on the expected distance.

Although DEER or FRET techniques do not allow 3D structure determination of proteins at atomic resolution, they nonetheless enable validation of 3D XRD structures of proteins under more physiological conditions and even in in-cell environments.

Among the many applications of solution NMR spectroscopy and DEER experiments in structural biology, the work of the present thesis focused on solution NMR to study ligand binding affinities and DEER as a tool to study large conformational changes of proteins.

1.3 Incorporation of unnatural amino acids into proteins in *Escherichia coli*

Genetic encoding of unnatural amino acids (UAA) offers potential to alter activity and function of proteins and allows site-specific introduction of posttranslational modifications. In addition, UAAs can be incorporated to act as spectroscopic probes or to aid the introduction of such spectroscopic probes via biorthogonal reactions (Dumas et al., 2015; Lang and Chin, 2014; Liu and Schultz, 2010).

Genetically encoded incorporation of UAAs requires aminoacyl-tRNA synthetase (aaRS)/tRNA pairs that are orthogonal to the aaRS and tRNA molecules in the host organism. Orthogonality is most easily achieved by transferring aaRS/tRNA pairs from another kingdom of life into the host organism. UAA incorporation strategies include quadruplet codon decoding (Anderson et al., 2004; Neumann et al., 2010; Neumann et al., 2014; Wang et al., 2014) and reassignment of sense codons to UAAs (Link and Tirrell 2005; Mukai et al., 2015), but stop codon suppression with UAAs is most common. The amber stop codon is considered to be most suitable as it is the least frequent stop codon (Alff-Sternberger and Epstein, 1994; Wan et al., 2010; Wang et al., 2001).

The tyrosyl-RS (TyrRS)/tRNA pair from the archaebacterium *Methanocaldococcus jannaschii (Mj)* and the pyrrolysyl-RS (PylRS)/tRNA pair from archaeal methanogens such as *Methanosarcina barkeri* (*M. barkeri*), *Methanosarcina mazei* (*M. mazei*) and the gram positive bacterium *Desulfitobacterium hafniense* (*Dha*) are some of the most prominent orthogonal aaRS/tRNA pairs reported for use in *Escherichia coli* (*E. coli*) expression systems (Blight et al., 2004; Nozawa et al., 2009; Wang et al., 2000).

To incorporate an UAA into a protein, the active site of the orthogonal aaRS needs to specifically bind the UAA in a proper location and orientation for subsequent aminoacylation of its cognate tRNA, which must also be orthogonal with regard to the host translation machinery. Orthogonality means that the aaRS must not bind any of the canonical amino acids in its amino-acid binding pocket. Therefore, wild-type *Mj* TyrRS requires mutations in its active site to enable UAA binding and eliminate the binding of the canonical amino acid, tyrosine. As *Mj* TyrRS does not recognize the anticodon in great detail (Steer and Schimmel, 1999), it has been possible to mutate the anticodon loop of *Mj*-$tRNA^{Tyr}$ for improved amber or quadruplet codon suppression without compromising its affinity towards the *Mj* TyrRS. Furthermore, *Mj*-$tRNA^{Tyr}$ has been engineered to reduce any inherent affinity towards *E. coli* synthetases (Wang et al., 2000; Wang et al., 2001). It is also advantageous that *Mj* TyrRS does not possess an editing mechanism to deacylate *Mj*-$tRNA^{Tyr}$ following loading with the UAA (Wang et al., 2006). All these features make the *Mj* TyrRS/$tRNA_{CUA}$ pair attractive for genetic encoding of UAA incorporation.

In contrast, wild-type PylRS does not bind any of the canonical 20 amino acids in its active site and therefore requires no mutations to avoid competition with the natural amino acids. Furthermore, it is associated with its own natural amber suppressor tRNA (Polycarpo et al., 2004). While the natural function of wild-type PylRS is to encode pyrrolysine (the 22^{nd} natural amino acid), it accepts many other UAAs with diverse functional groups and its binding pocket has been further engineered to incorporate many more UAAs that are significantly different from its natural substrate, pyrrolysine (Dumas et al., 2015; Liu and Schultz, 2010). To incorporate novel UAAs with bulky functional groups, the larger amino acid binding pocket of PylRS has proven to be more suitable than the re-engineering of the smaller amino-acid binding pocket of *Mj* TyrRS. Moreover, the archaeal PylRS is orthogonal both in prokaryotic (*E. coli*) and eukaryotic cells (e.g. yeast, *Drosophila melanogaster* and mammalian cells). This provides an additional

advantage to the PylRS/tRNA pair for genetic code expansion (Bianco et al., 2012; Hancock et al., 2010; Mukai et al., 2008; Srinivasan et al., 2002).

The wide interest in incorporating UAAs with diverse functional groups triggered the evolution of many synthetase mutants capable of charging their cognate tRNAs with specific UAAs. Evolution of new functional synthetase mutants, however, is laborious. It is thus of great interest that polyspecific aaRS enzymes have been found that are capable of charging a number of structurally similar UAAs while still discriminating against natural amino acids. In particular, *p*-cyanophenylalanyl-tRNA synthetase (*p*-CNFRS), G2 synthetase and AcKRS are aaRSs that have been demonstrated to be polyspecific (Chin et al., 2002; Cooley et al., 2014; Guo et al., 2014).

Today, hundreds of different UAAs have been incorporated into proteins using these aaRS/tRNA pairs and their mutants. The present work used UAAs that act as spectroscopic probes by themselves or support posttranslational bioorthogonal conjugation with spectroscopic probes.

2. Nuclear magnetic resonance spectroscopy to study ligand binding affinities

The strength of interaction between a protein and a ligand can quantitatively be described by the dissociation constant K_d. For a reversible interaction between a protein (P) and a ligand (L),

$$P + L \rightleftharpoons PL \qquad (1)$$

with the ligand having a single binding site on the protein, the dissociation constant is defined by

$$K_d = \frac{k_{off}}{k_{on}} = \frac{[P][L]}{[PL]} \qquad (2)$$

where [P], [L] and [PL] denote the concentrations of free protein, free ligand and protein-ligand complex, respectively, and k_{on} and k_{off} are the rate constants of the forward and backward reactions, respectively.

2.1 2D HSQC spectra to study ligand binding affinities

Titration of a ligand into a protein solution results in chemical shift perturbations of the protein. If the protein is isotope-labelled, the changes in chemical shifts can readily be measured in a 2D heteronuclear single-quantum coherence (HSQC) spectrum. This is termed chemical shift mapping and increasing changes in chemical shifts with increasing

concentration of ligand allows calculating the dissociation constant (Becker et al., 2018b; Gossert and Jahnke, 2016; Williamson, 2013). In addition, the chemical shift changes highlight the ligand binding site on the protein, if the protein resonance assignments are available.

Numerous studies have been published, where chemical shift mapping was used to monitor ligand binding and ligand binding sites. For example, peptide binding to the C-terminal SH3 (C-SH3) domain of the *Caenorhabditis elegans* protein Sem-5 elicits chemical shift perturbations for many of the ^{15}N-HSQC cross-peaks upon addition of substrate peptide, but not for all (Figure 1; Ferreon and Hilser, 2003). Comparison with the crystal structures with and without the peptide confirmed that the cross-peaks with the largest chemical shift changes are located near the peptide binding site of the protein (Figure 2).

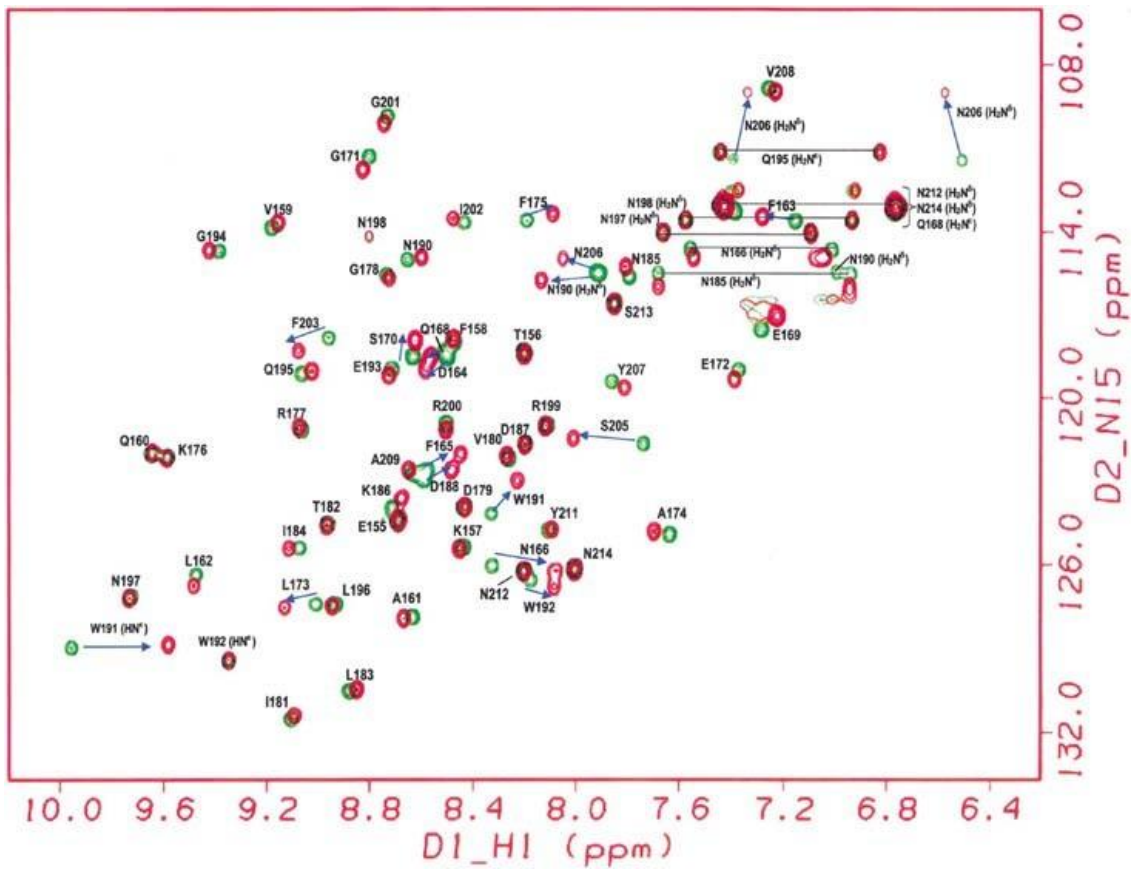


Figure 1. Assigned ^{1}H-^{15}N HSQC spectra of Sem-5 C-SH3 without peptide (green) and with peptide (red). Chemical shift perturbations are shown by blue arrows. Many resonances are not affected, but significant chemical shift changes are observed for residues near the peptide binding site (reproduced from Ferreon and Hilser, 2003).

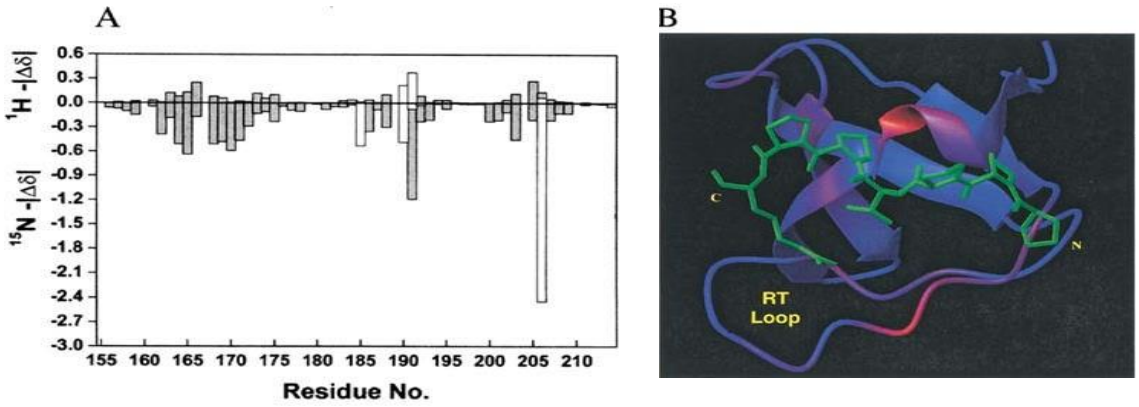


Figure 2. Mapping chemical shift changes onto a protein structure. (A) Changes of ^{1}H (top graph) and ^{15}N (bottom) chemical shifts of backbone amides following addition of peptide substrate plotted versus the amino-acid sequence. (B) Color mapping of the ^{1}H chemical shift changes onto the protein structure with the peptide (green) overlaid in the position as seen in the X-ray structure (reproduced from Ferreon and Hilser, 2003).

Chemical shift changes present the most basic information that can be derived from studies of protein-ligand interactions by solution NMR. More advanced experiments are required to extract more detailed information, such as the conformation of the bound ligand and its binding mode (Nitsche and Otting, 2018; Pintacuda et al., 2007; Zuiderweg, 2002). Even in their most basic form, however, the chemical shift changes observed in 2D HSQC spectra of the protein upon titration with ligand can be used to determine the ligand binding affinity. Two main scenarios must be distinguished. The slow exchange regime prevails, if the ligand binds tightly to the protein and k_{off} is slower than the difference in frequency between the bound and free state of the protein. In this case, titration of the ligand into protein results in gradual disappearance of the free protein signal and concomitant appearance of a new signal of the ligand-bound protein (Figure 3). In this situation, the intensities of the signals from the free and bound state are directly proportional to the concentrations of the respective species and K_d can be determined from equation 3 (Gossert and Jahnke, 2016),

$$\Delta I = \Delta I_{max} \frac{([L] + [P] + K_d) - \sqrt{([L] + [P] + K_d)^2 - 4[L][P]}}{2[P]} \qquad (3)$$

where ΔI is the observed intensity difference, ΔI_{max} the maximal intensity difference at saturation and [L] and [P] are the total concentrations of ligand and protein, respectively.

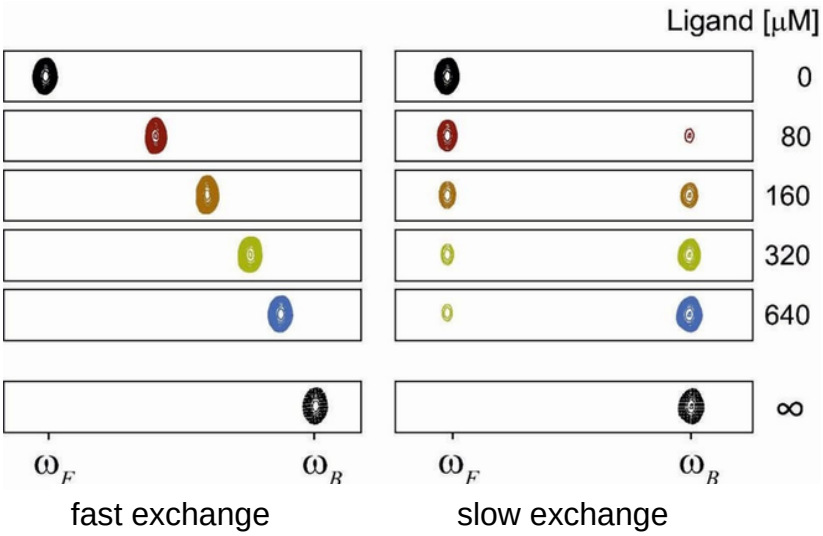


Figure 3. Schematic diagram showing changes of an HSQC cross-peak upon titration with a ligand. For weak binding in fast exchange, the signal of the free-protein peak at ω_F gradually moves towards the fully ligand-saturated protein peak at ω_B. For tight binding in slow exchange, only the relative signal intensities of free and bound protein peaks change (reproduced from Gossert and Jahnke, 2016).

In contrast, in the fast exchange regime the ligand is bound only weakly to the protein, k_{off} is faster than the frequency difference between the protein signals in the bound and free states, and titration of the ligand into the protein solution results in gradual shifts of the free-protein signal to the position of the ligand-bound protein (Figure 3). In this case, the frequency of the signal at any titration point represents the weighted average of the free-protein and bound-protein shifts and K_d can be determined using equation 4 (Gossert and Jahnke, 2016),

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

where 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. In an analogous way, K_d values can be obtained by ligand-observed methods, where the chemical shift changes in the ligand are monitored upon addition of

protein. Ligand-based methods are advantageous, when the ligand has a weak binding affinity to the protein, permitting use of excess ligand.

More complicated situations arise, when a ligand binds to multiple sites on the protein with different affinities. Such cases are manifested by changes in ^{15}N-HSQC cross-peak positions that do not follow a straight line with increasing ligand concentration.

2.2 NMR probes to measure ligand binding affinities

Isotope-labelled proteins allow the observation of well-resolved cross-peaks in HSQC spectra, but the production of suitably isotope-labelled samples can be expensive. Therefore, 1D ^{19}F NMR has been introduced as an alternative to monitor fluorine-substituted tryptophan, *p*-trifluoromethoxy-L-phenylalanine, trifluoromethyl-L-phenylalanine or 2,2,2-trifluoroethanethiol ligated to a cysteine residue in the protein (Cellitti et al., 2008; Jackson et al., 2007; Liu et al., 2012; Loscha et al., 2012). This introduces a new challenge, namely the introduction of fluorinated amino acids into proteins. Natural proteins do not contain fluorine, which greatly simplifies the ^{19}F-NMR spectra of selectively labelled proteins. The spectral resolution gained in this way is important, as ^{19}F-relaxation is governed by chemical shift anisotropy (CSA), which broadens the ^{19}F-NMR signals. Therefore, to minimize the effects arising from CSA, ^{19}F-NMR spectra are often recorded at low magnetic fields, but this compromises the sensitivity of the measurements.

Site-specific introduction of a chemical probe with a readily detectable NMR signal has been identified as an outstanding strategy to study ligand binding to protein. In particular, the *tert*-butyl and trimethylsilyl (TMS) groups both have nine degenerate methyl protons. Due to the rotations of the methyl groups, both of these functional groups result in intense and narrow singlets in the 1D ^{1}H-NMR spectrum, which can be detected without assigning the entire spectrum (Becker et al., 2018a; Chen et al., 2015; Loh et al., 2018). These *tert*-butyl and TMS functional groups have been introduced into proteins with help of UAAs and chemical probes.

O-*tert*-butyl-tyrosine (Figure 4A) can be incorporated into proteins by *p*-CNFRS evolved from the *Mj* TyrRS and its cognate tRNA$_{CUA}$ that recognizes the TAG stop codon (Chin et al., 2002). The O-*tert*-butyl group of this amino acid produces a narrow and intense single peak in the 1D ^{1}H-NMR spectrum between about 1.0 and 1.5 ppm (Chen et al., 2015). This intense peak has been identified in a protein hexamer with a molecular

weight larger than 300 kDa, although background resonances from the protein can potentially mask this signal. Nevertheless, the *tert*-butyl group has successfully been used in proteins not only to measure ligand binding affinities but also to observe NOEs between protein and ligands (Chen et al., 2015; Jabar et al., 2017). *tert*-butyl phenylalanine (Figure 4B) similarly results in an intense and narrow ^{1}H-NMR signal comparable to the signal produced by *O-tert*-butyl tyrosine (Loh et al., 2018).

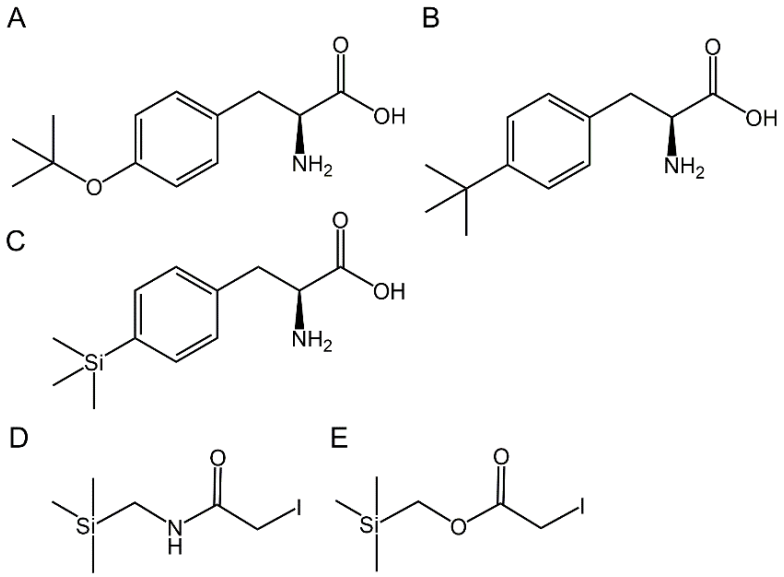


Figure 4. Chemical structures of O-tert-butyl and TMS-based NMR probes used in studying protein-ligand interactions. Structure of A) O-tert-butyl tyrosine, B) tert-butyl phenylalanine, C) 4-(trimethylsilyl) phenylalanine, D) an amide TMS tag and E) an ester TMS tag.

Recently, Loh et al. (2018) succeeded in incorporating 4-(trimethylsilyl)-phenylalanine (TMSf) (Figure 4C) into proteins using the same *p*-CNFRS/tRNA$_{CUA}$ pair to produce a narrow and intense signal. Importantly, the TMS resonance appears near 0 ppm and therefore is not easily masked by background signals from the protein. This offers unambiguous detection of the intense singlet arising from nine degenerate ^{1}H spins of the methyl groups attached to silicon. Unfortunately, the signal proved to be sensitive to extreme line broadening upon titration with ligands, which may be attributed to the rigidity of the TMS group attached to phenylalanine. Moreover, the *p*-CNFRS/tRNA$_{CUA}$ pair failed to give good yields of TMSf incorporation.

Consequently, alkylating TMS tags (Figure 4D, 4E) were introduced, first to measure intermolecular NOEs (Jabar et al., 2017) and, subsequently, to measure ligand

binding affinities (Becker et al., 2018a). These alkylating TMS tags can easily be ligated to solvent-exposed cysteine residues of the protein and, due to a more flexible attachment of the TMS group, produce narrow and intense singlets near 0 ppm. The TMS signal permits accurate ligand binding measurements especially in the slow exchange regime, where the intensities of the TMS peaks need to be integrated to calculate the K_d value from fits using equation 3. Accurate integration of the ^{1}H-NMR signal of a *tert*-butyl group would be much more difficult due to the presence of overlapping NMR signals. Becker et al. (2018a) demonstrated the use of the ester TMS tag of Figure 4E installed near the substrate binding sites of the Zika virus NS2B-NS3 protease (ZiPro) and human prolyl isomerase FK506 binding protein to measure K_d values of their respective inhibitors.

A major limitation of the alkylating TMS tags arises when the protein of interest contains native solvent-exposed cysteine residues which need to be mutated out before site-directed protein labeling with the TMS tags. Therefore, biorthogonal reactions that are not limited by the presence of native cysteine residues still need to be developed to expand the scope of alkylating TMS tags in studying ligand binding affinities to proteins, especially when the native cysteine residues of the protein are essential to maintain its structure and function.

The spectral resolution can also be improved by site-specific introduction of isotopically labelled amino acids. For instance, incorporation of ^{13}C-labelled *O-tert*-butyl-tyrosine enables selective detection of the signal of the *tert*-butyl group in a ^{13}C-HSQC spectrum (Jabar et al., 2017). While this helps to detect the *tert*-butyl resonance in a background of other ^{1}H-NMR resonances, the chemical synthesis of ^{13}C-labelled *O-tert*-butyl tyrosine is expensive. Site-selective incorporation of isotope-labelled natural amino acids has also been reported. To distinguish from unlabelled amino acids of the same type, the isotope labelled amino acid can be synthesized with a photo-sensitive protection group for site-specific incorporation as a UAA, releasing the labelled natural amino acid by photolysis. This approach has been demonstrated for the incorporation of ^{15}N-labelled tyrosine (Cellitti et al., 2008). In principal, the strategy can be applied to achieve site-specific incorporation of any of the 20 canonical amino acids in isotope-labelled form. Notably, however, suitable synthetase mutants for the requisite photocaged amino acids are usually not available, limiting the approach to the incorporation of photocaged tyrosine, serine, cysteine and lysine (Deiters et al., 2006; Gautier et al., 2010; Lemke et al., 2007; Nguyen et al., 2014; Wu et al., 2004).

The present work introduces a photocaged selenocysteine UAA, deprotection of which yields selenocysteine (Sec). Given that Sec can be deselenized into alanine and serine (Dery et al., 2015), isotope labelled photocaged Sec incorporation thus in principle enables site-selective incorporation of isotope-labelled alanine.

3. Double electron-electron resonance to measure nanometre distances in proteins

As discussed above, distance measurements on the nanometre scale are a very useful tool for studying large conformational changes of proteins and other biological macromolecules. For instance, structural changes of proteins resulting from interactions with ligands, proteins or DNA can be followed by measuring distances between two carefully selected sites of the protein. In addition, changes in structural flexibility can be assessed by the widths of the distance distributions obtained and this can be equally important as measuring average distances.

DEER, also called pulse electron-electron double-resonance (PELDOR), is an EPR technique that delivers distance distributions between two paramagnetic centres that are usually provided by chemical spin labels (SL) site-specifically attached to a biological macromolecule (Borbat and Freed, 2013). The primary parameter measured by these experiments is the dipolar interaction frequency, ω_{dd}, between the two paramagnetic centres. Fundamentally, the dipolar interaction between the two SLs A and B depends on the distance r_{AB} as shown in equation 5,

$$\omega_{\mathrm{dd}} = \frac{\mu_0}{4\pi\hbar}\frac{g_A g_B \mu_B^2}{r_{AB}^3}\,(3\,cos^2\theta - 1) \qquad (5)$$

where μ_0 is the permeability of vacuum, $\hbar$ is Planck's constant divided by 2π, g_A and g_B are the gyromagnetic ratios of spins A and B, respectively, μ_B is the Bohr magneton and θ is the angle between the distance vector r_{AB} and the external magnetic field. The technique works with frozen solutions, where all molecules are randomly oriented, yielding a powder pattern from which the ω_{dd} value can be derived. In practice, the data are evaluated by fitting a distribution of distances and angles to derive the final distance distribution. The technique typically measures distances in the range of 2–8 nm (Feintuch et al., 2015) and thus is particularly suitable to study large conformational changes of proteins.

It is an important feature of DEER experiments that the distance distributions obtained are quantitative. Although the DEER measurements are carried out in frozen solutions where large-amplitude protein dynamics do not exist, different protein conformations trapped in the frozen solution contribute to the width of the distance distribution and in this way deliver information on the structural variability of the spin-labelled protein. In principle, DEER measurements are also possible at room temperature, where protein dynamics are conserved, but the very short electron relaxation times of most SLs under these conditions present a serious limitation. Nonetheless, a few SLs with longer relaxation times have been shown to allow DEER measurements at room temperature (Meyer et al., 2015; Shevelev et al., 2014; Yang et al., 2012a).

Proteins containing intrinsic paramagnetic centres such as paramagnetic transition metal ions (Astashkin et al., 1998; Ezhevskaya et al., 2013; Abdullin et al., 2015; Kaminker et al., 2015) or radicals (Bennati et al., 2003; Denysenkov et al., 2006) can readily be used in DEER distance measurements without further modifications. In general, however, paramagnetic centres must be introduced artificially. Site-specific introduction of such centres has emerged as a powerful method to obtain structural information in otherwise diamagnetic biomolecules (Hubbel el al., 2000).

3.1 Paramagnetic tags used in DEER experiments

As most natural proteins are diamagnetic, a key step for DEER measurements is the introduction of paramagnetic centres into the protein. This is achieved by site-directed spin labelling (SDSL). Nitroxide radicals were the first SLs developed (Stone et al., 1965) and have widely been used to study conformational changes of proteins (Hubbell et al., 2000). For DEER distance measurements in bacterial and eukaryotic cells, however, nitroxide SLs are less suitable because of chemical reduction of the nitroxide radicals into EPR-silent hydroxylamines by the reducing environment of the cytoplasm. Better protected nitroxide SLs have also been introduced, which minimize the susceptibility of the radical towards chemical reduction (Haugland et al., 2018). In addition, orientation selection effects at frequencies above X-band frequency (~9.5 GHz) complicate the interpretation of DEER measurements, if the orientation of the nitroxides relative to the protein is in any way conformationally restrained. To overcome these limitations, SLs have been developed, which are based on paramagnetic metal ions (Gd^{3+}, Mn^{2+} and Cu^{2+}) that are resistant towards reduction (Banerjee et al., 2012; Cunningham et al., 2015b; Potapov et al., 2010). Gd^{3+}-based SLs have become particularly attractive candidates for

DEER measurements as they allow DEER measurements at W-band (approximately 95 GHz) in the absence of orientation selection effects, which would otherwise complicate the analysis of DEER data. W-band measurements increase sensitivity and therefore measurements can be carried out with only small amounts (>0.09 nmols) of spin-labelled samples (Polyhach et al., 2007; Potapov et al., 2010). Triarylmethyl (TAM)-based SLs attached to two cysteine residues have also been used for distance measurements (Yang et al., 2012b), but these tags are very large in size.

Distance measurements between two different types of SLs have also been performed and found to offer potentially increased sensitivity compared to the use of identical SLs. Examples include Gd^{3+}-nitroxide (Garbuio et al., 2013; Lueders et al. 2011; Kaminker et al., 2012; Kaminker et al., 2013; Wu et al., 2016), Mn^{2+}-nitroxide (Akhmetzyanov et al., 2015; Kaminker et al., 2015; Meyer et al., 2016; Wu et al., 2016), Gd^{3+}-Mn^{2+} (Wu et al., 2016) and trityl-nitroxide (Joseph et al., 2016; Shevelev et al., 2015; Wu et al., 2016) distance measurements.

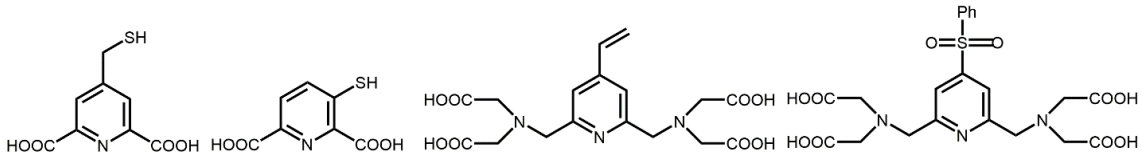
Typical SLs based on a paramagnetic metal are bi-functional, containing a metal ion binding site and a reactive group for ligation with the protein. If the tether between protein backbone and metal complex is flexible, this broadens the width of the distance distribution and compromises its interpretation in terms of the structural variability of the protein. Nonetheless, the tag-specific contribution can accurately be predicted by modelling the distance distribution of the tag based on the crystal structure of the protein and all possible conformations of the tag that avoid clashes with the protein (Abdelkader et al. 2015b; Kazmier et al., 2014a; 2014b).

In order to improve the structural resolution of the DEER data, the motional freedom of the SL needs to be restricted. Similar SLs with different paramagnetic metal ions are commonly used in paramagnetic NMR to generate PCSs and PREs for protein structure analysis. Therefore, rigid metal tags are of interest not only for DEER experiments, but also for other biophysical methods, including paramagnetic NMR and FRET. There is a continuous quest for new tags and labelling strategies to deliver maximal rigidity of the tag with respect to the protein it is bound to.

3.1.1 Gd^{3+} spin labels for DEER distance measurements

Early attempts to label proteins targeted cysteine residues for conjugation with 4-mercaptomethyldipicolinic acid (4MMDPA) or 3-mercaptodipicolinic acid (3MDPA) (Man et al., 2010; Su et al., 2008) (Figure 5) and this was successfully used for DEER

experiments with Gd^{3+} (Potapov et al., 2010). DPA forms a tridentate lanthanide binding motif with a K_d value in the low nanomolar range (Grenthe, 1980). These tags, however, required the thiol groups of the protein to be activated with Ellman's reagent (5,5-dithiobis-(2-nitrobenzoic acid), DTNB) before adding the DPA compounds to the protein. To overcome this cumbersome reaction step and obtain a stable ligation product suitable for in-cell DEER experiments, Qi et al. (2014) introduced 4-vinyl-PyMTA (Figure 5). There, the vinyl group reacts spontaneously by Michael addition with the thiol group of cysteine residues to form a stable thioether bond. Furthermore, the PyMTA moiety enables heptadentate coordination of the Gd^{3+} ion, which produces a more stable complex than the DPA moiety. As a drawback, the 4-vinyl-PyMTA tag generates a longer and more flexible tether between the protein backbone and the paramagnetic metal centre. In the most recent demonstration by Yang et al. (2017a), the 4-phenylsulfonyl-PyMTA tag (4PS-PyMTA) (Figure 5) produces a much shorter linkage to the protein backbone via a shorter thioether bond to restrict the mobility of the coordinated metal ion. These tags are thus highly optimized to place the paramagnetic metal ions close to the protein backbone and produce accurate DEER distance measurements. As a general disadvantage, however, the tagged proteins need to be accurately titrated with a Gd^{3+} ion solution in 1:1 ratio to obtain faithful DEER results.



4MMDPA 3MDPA 4-vinyl-PyMTA 4PS-PyMTA

Figure 5. Structures of Gd^{3+} chelating tags used in DEER distance measurements.

Problems associated with Gd^{3+} titration are avoided by tags that are synthesized with a stably bound Gd^{3+} ion as part of the tag. Cyclen-based tags loaded with a lanthanide ion are particularly interesting as the lanthanide complexes are exceptionally stable. In addition, the bulky cyclen framework helps to restrict the mobility of the Gd^{3+} complex relative to the protein. Over recent years, the cyclen framework has been derivatized with (i) different pendant groups to produce tags with different chemical and spectroscopic properties and (ii) different linker moieties for different attachment strategies, which result in different linker lengths between the protein backbone and the paramagnetic metal centre (Figure 6).

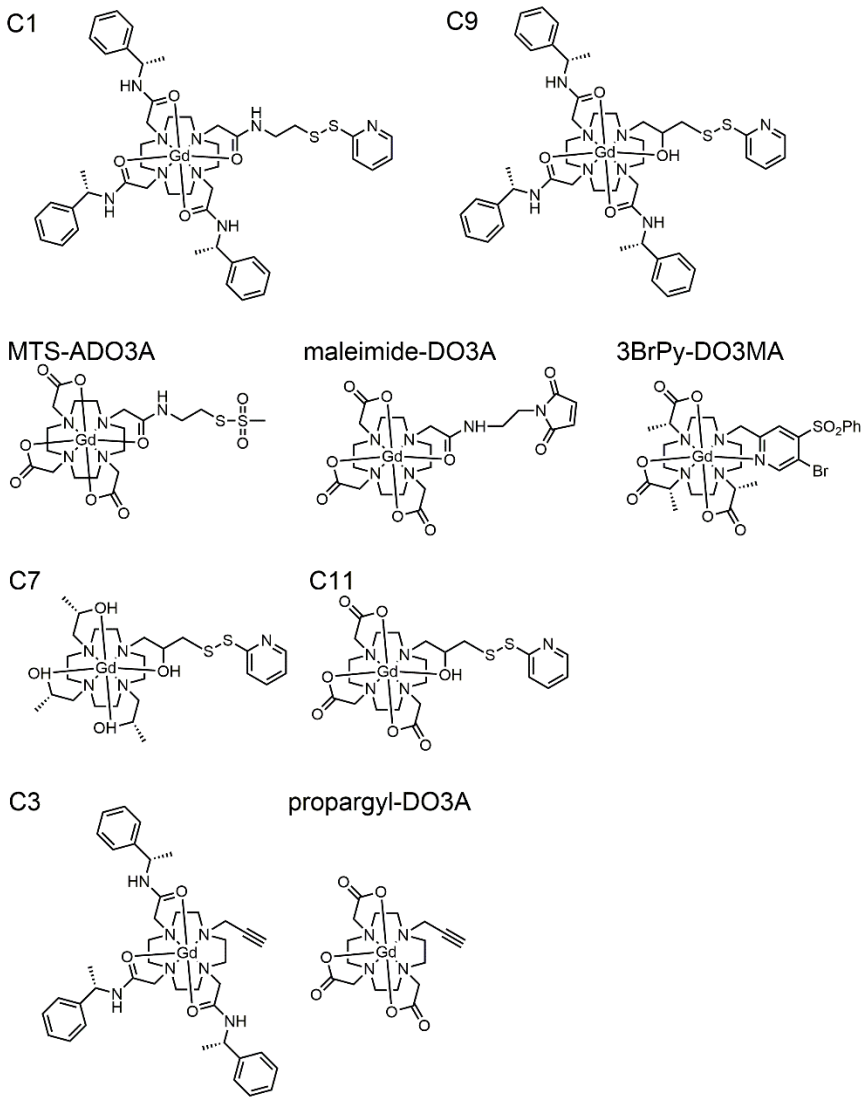


Figure 6. Chemical structures of a selection of cyclen-based Gd^{3+} tags used in DEER distance measurements. All the tags are cyclen-based Gd^{3+} tags but have different pendant groups and linker arms.

The EPR properties of Gd^{3+} tags depend on the chemical structures and coordination properties of the metal ion in ways that are not well understood. DEER distance measurements depend not only on the spectroscopic properties of the tags but, obviously, also on the characteristics of the tether connecting the paramagnetic centre with the protein backbone.

The C1 tag is the first cyclen-based Gd^{3+} tag used in DEER distance measurements (Figure 6; Yagi et al., 2011). It has an activated disulfide bond for spontaneous reaction with solvent-exposed cysteine residues. The tag features a positive overall net charge and its bulky and hydrophobic phenyl pendants were expected to limit

the flexibility of the paramagnetic centre relative to the protein. To further restrict the mobility of the paramagnetic centre attached to the protein backbone, Abdelkader et al. (2015b) introduced the C9 tag (Figure 6), which is structurally similar to the C1 tag except for the length of the tether connecting the protein backbone with the paramagnetic centre. As anticipated, the shorter tether in the C9 tag significantly reduced the conformational space accessible to the Gd^{3+} ion (Figure 7). Therefore, following attachment to the ERp29 homodimer to measure interdomain distances, the C9 tag delivered distance distributions significantly narrower than those delivered by the C1 tag.

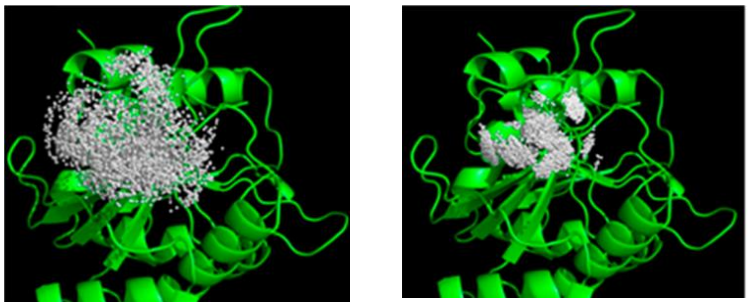


Figure 7. Comparison of the conformational space accessible to the Gd^{3+} metal centre in (A) the C1 tag and (B) the C9 tag. Simulated metal positions are shown as white spheres. (Reproduced from Abdelkader et al., 2015b).

In contrast, MTS-ADO3A and maleimide-DO3A tags that had been developed before the C9 tag failed to deliver narrow DEER distance distributions. Compared with the bulky phenyl pendants of the C1 and the C9 tags, the MTS-ADO3A and maleimide-DO3A tags contain three carboxylate pendants, yielding a neutral net charge and sterically compact structures. Ultimately, the observed broad widths of the DEER distance distributions (full width at half height of 1.5 nm for W-band DEER results of ubiquitin S20C/G35C-Gd^{3+}-maleimide-DO3A *in vitro* and in cells; full width at half-height calculated to be about 1.5 nm for a transmembrane helical peptide tagged with Gd^{3+}-ADO3A but much broader distance distribution observed by W-band DEER) were attributed to the length and flexibility of the tethers of these tags (Manukovsky et al., 2015; Martorana et al., 2014). Notably, however, the chemical stability of the tether produced by the maleimide-DO3A tag is advantageous for in-cell ERR measurements.

Based on the hypothesis that the positive charge of the C1 and C9 tags and the hydrophobicity of their pendants result in undesired, difficult-to-control interactions with the protein surface, the C7 and C8 tags were designed with hydrophilic 2-hydroxypropyl pendants by Prokopiou et al. (2018). The C7 and C8 tags are enantiomers of each other and feature the same linker as the C9 tag. In spite of their smaller size, they delivered

equally narrow DEER distance distributions as the C9 tag, showing that bulkiness of the pendants is not essential to deliver good localization of the metal ion. As a complicating factor, however, the W-band EPR spectra of the C7 and C8 tags proved to be dependent on the local protein environment and buffer conditions. Mahawaththa et al. (2018) subsequently introduced a new DO3A tag, C11, which carries no net charge or bulky hydrophobic phenyl groups in the pendants (as opposed to the C9 tag) or hydroxyl groups (as opposed to the C7 and C8 tags), which are sensitive to the local environment. Importantly, however, the C11 tag features the same short tether as the C9 tag, so that differences in the widths of the DEER distance distributions should reflect the effect of the pendants and the overall charge. Remarkably, the C11 tag delivered significantly narrower DEER distance distributions for cysteine mutants of the Zika virus NS2B-NS3 protease, confirming the hypothesis that charge-charge interactions or hydrophobic interactions imposed by the C9 tag have the potential of perturbing the protein structure or biasing the positioning of the Gd^{3+} ion. Comparison of the DEER distance measurements between the C3 and propargyl-DO3A tags likewise supported the above hypothesis (Mahawaththa et al., 2018).

Based on the combined experience accumulated, one would expect that the DO3MA-3BrPy tag, which has also been identified as a good candidate for in-cell EPR (Yang et al., 2017b), should produce the narrowest and most accurate DEER distance distributions conceivable to a tag attached to cysteine residues as it (i) lacks phenyl groups in the pendants, which could result in undesired hydrophobic interactions with the protein surface, (ii) carries no net charge, which could participate in electrostatic interactions with neighboring residues of the protein and (iii) has a very short and rigid tether between the cysteine sulfur atom and the paramagnetic centre.

Apart from the structural features of the tags, also their spectroscopic properties need to be considered. As demonstrated by Dalaloyan et al. (2015), the spectroscopic properties of the tags, particularly the zero field splitting (ZFS), affect their DEER performance and the widths of the distance distributions. The magnitude of the ZFS is governed by the symmetry of the Gd^{3+} coordination. Tags with small ZFS are characterized by narrow EPR spectra, resulting in higher EPR sensitivities. For such SLs, however, the weak dipolar coupling approximation is invalid for shorter Gd^{3+}–Gd^{3+} distances (below 4 nm) and the standard DEER data analysis artificially broadens the width of the distance distribution. Therefore, tags with a larger ZFS are preferred to probe short Gd^{3+}–Gd^{3+} distances (Collauto et al., 2016; Dalaloyan et al., 2015).

The standard analysis of DEER data uses the weak coupling approximation, neglecting pseudo-secular terms. As the weak coupling approximation is not strictly valid for Gd^{3+} SLs with small ZFS, this results in artificial broadening of the distance distributions (Ching et al., 2015; Dalaloyan et al., 2015). To overcome this problem, relaxation-induced dipolar modulation enhancement (RIDME) experiments have been proposed. The dipolar evolution traces obtained by RIDME experiments are less sensitive to the effects of the pseudo-secular terms (Collauto et al., 2016). RIDME experiments can be carried out at W-band frequencies, where they deliver very high concentration sensitivities (Polyhach et al., 2007; Potapov et al., 2010). In addition, when compared to W-band DEER, RIDME experiments also feature larger dipolar modulation depths, which contribute to increased sensitivity (Razzaghi et al., 2014). Thus, RIDME experiments at W-band tend to be more sensitive than corresponding DEER experiments (Razzaghi et al., 2014) and, for tags with small ZFS, narrower distance distributions have been observed with RIDME than DEER experiments. Unfortunately, standard data analysis software applied to RIDME data also produces harmonics of the dipolar interaction frequencies, making it near-impossible to identify the correct distance distributions (Collauto et al., 2016). The ZFS parameters as well as the electron relaxation times of the tags thus need to be considered when comparing the performance of the tags in DEER distance measurements. A more detailed discussion is presented in Chapter 3.

To address the problem of imprecise localization of the paramagnetic centre, Fleissner et al. (2011) designed a nitroxide tag for double-arm attachment to two cysteine residues. This delivered outstanding results with very narrow distance distributions. The same strategy should be applicable to Gd^{3+} tags. Prior to the present thesis, DOTA-based double-arm lanthanide tags had been used for obtaining large PCSs in paramagnetic NMR (Keizers et al., 2007; Lee et al., 2016), but their Gd^{3+} analogs had not been reported or used for DEER distance distributions.

The ideal Gd^{3+} tag with the spectroscopic and chemical properties to deliver narrow distance distributions under all circumstances is yet to be discovered. In general, double-arm attachment of Gd^{3+} tags and single-arm Gd^{3+} tags with shorter tethers such as the DO3MA-3BrPy and 4PS-PyMTA tags have the potential to restrict the conformational space accessible to the Gd^{3+} ion to get narrower distance distributions. This thesis assesses both approaches.

4. Site-directed protein labelling for NMR and DEER spectroscopy

As described above, DEER experiments and paramagnetic NMR of proteins require the introduction of paramagnetic SLs at defined sites. The problem of site-specific introduction of SLs is closely related to the problem of incorporating general NMR probes into proteins that can be used to study ligand binding and measure intermolecular NOEs. Optimizing ligation strategies is as important as defining the optimal spectroscopic and chemical properties to expand the accuracy and scope of these techniques. When attaching a tag to a protein, high ligation yields and bio-orthogonality are of prime importance.

4.1 Site-directed protein labelling on cysteine

Most tags have been designed for reaction with thiol groups for attachment to cysteine residues in the target protein, either via covalent disulfide (Figure 8A and B) or thioester bonds (Figure 8C and D). Cysteine stands out as the most nucleophilic and oxidation sensitive amino acid among the 20 canonical amino acids. Using cysteine for tagging, however, requires the removal of naturally occurring cysteine residues and introduction of new cysteine residues at the desired locations, which is usually achieved by site-directed mutagenesis.

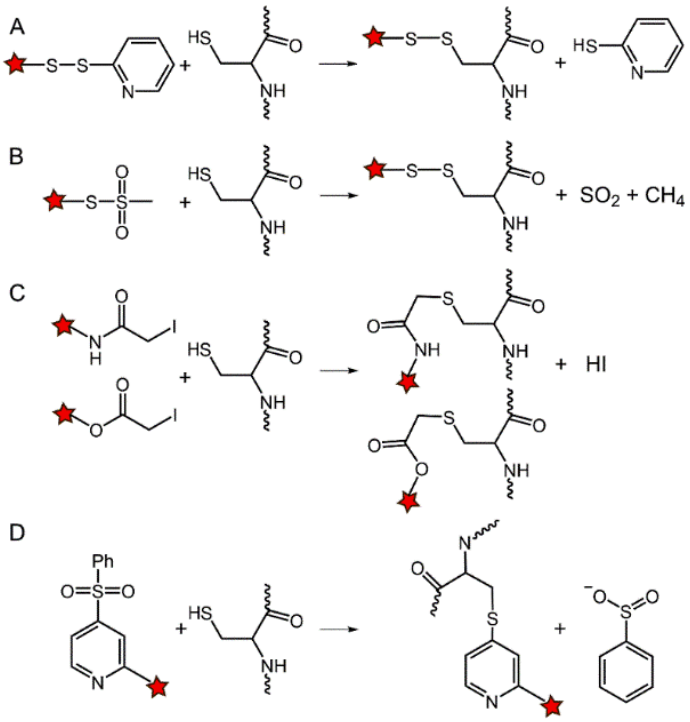


Figure 8. Common strategies used to modify cysteine residues in proteins. Stars indicate different chemical groups. A and B) Conjugation with formation of disulfide bonds. C and D) Conjugation with formation of thioether bonds.

Ligation to cysteine residues is not applicable for cysteine-rich proteins, particularly if any naturally occurring cysteine residues are essential for maintaining the structure or function of the protein. In recent years, many alternative strategies have been proposed to achieve biorthogonal labelling of proteins and many of the methods take advantage of UAAs that can be site-specifically incorporated into proteins by genetic encoding. For instance, a photocaged cysteine (PCC) can be incorporated as a UAA (Nguyen et al., 2014). This enables a protocol, where the native cysteine residues are chemically protected, the PCC residue is decaged and chemically modified, and the native cysteine residues are recovered by deprotection (Yang et al., 2016). Charging of the cognate amber suppressor $tRNA_{CUA}$ with the PCC amino acid is elegantly achieved by a mutant *M. barkeri* PylRS, named PCC2RS (Nguyen et al., 2014), but the number of steps involved in selective cysteine labelling by this protocol renders the overall process unattractive.

4.2 Site-specific protein labelling using UAAs

UAAs that directly chelate a paramagnetic metal ion (Nguyen et al., 2011) or UAAs containing a nitroxide group (Schmidt et al., 2014) have been identified that can be directly incorporated into proteins by genetic encoding like UAAs with other NMR-active functional groups (Cellitti et al., 2008; Chen et al., 2015; Jackson et al., 2007; Liu et al., 2012; Loh et al., 2018). This approach is limited by the availability of the requisite orthogonal aaRS/tRNA pairs. Alternatively, a *p*-azido-phenylalanine residue can be site-specifically incorporated into proteins as a UAA by *p*-CNFRS (Chin et al., 2002). This allows subsequent ligation with different alkyne lanthanide binding tags in a bio-orthogonal, Cu(I)-catalyzed click reaction. The approach has been used to obtain PCSs (Loh et al., 2013; Loh et al., 2015) and DEER distance measurements (Abdelkader et al., 2015a). Its strength lies in excellent bio-orthogonality, but the ligation reaction results in a long and flexible tether between the protein backbone and the paramagnetic moiety, thereby affecting the precision and extent of protein dependent information that can be extracted from DEER distance distributions.

4.3 Paramagnetic metals in specific protein motives

Substitution of a metal ion naturally occurring in a metalloprotein with a paramagnetic metal ion presents another simple but not universally applicable approach to introduce paramagnetic centres into proteins (Bertini et al., 2003). Mn^{2+} ions have been attached by two poly-histidine motives (Ching et al., 2016) and Cu^{2+} ions have been immobilized by two double-histidine motives (Cunningham et al., 2015a; Lawless et al., 2017). These strategies allow distance measurements without requiring the presence of native cysteine residues in the protein, but the poly-histidine motives are structurally labile and the short electronic relaxation times of Cu^{2+} limit the distances that can be measured.

4.4 Site-directed protein labelling of a selenocysteine residue

A generally applicable strategy that achieves site-specific labelling of a protein in the presence of native, solvent-exposed cysteine (Cys) residues and positions the probe close to the protein backbone would be of great value. This can be achieved by selenocysteine (Sec), which is more reactive than cysteine.

Sec has similar chemical properties as cysteine but the pK_a values of Sec and cysteine (5.2 versus 8.5) make Sec more reactive than cysteine especially at slightly acidic

pH conditions (Cheng et al., 2006a,b; Liu et al., 2017; Rakauskaitė et al., 2015). Sec is also referred to as the 21st natural amino acid, but it is rare in both eukaryotic and prokaryotic proteins. The natural translational mechanism for Sec incorporation is complex. Sec insertion naturally occurs at a UGA stop codon and relies 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 (Sec insertion sequence or SECIS) in the vicinity of the target UGA codon. The requisite mRNA stem-loop structure formed by the SECIS element restricts the sequence context where Sec residues can be positioned (Johansson et al., 2005; Mousa et al., 2017). For the same reason, site-specific Sec incorporation into recombinant proteins is challenging to achieve.

Many alternative strategies have been proposed to achieve site-specific Sec incorporation into proteins. Prior to this thesis, a straightforward strategy to yield pure Sec in good yields in *E. coli* still needed to be developed. Selenocysteine incorporation as a photocaged unnatural amino acid, DMNB-Sec, in yeast had been reported (Rakauskaitė et al., 2015). UV illumination of DMNB-Sec yields Sec with high purity. Incorporation of Sec as a photocaged UAA not only bypasses the complex natural Sec incorporation machinery, but also helps to control the reactivity of the selenium, which is prone to form Se–Se bonds under the redox conditions prevailing in the cytosol. It is thus of great interest to develop a similar Sec incorporation strategy suitable for selenoprotein production in *E. coli*, which is the most frequently used organism for recombinant protein synthesis. Unfortunately, the LeuRS/$tRNA_{CUA}$ synthetase pair incorporating the previously published photocaged Sec (Rakauskaitė et al., 2015) is not orthogonal in *E. coli.*

5. Photocaged UAAs in genetic code expansion

Appending a light-sensitive protection group to natural amino acids and incorporating them as photocaged UAAs is of increasing interest in genetic code expansion both to regulate protein functionality (Arbely et al., 2012; Deiters et al., 2006; Gautier et al., 2010; Lemke et al., 2007; Nguyen et al., 2014) and to incorporate amino acids that are otherwise difficult to incorporate by the natural decoding mechanism (Rakauskaitė et al., 2015). Specifically, incorporation of photocaged tyrosine, lysine, serine and cysteine UAAs has been reported to block the specific activity of a protein until it is activated by irradiation with non-destructive UV light (Arbely et al., 2012; Deiters et al., 2006; Gautier

et al., 2010; Lemke et al., 2007; Nguyen et al., 2014). In addition, use of isotopically labelled photocaged tyrosine has been demonstrated for site-specific incorporation of isotope-labelled tyrosine (Cellitti et al., 2008).

The LeuRS synthetase mutant capable of incorporating photocaged Sec had first been evolved to incorporate photocaged serine (Lemke et al., 2007) and only later was found to incorporate cysteine and even Sec analogues of the photocaged serine (Rakauskaitė et al., 2015). Based on these observations, it can be hypothesized that the *M. barkeri* PylRS mutant PCC2RS, which had been evolved to incorporate a PCC UAA (Nguyen et al., 2014), also has the potential to incorporate the Sec analogue, which would provide an alternative and convenient route for Sec incorporation in *E. coli.* Access to proteins containing photocaged Sec is appealing because the highly reactive Sec residue is set free only after protein purification and only temporarily, avoiding undesired reactions in the bacterial cytosol.

6. Cell-free protein synthesis to incorporate UAAs

The wild-type PylRS/$tRNA_{CUA}$ pair and its active-site mutants are exceptionally promiscuous with regard to the variety of UAAs they can accept while maintaining fidelity and orthogonality (Dumas et al., 2015; Liu and Schultz 2010). As a drawback, however, PylRS is catalytically less active than the canonical aaRSs (Suzuki et al., 2017) or *Mj* TyrRS, and this affects the protein yields that can be achieved with UAAs. As a consequence, producing full-length proteins in good yields is challenging with these synthetases, which can be problematic for expensive UAAs.

Cell-free protein synthesis (CFPS) offers a potential solution to address the problem of expensive UAAs, as CFPS is carried out in significantly smaller volumes compared to *in vivo* protein synthesis. In addition, CFPS avoids potential problems with toxic UAAs and charged UAAs, which cannot easily be taken up across the cell membrane. Moreover, the sparing use of amino acids in CFPS reactions is of great interest when the incorporation of UAAs is to be combined with the use of expensive isotope-labelled amino acids, which has important applications in protein NMR.

CFPS with the PylRS/$tRNA_{CUA}$ pair requires PylRS and $tRNA_{CUA}$ in the cell-free reaction mixture in addition to the usual components that are present in conventional CFPS reaction. To provide this, PylRS needs to be recombinantly expressed and purified in good concentration prior to CFPS. Despite the interest in CFPS with PylRS enzymes, however, this has been reported only in very few studies (Chemla et al., 2015; Seki et al.,

2018), because of the challenges associated with purifying full-length PylRS in active form. The main problem has been attributed to the low solubility (~2 mg/mL; Suzuki et al., 2017) of the N-terminal tRNA-recognition domain of the synthetase (Jiang and Krzycki, 2012; Yanagisawa et al., 2006) and a crystal structure of full-length PylRS has never been reported. A crystal structure of the N-terminal domain (NTD) of *M. mazei* PylRS bound to *M. mazei* $tRNA^{Pyl}$, however, has recently been published. The structure shows that the NTD folds into a compact globular structure, which is stabilized by a zinc ion (Suzuki et al., 2017). Purification of full-length PylRS was attempted in this thesis, because this would be important not only for CFPS but also for studies of its structure and dynamics by biophysical techniques such as NMR, which would in turn be helpful for predicting possible modifications to improve the enzymatic activity.

References

Abdelkader, E. H., Feintuch, A., Yao, X., Adams, L. A., Aurelio, L., Graham, B., Goldfarb, D. and Otting, G. (2015a) Protein conformation by EPR spectroscopy using gadolinium tags clicked to genetically encoded *p*-azido-L-phenylalanine. *Chem. Commun*. **51**, 15898–15901.

Abdelkader, E. H., Lee, M. D., Feintuch, A., Ramirez-Cohen, M., Swarbrick, J. D., Otting, G., Graham, B. and Goldfarb, D. (2015b) A new Gd^{3+} spin label for Gd^{3+}-Gd^{3+} distance measurements in proteins produces narrow distance distributions. *J. Phys. Chem. Lett.* **6**, 5016–5021.

Abdullin, D., Duthie, F., Meyer, A., Müller, E. S., Hagelueken, G. and Schiemann, O. (2015) Comparison of PELDOR and RIDME for distance measurements between nitroxides and low-spin Fe(III) ions. *J. Phys. Chem. B* **119**, 13534−13542.

Akhmetzyanov, D., Plackmeyer, J., Endeward, B., Denysenkov, V. and Prisner, T. F. (2015) Pulsed electron-electron double resonance spectroscopy between a high-spin Mn^{2+} ion and a nitroxide spin label. *Phys. Chem. Chem. Phys*. **17**, 6760−6766.

Alff-Steinberger, C. and Epstein R. J. (1994) Codon preference in the terminal region of *E. coli* genes and evolution of stop codon usage. *Theor. Biol*. **168**, 461–463.

Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S. and Schultz, P. G. (2004) An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. USA* **101**, 7566–7571.

Arbely, E., Torres-Kolbus, J., Deiters, A. and Chin, J. W. (2012) Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **134**, 11912–11915.

Astashkin, A. V., Hara, H. and Kawamori, A. (1998) The pulsed electron-electron double resonance and "2+1" electron spin echo study of the oriented oxygen-evolving and Mn-depleted preparations of photosystem II. *J. Chem. Phys.* **108**, 3805–3812.

Bai, X. C., McMullan, G. and Scheres, S. H. W. (2015) How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57.

Banerjee, D., Yagi, H., Huber, T., Otting, G. and Goldfarb, D. (2012) Nanometer range distance measurement in a protein using Mn^{2+} tags. *J. Phys. Chem. Lett*. **3**, 157–160.

Becker, W., Adams, L. A., Graham, B., Wagner, G. E., Zangger, K., Otting, G. and Nitsche, C. (2018a) Trimethylsilyl tag for probing protein-ligand interactions by NMR. *J. Biomol. NMR* **70**, 211−218.

Becker, W., Bhattiprolu, K. C., Gubensäk, N. and Zangger, K. (2018b) Investigating protein–ligand interactions by solution nuclear magnetic resonance spectroscopy. *ChemPhysChem* **19**, 895–906.

Bennati, M., Weber, A., Antonic, J., Perlstein, D. L., Robblee, J. and Stubbe, J. (2003) Pulsed ELDOR spectroscopy measures the distance between the two tyrosyl radicals in the R2 subunit of the *E. coli* ribonucleotide reductase. *J. Am. Chem. Soc*. **125**, 14988–14989.

Bertini, I., Gelis, I., Katsaros, N., Luchinat, C. and Provenzani, A. (2003) Tuning the affinity for lanthanides of calcium binding proteins. *Biochemistry* **42**, 8011–8021.

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.

Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Green-Church, K. B., Chan, M. K. and Krzycki, J. A. (2004) Direct charging of $tRNA_{CUA}$ with pyrrolysine *in vivo* and *in vitro*. *Nature* **431**, 333–335.

Borbat, P. P. and Freed, J. H. (2013) Pulse dipolar electron spin resonance: distance measurements. *Struct. Bond.* **152**, 1–82.

Cellitti, S. E., Jones, D. H., Lagpacan, L., Hao, X., Zhang, Q., Hu, H., Brittain, S. M., Brinker, A., Caldwell, J., Bursulaya, B., Spraggon, G., Brock, A., Ryu, Y., Uno, T., Schultz, P. G. and Geierstanger, B. H. (2008**)** *In vivo* incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **130**, 9268−9281.

Chao, L. K., Clegg, R. M. and Begley, T. P. (2008) Förster resonance energy transfer (FRET) for proteins. Wiley Encyclopedia of Chemical Biology: John Wiley & Sons, Inc.

Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V. and Alfonta, L. (2015) Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. *Biotech. Bioeng.* **112**, 1663–1672.

Chen, W. N., Kuppan, K. V., Lee, M., Jaudzems, K., Huber, T. and Otting, G. (2015) O-*tert*-butyltyrosine, an NMR tag for high-molecular weight systems and measurements of submicromolar ligand binding affinities. *J. Am. Chem. Soc*. **137**, 4581−4586.

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

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

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L. and Schultz, P. G. (2002) Addition of *p*-azido-L-phenylaianine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* **124**, 9026–9027.

Ching, H. Y. V., Demay-Drouhard, P., Bertrand, H. C., Policar, C., Tabares, L. C. and Un, S. (2015) Nanometric distance measurements between Mn(II) DOTA centers. *Phys. Chem. Chem. Phys*. **17**, 23368–23377.

Ching, H. Y. V., Mascali, F. C., Bertrand, H. C., Bruch, E. M., Demay-Drouhard, P., Rasia, R. M., Policar, C., Tabares, L. C. and Un, S. (2016) The use of Mn(II) bound to His-tags as genetically encodable spin-label for nanometric distance determination in proteins. *J. Phys. Chem. Lett.* **7**, 1072–1076.

Collauto, A., Frydman, V., Lee, M. D., Abdelkader, E. H., Feintuch, A., Swarbrick, J. D., Graham, B., Otting, G. and Goldfarb, D. (2016) RIDME distance measurements using Gd(iii) tags with a narrow central transition. *Phys. Chem. Chem. Phys.* **18**, 19037–19049.

Cooley, R. B., Karplus, P. A. and Mehl, R. A (2014) Gleaning unexpected fruits from hard-won synthetases: Probing principles of permissivity in non-canonical amino acid-tRNA synthetases. *ChemBioChem* **12**, 1810–1819.

Cunningham, T. F., Putterman, M. R., Desai, A., Horne, W. S. and Saxena, S. (2015a) The double-histidine Cu^{2+}-binding motif: a highly rigid, site-specific spin probe for electron spin resonance distance measurements. *Angew. Chem. Int. Ed*. **54**, 6330–6334.

Cunningham, T. F., Shannon, M. D., Putterman, M. R., Arachchige, R. J., Sengupta, I., Gao, M., Jaroniec, C. P. and Saxena, S. (2015b) Cysteine-specific Cu^{2+} chelating tags used as paramagnetic probes in double electron electron resonance. *J. Phys. Chem. B.* **119**, 2839–2843.

Dalaloyan, A., Qi, M., Ruthstein, S., Vega, S., Godt, A., Feintuch, A. and Goldfarb, D. (2015) Gd(iii)-Gd(iii) EPR distance measurements-the range of accessible distances and the impact of zero field splitting. *Phys. Chem. Chem. Phys*. **17**, 18464–18476.

Deiters, A., Groff, D., Ryu, Y., Xie, J. and Schultz, P. G. (2006) A genetically encoded photocaged tyrosine. *Angew. Chem. Int. Ed.* **45**, 2728–2731.

Denysenkov, V. P., Prisner, T. F., Stubbe, J. and Bennati, M. (2006) High-field pulsed electron–electron double resonance spectroscopy to determine the orientation of the

tyrosyl radicals in ribonucleotide reductase. *Proc. Natl. Acad. Sci. USA* **103**, 13386–13390.

Dery, S., Reddy, P. S., Dery, L., Mousa, R., Dardashti, R. N. and Metanis, N. (2015) Insights into the deselenization of selenocysteine into alanine and serine. *Chem. Sci.* **6**, 6207–6212.

Dumas, A., Lercher, L., Spicer, C. D. and Davis, B. G. (2015) Designing logical codon reassignment - expanding the chemistry in biology. *Chem. Sci.* **6**, 50–69.

Ezhevskaya, M., Bordignon, E., Polyhach, Y., Moens, L., Dewilde, S., Jeschke, G. and Van Doorslaer, S. (2013) Distance determination between low-spin ferric haem and nitroxide spin label using DEER: the neuroglobin case. *Mol. Phys.* **111**, 2855−2864.

Feintuch, A., Otting, G. and Goldfarb, D. (2015) Gd^{3+} spin labeling for measuring distances in biomacromolecules: why and how? *Methods Enzymol*. **563**, 415–457.

Ferreon, J. C. and Hilser, V. J. (2003) Ligand-induced changes in dynamics in the RT loop of the C-terminal SH3 domain of Sem-5 indicate cooperative conformational coupling. *Protein Sci.* **12**, 982–996.

Fleissner, M. R., Bridges, M. D., Brooks, E. K., Cascio, D., Kálai, T., Hideg, K. and Hubbell, W. L. (2011) Structure and dynamics of a conformationally constrained nitroxide side chain and applications in EPR spectroscopy. *Proc. Natl. Acad. Sci. USA* **108**, 16241–16246.

Freedberg, D. I. and Selenko, P. (2014) Live cell NMR. *Annu. Rev. Biophys.* **43**, 171–192.

Garbuio, L., Bordignon, E., Brooks, E. K., Hubbell, W. L., Jeschke, G. and Yulikov, M. (2013) Orthogonal spin labeling and Gd(III)-nitroxide distance measurements on bacteriophage T4-lysozyme. *J. Phys. Chem. B* **117**, 3145−3153.

Gautier, A., Nguyen, D. P., Lusic, H., An, W., Deiters, A. and Chin, J. W. (2010) Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* **132**, 4086–4088.

Gossert, A. D. and Jahnke, W. (2016) NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc*. **97**, 82–125.

Grenthe, I. J. (1980) Stability relationships among the rare earth dipicolinates. *J. Am. Chem. Soc*. **83**, 360–364.

Grohmann, D., Klose, D., Klare, J. P., Kay, C. W. M., Steinhoff, H. J. and Werner, W. (2010) RNA-binding to archaeal RNA polymerase subunits F/E: A DEER and FRET study. *J. Am. Chem. Soc.* **132**, 5954–5955.

Guo, L. –T., Wang, Y. –S., Nakamura, A., Eiler, D., Kavran, J. M., Wong, M., Kiessling, L. L., Steitz, T. A., O'Donoghue, P. and Söll, D. (2014) Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proc. Natl. Acad. Sci. USA* **111**, 16724–16729.

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.

Haugland, M. M., Lovett, J. E. and Anderson, E. A. (2018) Advances in the synthesis of nitroxide radicals for use in biomolecule spin labelling. *Chem. Soc. Rev.* **47**, 668–680.

Hubbell, W. L., Cafiso, D. S. and Altenbach, C. (2000) Identifying conformational changes with site-directed spin labelling. *Nat. struct. Biol.* **7**, 735–739.

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

Jackson, J. C., Hammill, J. T. and Mehl, R. A. (2007) Site-specific incorporation of a ^{19}F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc.* **129**, 1160−1166.

Jeschke, G., Wegener, C., Nietschke, M., Jung, H. and Steinhoff, H. (2004) Interresidual distance determination by four-pulse double electron-electron resonance in an integral membrane protein: the Na^+/prolin transporter PutP of *Escherichia coli*. *Biophys. J*. **86**, 2551–2557.

Jeschke, G. (2012) DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.* **63**, 419–446.

Jiang, R. and Krzycki, J. A. (2012) PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine**.** *J. Biol. Chem*. **287**, 32738–32746.

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

Joseph, B., Tormyshev, V. M., Rogozhnikova, O. Y., Akhmetzyanov, D., Bagryanskaya, E. G. and Prisner, T. F. (2016) Selective high-resolution detection of membrane protein-ligand interaction in native membranes using trityl-nitroxide PELDOR. *Angew. Chem. Int. Ed*. **55**, 11538−11542.

Kaminker, I., Yagi, H., Huber, T., Feintuch, A., Otting, G. and Goldfarb, D. (2012) Spectroscopic selection of distance measurements in a protein dimer with mixed nitroxide and Gd^{3+} spin labels. *Phys. Chem. Chem. Phys*. **14**, 4355−4358.

Kaminker, I., Tkach, I., Manukovsky, N., Huber, T., Yagi, H., Otting, G., Bennati, M. and Goldfarb, D. (2013) W-band orientation selective DEER measurements on a Gd^{3+}/nitroxide mixed-labeled protein dimer with a dual mode cavity. *J. Magn. Reson.* **227**, 66−71.

Kaminker, I., Bye, M., Mendelman, N., Gislason, K., Sigurdsson, S. T. and Goldfarb, D. (2015) Distance measurements between manganese(II) and nitroxide spin-labels by DEER determine a binding site of Mn^{2+} in the HP92 loop of ribosomal RNA. *Phys. Chem. Chem. Phys*. **17**, 15098−15102.

Kazmier, K., Sharma, S., Quick, M., Islam, S. M., Roux, B., Weinstein, H., Javitch, J. A. and Mchaourab, H. S. (2014a) Conformational dynamics of ligand-dependent alternating access in LeuT. *Nat. Struct. Mol. Biol*. **21**, 472−479.

Kazmier, K., Sharma, S., Islam, S. M., Roux, B. and Mchaourab, H. S. (2014b) Conformational cycle and ion-coupling mechanism of the Na+/hydantoin transporter Mhp1. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14752−14757.

Keizers, P. H. J., Desreux, J. F., Overhand, M. and Ubbink, M. (2007) Increased paramagnetic effect of a lanthanide protein probe by two-point attachment. *J. Am. Chem. Soc.* **129**, 9292–9293.

Lang, K. and Chin, J. W. (2014) Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **114**, 4764–4806.

Lawless, M. J., Ghosh, S., Cunningham, T. F., Shimshi, A. and Saxena, S. (2017) On the use of the Cu^{2+}-iminodiacetic acid complex for double histidine based distance measurements by pulsed ESR. *Phys. Chem. Chem. Phys*. **19**, 20959–20967.

Lee, M. D., Dennis, M. L., Swarbrick, J. D. and Graham, B. (2016) Enantiomeric two-armed lanthanide-binding tags for complementary effects in paramagnetic NMR spectroscopy. *Chem. Commun*. **52**, 7954–7957.

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.

Link, A. J. and Tirrell, D. A. (2005) Reassignment of sense codons *in vivo*. *Methods* **36**, 291–298.

Liu, C. C. and Schultz, P. G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444.

Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. and Wüthrich, K. (2012) Biased signaling pathways in β2-adrenergic receptor characterized by ^{19}F-NMR. *Science* **335**, 1106−1110.

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

Loh, C. T., Adams, L. A., Graham, B. and Otting, G. (2018) Genetically encoded amino acids with *tert*-butyl and trimethylsilyl groups for site-selective studies of proteins by NMR spectroscopy. *J. Biomol. NMR* **71**, 287–293.

Loh, C. T., Ozawa, K., Tuck, K., Barlow, N., Huber, T., Otting, G. and Graham, B. (2013) Lanthanide tags for site-specific ligation to an unnatural amino acid and generation of pseudocontacts shifts in proteins. *Bioconjugate Chem*. **24**, 260–268.

Loh, C. T., Graham, B., Abdelkader, E. H., Tuck, K. L. and Otting, G. (2015) Generation of pseudocontact shifts in proteins with lanthanides using small "clickable" nitrilotriacetic acid and iminodiacetic acid tags. *Chem. Eur. J.* **21**, 5084–5092.

Loscha, K. V., Herlt, A. J., Qi, R., Huber, T., Ozawa, K. and Otting, G. (2012) Multiple-site labeling of proteins with unnatural amino acids. *Angew. Chem. Int. Ed*. **51**, 2243–2246.

Lueders, P., Jeschke, G. and Yulikov, M. (2011) Double electron-electron resonance measured between Gd^{3+} ions and nitroxide radicals. *J. Phys. Chem. Lett*. **2**, 604−609.

Mahawaththa, M. C., Lee, M. D., Giannoulis, A., Adams, L. A., Feintuch, A., Swarbrick, J. D., Graham, B., Nitsche, C., Goldfarb, D. and Otting, G. (2018) Small neutral Gd(III) tags for distance measurements in proteins by double electron–electron resonance experiments. *Phys. Chem. Chem. Phys*., submitted.

Man, B., Su, X. –C., Liang, H., Simonsen, S., Huber, T., Messerle, B. A. and Otting, G. (2010) 3-Mercapto-2,6-pyridinedicarboxylic acid, a small lanthanide-binding tag for protein studies by NMR spectroscopy. *Chem. Eur. J*. **16**, 3827–3832.

Manukovsky, N., Frydman, V. and Goldfarb, D. (2015) Gd^{3+} spin labels report the conformation and solvent accessibility of solution and vesicle-bound melittin. *J. Phys. Chem. B.* **119**, 13732–13741.

Martorana, A., Bellapadrona, B., Feintuch, A., Di Gregorio, E., Aime, S. and Goldfarb, D. (2014) Probing protein conformation in cells by EPR distance measurements using Gd^{3+} spin labeling. *J. Am. Chem. Soc*. **136**, 13458–13465.

McCoy, M. and Wyss, D. F. (2002) Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J. Am. Chem. Soc*. **124**, 11758–11763.

McDermott, A. (2004) Structural and dynamic studies of proteins by solid-state NMR spectroscopy: rapid movement forward. *Curr. Opin. Struct. Biol.* **14**, 554–561.

Meyer, V., Swanson, M. A., Clouston, L. J., Boratyński, P. J., Stein, R. A., Mchaourab, H. S., Rajca, A., Eaton, S. S. and Eaton, G. R. (2015). Room-temperature distance measurements of immobilized spin-labeled protein by DEER/PELDOR. *Biophys. J*. **108**, 1213–1219.

Meyer, A. and Schiemann, O. (2016) PELDOR and RIDME measurements on a high-spin manganese(II) bisnitroxide model complex. *J. Phys. Chem. A* **120**, 3463−3472.

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

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.

Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., Kira, S., Yanagisawa, T., Yokoyama, S., Hoshi, H., Kobayashi, T. and Sakamoto, K. (2015) Rassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.* **43**, 8111–8122.

Neuman, H., Hahn, L. E. and Lammers, C. (2014) Optimized plasmid systems for the incorporation of multiple different unnatural amino acids by evolved orthogonal ribosomes. *ChemBioChem.* **15**, 1800–1804.

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. and Chin, J. W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444.

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.

Nguyen, T. H. D., Ozawa, K., Stanton-Cook, M., Barrow, R., Huber, T. and Otting, G. (2011) Pseudocontact shifts in protein NMR spectra generated using a genetically encoded Co^{2+}-binding amino acid. *Angew. Chem. Int. Ed*. **50**, 692–694.

Nitsche, C. and Otting, G. (2018) NMR studies of ligand binding. *Curr. Opin. Struct. Biol.* **48**, 16–22.

Nogales, E. and Scheres, S. H. W. (2015) Cryo-EM: A unique tool for the visualization of macromolecular complexity. *Mol. Cell* **58**, 677–689.

Nozawa, K., O'Donoghue, P., Gundllapalli, S., Araiso, Y., Ishitani, R., Umehara, T., Söll, D. and Nureki, O. (2009) Pyrrolysyl-tRNA synthetase-$tRNA^{Pyl}$ structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163–1167.

Pintacuda, G., John, M., Su, X. –C. and Otting, G. (2007) NMR structure determination of protein-ligand complexes by lanthanide labeling. *Acc. Chem. Res*. **40**, 206–212.

Piston, D. W. and Kremers, G. J. (2007) Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem. Sci.* **32**, 407–414.

Polyhach, Y., Godt, A., Bauer, C. and Jeschke, G. (2007) Spin pair geometry revealed by high-field DEER in the presence of conformational distributions. *J. Magn. Reson*. **185**, 118−129.

Polycarpo, C., Ambrogelly, A., Bérube, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., Wood, J. L. and Söll, D. (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. USA* **101**, 12450–12454.

Potapov, A., Yagi, H., Huber, T., Jergic, S., Dixon, N. E., Otting, G. and Goldfarb, D. (2010) Nanometer scale distance measurements in proteins using Gd^{3+} spin labelling. *J. Am. Chem. Soc*. **132**, 9040–9048.

Prokopiou, G., Lee, M. D., Collauto, A., Abdelkader, E. H., Bahrenberg, T., Feintuch, A., Ramirez-Cohen, M., Clayton, J., Swarbrick, J. D., Graham, B., Otting, G. and Goldfarb, D. (2018) Small Gd(III) tags for Gd(III)-Gd(III) distance measurements in proteins by EPR spectroscopy. *Inorg. Chem*. **57**, 5048–5059.

Qi, M., Gross, A., Jeschke, G., Godt, A. and Drescher, M. (2014) Gd(III)-PyMTA label is suitable for in-cell EPR. *J. Am. Chem. Soc.* **136**, 15366–15378.

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.

Razzaghi, S., Qi, M., Nalepa, A. I., Godt, A., Jeschke, G., Savitsky, A. and Yulikov, M. (2014) RIDME spectroscopy with Gd(III) centers. *J. Phys. Chem. Lett*. **5**, 3970–3975.

Schmidt, M. J., Borbas, J., Drescher, M. and Summerer, D. (2014) A genetically encoded spin label for electron paramagnetic resonance distance measurements. *J. Am. Chem. Soc.* **136**, 1238–1241.

Seki, E., Yanagisawa, T. and Yokoyama, S. (2018) Cell-free protein synthesis for multiple site-specific incorporation of noncanonical amino acids using cell extracts from RF-1 deletion *E. coli* strains. *Methods Mol. Biol.* **1728**, 49–65.

Shevelev, G. Y., Krumkacheva, O. A., Lomzov, A. A., Kuzhelev, A. A., Rogozhnikova, O. Y., Trukhin, D. V., Troitskaya, T. I., Tormyshev, V. M., Fedin, M. V., Pyshnyi, D. V. and Bagryanskaya, E. G. (2014). Physiological-temperature distance measurement in nucleic acid using triarylmethyl-based spin labels and pulsed dipolar EPR spectroscopy. *J. Am. Chem. Soc.* **136**, 9874–9877.

Shevelev, G. Y., Krumkacheva, O. A., Lomzov, A. A., Kuzhelev, A. A., Trukhin, D. V., Rogozhnikova, O. Y., Tormyshev, V. M., Pyshnyi, D. V., Fedin, M. V. and Bagryanskaya, E. G. (2015) Triarylmethyl labels: Toward improving the accuracy of EPR nanoscale distance measurements in DNAs. *J. Phys. Chem. B* **119**, 13641−13648.

Shi, Y. (2014) A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995–1014.

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.

Steer, B. A. and Schimmel, P. (1999) Major anticodon-binding region missing from an archaeal tRNA synthetase. *J. Biol. Chem.* **274**, 35601–35606.

Stone, T. J., Buckman, T., Nordio, P. L. and McConnel, H. M. (1965) Spin-labeled biomolecules. *Proc. Natl. Acad. Sci. USA* **54**, 1010–1017.

Su, X. –C., Man, B., Beeren, S., Liang, H., Simonsen, S., Schmitz, C., Huber, T., Messerle, B. A. and Otting, G. (2008) A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy. *J. Am. Chem. Soc*. **130**, 10486–10487.

Suzuki, T., Miller, C., Guo, L. –T., Ho, J. M. L., Bryson, D. I., Wang, Y. –S., Liu, D. R. and Söll, D. (2017) Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* **13**, 1261–1266.

Sze, K. H., Wu, Q., Tse, H. S. and Zhu, G. (2012) Dynamic nuclear polarization: new methodology and applications. *Top. Curr. Chem.* **326**, 215–242.

Thompson, L. K. (2002) Solid-state NMR studies of the structure and mechanisms of proteins. *Curr. Opin. Struct. Biol.* **12**, 661–669.

Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P. –J., Russell, D. H. and Liu, W. R. (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angew. Chem. Int. Ed.* **49**, 3211–3214.

Wang, K., Sachdeva, A., Cox, D. J., Wilf, N. M., Lang, K., Wallace, S., Mehl, R. A. and Chin, J. W. (2014) Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* **6**, 393–403.

Wang, L., Brock, A., Herberich, B. and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500.

Wang, L., Magliery, T. J., Liu, D. R. and Schultz, P. G. (2000) A new functional tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.* **122**, 5010–5011.

Wang, L., Xie, J. and Schultz, P. (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249.

Weiss, S. (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.* **7**, 724–729.

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

Wu, N., Deiters, A., Cropp, T. A., King, D. and Schultz, P. G. (2004) A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 4306–14307.

Wu, Z., Feintuch, A., Collato, A., Adams, L. A., Aurelio, L., Graham, B., Otting, G. and Goldfarb, D. (2017) Selective distance measurements using triple spin labeling with Gd^{3+}, Mn^{2+} and a nitroxide. *J. Phys. Chem. Lett.* **8**, 5277−5282.

Wüthrich, K. (1989) Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* **243**, 45–50.

Wüthrich, K. (1990) Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* **265**, 22059–22062.

Yagi, H., Banerjee, D., Graham, B., Huber, T., Goldfarb, D. and Otting, G. (2011) Gadolinium tagging for high-precision measurements of 6 nm distances in protein assemblies by EPR. *J. Am. Chem. Soc.* **133**, 10418–10421.

Yanagisawa, T., Ishii, R., Fukunaga, R., Nureki, O. and Yokoyama, S. (2006) Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*. *Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun.* **62**, 1031–1033.

Yang, Z., Liu, Y., Borbat, P., Zweier, J., Freed, J. and Hubbell, W. (2012a) Room temperature electron spin resonance distance measurements in T4 lysozyme using trityl-based spin labels. *Biophys. J.* **102**, 405A.

Yang, Z., Liu, Y., Borbat, P., Zweier, J. L., Freed, J. H. and Hubbell, W. L. (2012b) Pulsed ESR dipolar spectroscopy for distance measurements in immobilized spin labeled proteins in liquid solution. *J. Am. Chem. Soc*. **134**, 9950–9952.

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.

Yang, Y., Gong, Y. –J., Litvinov, A., Liu, H. –K., Yang, F., Su, X. –C. and Goldfarb, D. (2017a) Generic tags for Mn(II) and Gd(III) spin labels for distance measurements in proteins. *Phys. Chem. Chem. Phys*. **19**, 26944–26956.

Yang, Y., Yang, F., Gong, Y. –J., Chen, J. –L., Goldfarb. D. and Su, X. –C. (2017b) A reactive, rigid Gd-III labeling tag for in-cell EPR distance measurements in proteins. *Angew. Chem. Int. Ed*. **56**, 2914–2918.

Zuiderweg, E. R. P. (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1–7.

CHAPTER 2

DOUBLE-ARM LANTHANIDE TAGS DELIVER NARROW Gd^{3+}–Gd^{3+} DISTANCE DISTRIBUTIONS IN DOUBLE ELECTRON–ELECTRON RESONANCE (DEER) MEASUREMENTS

This chapter is reproduced from the following published article:

Welegedara, A. P., Yang, Y., Lee, M. D., Swarbrick, J. D., Huber, T., Graham, B., Goldfarb, D. and Otting, G. (2017) Double-arm lanthanide tags deliver narrow Gd^{3+}-Gd^{3+} distance distributions in double electron-electron resonance (DEER) measurements. *Chem. Eur. J.* **23**, 11694–11702.

Contribution:

I prepared all protein samples for DEER analysis, including protein synthesis, purification, ligation with T1, T2, IDA-SH and C9 tags and preparation of EPR samples. The tags were synthesized by Dr M. D. Lee in the group of A/Prof. Bim Graham and Dr J. D. Swarbrick at the Monash Institute of Pharmaceutical Sciences, Monash University, Australia. The DEER distance measurements and data analysis were performed by Dr Y. Yang in the group of Prof. D. Goldfarb at the Department of Chemical Physics, Weizmann Institute of Science, Israel. Prof. T. Huber calculated predicted distance distributions for the double-arm tags. I interpreted the DEER distance measurements, modelled the tagged protein and predicted the distance distributions for the C9 tag, and prepared the initial version of the manuscript.

Protein Structures

Double-Arm Lanthanide Tags Deliver Narrow Gd^{3+}–Gd^{3+} Distance Distributions in Double Electron–Electron Resonance (DEER) Measurements

Adarshi P. Welegedara[+],[a] Yin Yang[+],[b] Michael D. Lee[+],[c] James D. Swarbrick,[c] Thomas Huber,[a] Bim Graham,*[c] Daniella Goldfarb,*[b] and Gottfried Otting*[a]

Abstract: Double-arm cyclen-based Gd^{3+} tags are shown to produce accurate nanometer scale Gd^{3+}–Gd^{3+} distance measurements in double electron–electron resonance (DEER) experiments by confining the space accessible to the metal ion. The results show excellent agreement with predictions both for the maximum and width of the measured distance distributions. For distance measurements in proteins, the tags can be attached to two cysteine residues located in positions i and $i+4$, or i and $i+8$, of an α-helix. In the latter case, an additional mutation introducing an aspartic acid at position $i+4$ achieves particularly narrow distribution widths. The concept is demonstrated with cysteine mutants of T4 lysozyme and maltose binding protein. We report the narrowest Gd^{3+}–Gd^{3+} distance distributions observed to date for a protein. By limiting the contribution of tag mobility to the distances measured, double-arm Gd^{3+} tags open new opportunities to study the conformational landscape of proteins in solution with high sensitivity.

Introduction

The Protein Data Bank contains the coordinates of tens of thousands of proteins determined by X-ray crystallography, but assessing their conformations and structural variability at atomic resolution in solution is much more difficult. For multidomain proteins and complexes between biological macromolecules, selective distance measurements on the nanometer scale offer a relatively straightforward way to probe the three-dimensional structure in solution and establish structure–function relationships. The determination of the structural variability underpinning the function of many proteins, however, adds a further challenge that requires not only measurements of average distances, but also faithful measurements of distance distributions.

The double electron–electron resonance (DEER) experiment (also called PELDOR) is a pulse EPR experiment for measuring specific nanometer distances in proteins and macromolecular assemblies. Importantly, the experiment measures not only distances, but entire distance distributions, yielding a unique tool to assess structural variability.

Analysis of the distance distribution and particularly its width is complicated by the fact that the width of the distribution reflects the structural variability of both the protein and the tag. Therefore, the important structural information encoded in the width of the distance distribution can be extracted only with a rigid tag that does not perturb the protein structure and allows a faithful prediction of the location of the spin label with respect to the protein backbone.

For molecules containing two Gd^{3+} ions, the DEER experiment measures the distance between the Gd^{3+} ions and can be conducted at W-band frequency (94 GHz), at which only small amounts (>0.09 nmols) of sample are required.[1,2] Measuring Gd^{3+}–Gd^{3+} distances is advantageous for the absence of orientation selection effects that complicate corresponding measurements with nitroxide tags at this magnetic field strength.[3–6] Moreover, the redox stability of Gd^{3+} makes it attractive for in-cell DEER distance measurements.[7–11] For Gd^{3+}–Gd^{3+} distance measurements in proteins, the Gd^{3+} ions must be introduced through suitable tag molecules that position a Gd^{3+} ion site-specifically in a defined location relative to the protein backbone. This is not trivial to achieve. For example, the cyclen-based **C1-Gd**,[12,13] **C9-Gd**,[14] and **DO3MA-3BrPy-Gd**[11] tags have been attached to single cysteine residues leading to several rotatable bonds in the tether between the Gd^{3+} ion and protein backbone; this in turn leads to uncertainty of

[a] *A. P. Welegedara,[+] Prof. T. Huber, Prof. G. Otting*
Research School of Chemistry
Australian National University, Canberra ACT 2601 (Australia)
E-mail: gottfried.otting@anu.edu.au

[b] *Dr. Y. Yang,[+] Prof. D. Goldfarb*
Department of Chemical Physics
Weizmann Institute of Science, Rehovot 7610001 (Israel)
E-mail: daniella.goldfarb@weizmann.ac.il

[c] *Dr. M. D. Lee,[+] Dr. J. D. Swarbrick, Dr. B. Graham*
Monash Institute of Pharmaceutical Sciences
Monash University, Parkville VIC 3052 (Australia)
E-mail: bim.graham@monash.edu

[+] *The first three authors contributed equally.*

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/chem.201702521.

the positions of the Gd^{3+} ions and results in broadened distance distributions. Previously, a double-arm nitroxide tag attached to two cysteine residues has been shown to restrict the motional freedom of the tag and to produce narrow distance distributions in DEER experiments.[15] The present work demonstrates the DEER performance of different double-arm strategies to immobilize Gd^{3+} ions.

The first approach used two double-arm cyclen-based tags, **T1-Gd** and **T2-Gd** (Figure 1 a and b),[16] which feature chiral 2-hydroxypropyl pendants to promote single tag conformations and two activated disulfide groups for facile ligation to two cysteine residues. Loaded with Tm^{3+} and Yb^{3+}, the **T1** and **T2** tags produced exceptionally large pseudocontact shifts in NMR measurements of different proteins, indicating good immobilization of the lanthanide ion, while leaving the protein structure intact.[16] The second approach used an iminodiacetic acid (IDA) derivative with an activated disulfide group. Two copies of this tag molecule (in the following referred to as **IDA-SH** tag) have been shown to combine to generate a hexadentate binding motif for a single metal ion when they are attached to two cysteine residues in positions *i* and $i+4$ of an α-helix.[17,18] It was reasoned that the short tether would position the Gd^{3+} ion relatively close to the protein backbone, which could facilitate the interpretation of Gd^{3+}–Gd^{3+} distances in terms of protein structure.

To benchmark the performance of the tags, we performed experiments with cysteine mutants of T4 lysozyme and maltose binding protein (MBP) and compared the results with data obtained with the single-arm **C9-Gd** tag (Figure 1 e), which has previously been shown to produce narrow distance distributions in DEER experiments.[14] For best comparison with distance distributions predicted by modeling on crystal structures, only intra-domain distances were examined.

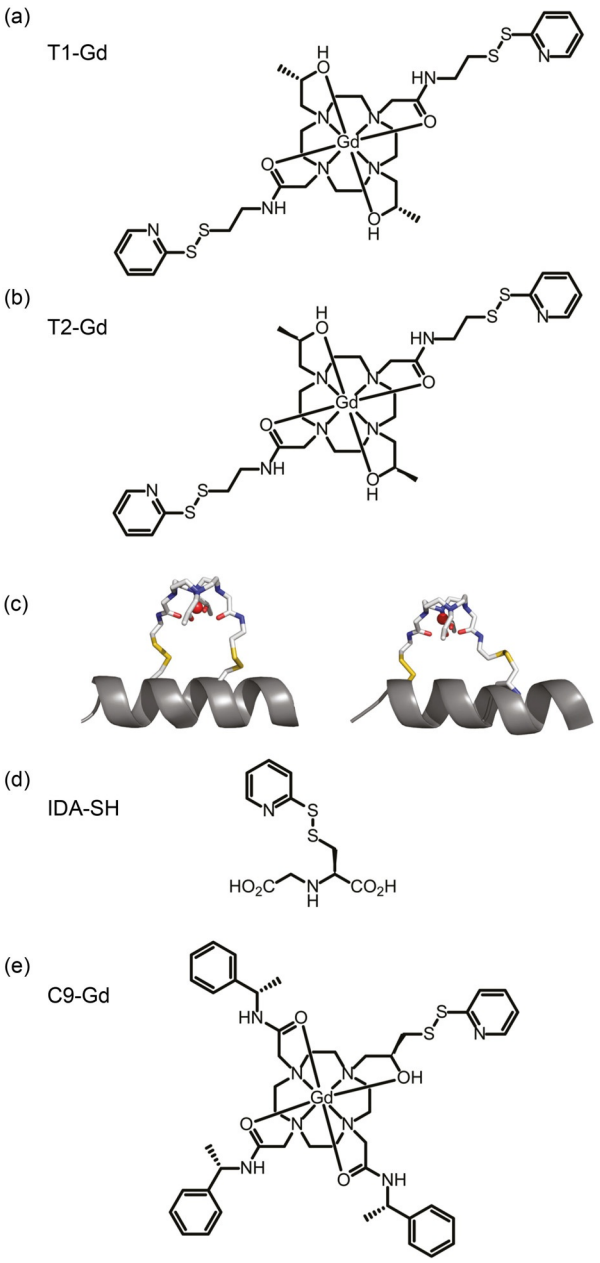


Figure 1. Structures of a) **T1-Gd** tag (2-hydroxypropyl arms have *S* stereochemistry), b) **T2-Gd** tag (2-hydroxypropyl arms have *R* stereochemistry), c) **T2-Gd** tag modeled onto two cysteine residues in positions *i* and $i+4$ (left) or positions *i* and $i+8$ (right) of an α-helix, d) the **IDA-SH** tag activated by a thiopyridyl group for spontaneous reaction with a cysteine thiol group, and e) the **C9-Gd** tag.

Results

Tagging strategies

In the first instance, tags were attached to cysteine residues located in positions *i* and $i+4$ of an α-helix, which in the following is referred to as the "*i*,$i+4$ attachment mode". This attachment mode has previously been used for the **CLaNP-5** tag, which is also a double-arm tag based on a cyclen-lanthanide complex and features the same covalent linkers with cysteine residues as also found in the **T1-Gd** and **T2-Gd** tags.[19,20] In addition, we tested the **T1-Gd** and **T2-Gd** tags in a different attachment mode, in which the cysteine residues were located in positions *i* and $i+8$ of an α-helix (the "*i*,$i+8$ attachment mode") with an aspartic acid residue positioned in position $i+4$ to provide an additional favorable electrostatic interaction with the overall positive charge of the tags (Figure 1 c). Like the other tags, the **IDA-SH** tag was synthesized with an activated disulfide group for spontaneous reaction with cysteine thiol groups and ligated to cysteine residues in positions *i* and $i+4$ of α-helices. Coordination of one Gd^{3+} ion by two **IDA-SH** tags was achieved by titrating the protein with $GdCl_3$ at a 1:2 (Gd^{3+}:**IDA-SH** tag) molar ratio, as reported previously for Co^{2+} coordination.[18]

Protein samples

Six mutants of T4 lysozyme were prepared. In all of them, the natural cysteine residues at positions 54 and 97 were mutated to threonine and alanine, respectively. These mutations are known to maintain the structure of the protein.[21] The additional mutations placed cysteine residues in α-helices. Two mu-

tants put cysteine residues in positions 72 and 76, 127 and 131, and 131 and 135. In the following, we refer to the C54T/C97A/D72C/R76C/D127C/V131C mutant as T4L-A and the C54T/C97A/D72C/R76C/V131C/K135C mutant as T4L-B. The third mutant, referred to as T4L-C, positioned cysteines in positions 72 and 80, and 127 and 135, while positioning aspartic acid residues in positions 76 and 131 (C54T/C97A/D72C/R76D/R80C/D127C/V131D/K135C mutant). The fourth mutant, T4L-D, was prepared for the single-arm **C9-Gd** tag and contained cysteine residues at positions 72 and 131 (C54T/C97A/D72C/V131C mutant). Finally, two additional variants of T4L-C were prepared to assess the role of the aspartic acid residue at position $i+4$ in the $i,i+8$ attachment mode. In the mutant T4L-E (C54T/C97A/D72C/R80C/D127C/K135C), the wild-type residues R76 and V131 were retained in the respective $i+4$ positions, while in the mutant T4L-F (C54T/C97A/D72C/R76D/R80C/D127C/K135C) R76 was changed to an aspartic acid but not V131.

For confirmation by a second system, we also prepared mutants of the maltose binding protein. In the following, we refer to the quadruple-mutant S233C/T237C/Y341C/T345C as MBP-A and to the double-mutant T237C/T345C as MBP-B.

Double-arm attachment of the **T1-Gd** and **T2-Gd** tags was confirmed by mass spectrometry. T4L-C showed no evidence for incomplete tagging, whereas mutants T4L-E and T4L-F failed to yield complete double-arm attachment at a level of about 30% and 10%, respectively, illustrating the benefit of an aspartic acid residue in position $i+4$.

ED-EPR spectra

Echo-detected EPR (ED-EPR) and DEER measurements were carried out on 13 different T4L-tag variants and four different MBP-tag constructs. The MBP samples were measured both in the absence and presence of maltose. Figure 2 presents the region of the Gd^{3+} central transition of the W-band ED-EPR spectra of the four different Gd^{3+} tags of Figure 1, following attachment to the T4L mutants T4L-A and T4L-D. The ED-EPR spectra of all other samples prepared in the present work are shown in the Supporting Information (Figures S1 and S2). The **C9-Gd** tag consistently delivered the narrowest central transition, indicating the smallest zero-field splitting (ZFS), while all other tags featured a considerably broader central transition with variations in the detailed line shape between the different tags and mutants. The variations in line shape may arise from different protonation states of the 2-hydroxypropyl pendants, which are also known to produce pH-dependent pseudocontact shifts in NMR experiments with Tb^{3+} complexes.[16,22]

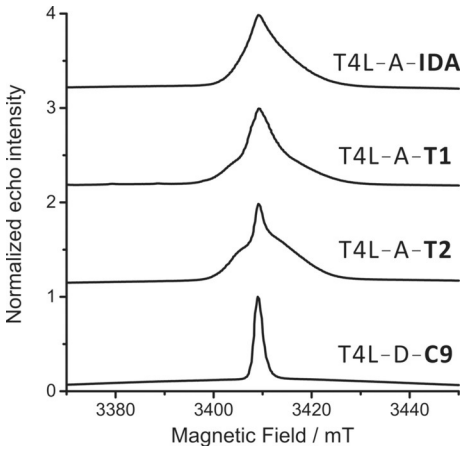


Figure 2. ED-EPR spectra of the central transition region, obtained for the Gd^{3+} ions in different tags attached to the T4 lysozyme mutants T4L-A and T4L-D. The ED-EPR spectra of the tags attached to the other mutants of T4L and MBP were similar and characteristic of the tag rather than the protein. All spectra were normalized to unity and are shifted by 1 for improved visualization and comparison.

Gd^{3+}–Gd^{3+} distance measurements: T4 lysozyme

Figure 3 shows the DEER results obtained for T4L with the four different Gd^{3+} tags used. As expected, based on their small structural differences, the **T1-Gd** and **T2-Gd** tags produced very similar distance distributions for all protein constructs de-

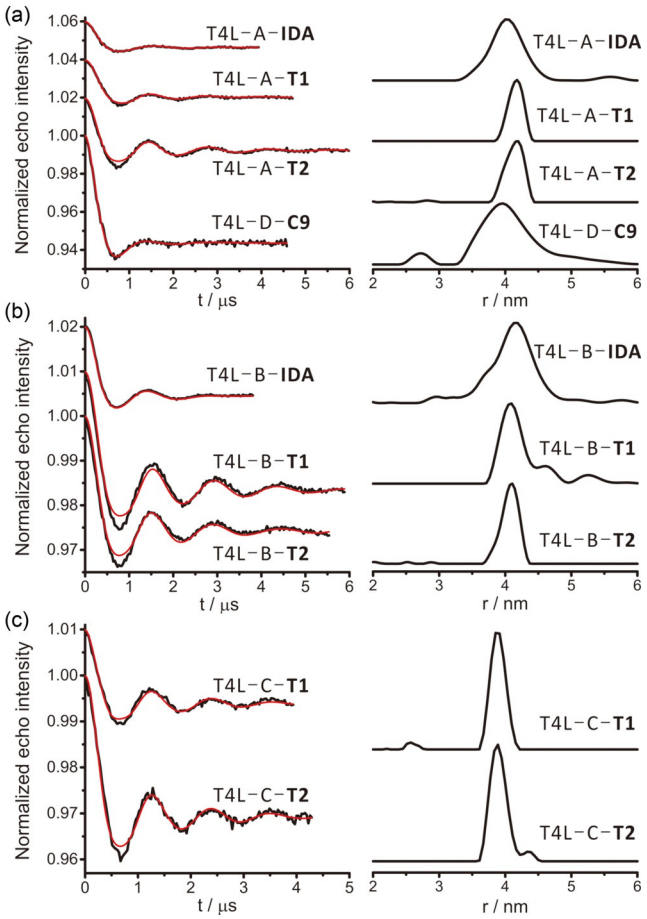


Figure 3. DEER results obtained with different Gd^{3+} tags on T4 lysozyme mutants. a) T4L-A and T4L-D. b) T4L-B. c) T4L-C. Left: Background corrected DEER traces (black) along with the fit obtained with the distance distributions shown on the right (red). The small mismatch between experimental and fitted traces obtained for the **T1-Gd** and **T2-Gd** tags may be attributed to the neglect of the dipolar pseudosecular terms in the data analysis[23] or the uncertainty associated with choosing the correct background decay. The uncorrected traces are shown in Figure S3 in the Supporting Information. Right: Analyzed distance distributions.

signed for double-arm tag attachment. The **IDA-SH** and **C9-Gd** tags produced significantly broader distance distributions with maxima within 0.3 nm of the maxima observed with the **T1-Gd** or **T2-Gd** tags. As the zero-field splitting of the **C9-Gd** tag is small and the central line narrow, distance distributions obtained with this tag are prone to artificial broadening arising from neglecting the pseudo-secular terms of the dipolar interaction in the data analysis.[23] We checked this hypothesis by carrying out relaxation induced dipolar modulation enhancement (RIDME) experiments, which produce dipolar evolution traces that are much less sensitive to the effects of the pseudo-secular terms than the DEER experiments.[24] The RIDME data, recorded for T4L-B with the **T1-Gd** tag and for T4L-D with the **C9-Gd** tag, showed a clear contribution from a second harmonic when analyzed with DeerAnalysis[25] (Figure 4) and also a third weaker peak, the position of which

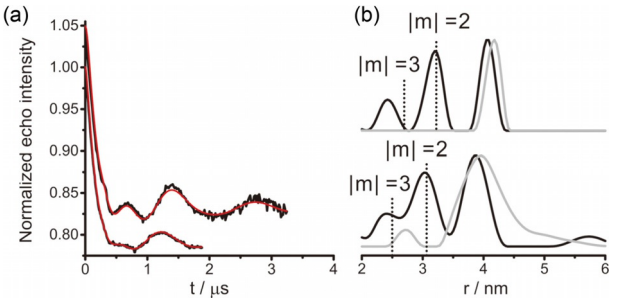


Figure 4. RIDME results obtained with T4L-B-**T1-Gd** and T4L-D-**C9-Gd**. a) Background corrected RIDME traces (black) along with the fit obtained with the distance distributions shown on the right (red). The primary DEER traces are shown in Figure S4 in the Supporting Information. b) Comparison of the distance distributions obtained from the analysis of RIDME (black) and DEER (grey) data. The dotted lines show the positions expected for the second and third harmonics calculated from the main peak.

does not quite match the expected third harmonic and therefore cannot be assigned unambiguously. Most importantly, however, the sample tagged with **C9-Gd** showed a distance distribution for the first harmonic that was significantly narrower than in the DEER experiment (full width at half amplitude of about 0.5 nm versus 0.8 nm in the DEER results), whereas the RIDME and DEER measurements of T4L-B with **T1-Gd** tag showed very similar distance distributions apart from a minor (0.1 nm) shift in the maximum (Figure 4). RIDME measurements of T4L-C with the **T2-Gd** tag did not reveal any narrowing with respect to DEER either (Figure S5 in the Supporting Information). We attribute the conservation in distribution width to the significantly broader EPR line width associated with the **T1-Gd** and **T2-Gd** tags, which allows the dipolar pseudo-secular terms to be neglected without risking the artificial broadening associated with a data analysis carried out under the weak coupling approximation.[23] The comparison between RIDME and DEER data also suggests that the narrow component observed in some of the ED-EPR spectra measured with the **T1-Gd** and **T2-Gd** tags (Figure 2, and Figures S1, and S2 in the Supporting Information) has no significant impact on the distance distribution width.

The larger distribution width observed in the RIDME experiments with the **C9-Gd** versus the **T1-Gd** tag indicates that the single-arm attachment of the **C9-Gd** tag broadens the distance distribution due to its greater conformational flexibility. Disappointingly, the DEER results obtained with the **IDA-SH** tag were always relatively broad, despite a broad central line in the ED-EPR spectrum (Figure 2), which suggests little sensitivity to pseudo-secular terms. The broad distance distributions obtained with the **IDA-SH** tag probably reflect multiple different complexation modes of the Gd^{3+} ion, as indicated by previous NMR experiments.[18]

Except for the samples with the **IDA-SH** tag, in which the precise complexation mode is uncertain, we benchmarked the experimental results against theoretical predictions by modeling the tags on the crystal structures with different rotamer states of the linker segments between the Gd^{3+} complex and protein backbone, and used the rotamer libraries to compute distance distributions. Figure 5 shows that, with a single exception, the maxima of the experimentally determined distance distributions were all within 0.3 nm of the simulated distance distributions. The exception is T4L-B tagged with **T1-Gd** or **T2-Gd**, in which the experimental distance was up to 0.5 nm longer than the distance predicted by modeling. In contrast, the **C9-Gd** tag attached to T4L-D showed a shorter experimental distance than predicted by modeling (Figure 5 a). Excellent agreement, within 0.1 nm, between experimental and predicted distance and distribution width was obtained for the **T1-Gd** and **T2-Gd** tags attached to T4L-C in the *i*,*i*+8 attachment

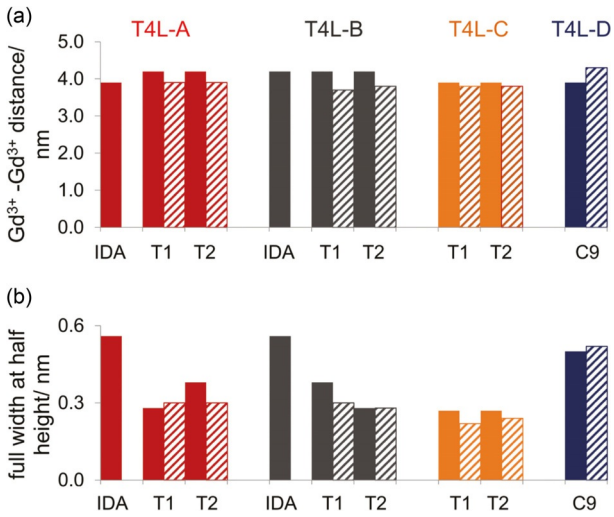


Figure 5. Bar graphs of the experimental and modeled values of distance and width of the distance distributions in different T4 lysozyme mutants. The distance was read from the maximum of a distance distribution and the width is reported as the full width at half-height. Experimental and modeled data are represented by filled and hatched bars, respectively. Distances were modeled using the crystal structure 2LZM.[26] Data are shown for the mutants T4L-A (red), T4L-B (black), T4L-C (orange), and T4L-D (blue). The numerical values are reported in Tables S1 and S2 in the Supporting Information. RIDME results are reported for the **C9-Gd** tag. a) Gd^{3+}–Gd^{3+} distances for the tags indicated below the graph. b) Widths of the distance distributions at half amplitude.

mode. This indicates that a smaller conformational space available for the tag not only delivers narrow distance distributions, but also improves the accuracy with which Gd^{3+}–Gd^{3+} distance distributions can be predicted.

To assess the importance of an aspartic acid residue in position $i+4$ in the $i,i+8$ attachment mode, we also measured DEER and RIDME data of the mutants T4L-E and T4L-F with **T1-Gd** and **T2-Gd** tags. T4L-E contains wild-type residues in both $i+4$ positions, while T4L-F contains an aspartic acid mutation at one of the $i+4$ positions and a wild-type residue (valine) at the other. Both mutants produced consistently broader distance distributions than the T4L-C mutant (Figures S6–S9 in the Supporting Information), highlighting the value of an aspartic acid residue in position $i+4$ when using **T1-Gd** or **T2-Gd** tags in the $i,i+8$ attachment mode on an α-helix.

Gd^{3+}–Gd^{3+} distance measurements: maltose binding protein (MBP)

The DEER results obtained for MBP with and without maltose are shown in Figure 6. The distance distributions were significantly broader than for the T4 lysozyme and of similar width

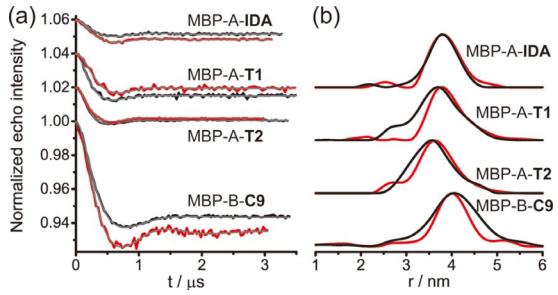


Figure 6. DEER results obtained with different tags on MBP mutants with (red) and without (black) one equivalent of maltose. a) Background corrected DEER traces along with the fit obtained with the distance distributions shown on the right (grey). The primary DEER traces are shown in Figure S10 and corresponding RIDME results in Figure S11 in the Supporting Information. b) Analyzed distance distributions.

for the **IDA-SH**, **T1-Gd**, and **T2-Gd tags**. The **C9-Gd** tag also produced a broad distance distribution and RIDME measurements (shown in Figure S4 in the Supporting Information) suggest that the width is intrinsic and due to a multitude of conformations and not due to pseudo-secular term contributions. The comparison with the calculated distance distributions shows a better agreement between experimental and modeled distances than for T4 lysozyme, with no difference greater than 0.1 nm (Figure 7a). The addition of maltose did not change the distances significantly, but the widths of the distance distributions became narrower (except for the sample tagged with **IDA-SH**, for which the change in distance distribution width was very small). As in the case of T4 lysozyme, the predicted distribution widths were consistently narrower for the double-arm tags than the **C9-Gd** tag, but they notably underestimated the experimental distribution widths for the double-armed **T1-Gd** and **T2-Gd** tags (Figure 7b), although the computations accepted tag conformations with interatomic distances slightly shorter than the sum of van der Waals radii. This result suggests that the MBP domain tagged in the present work is more flexible than the C-terminal domain of T4 lysozyme and that the binding of maltose rigidifies the protein to some degree, but not completely. In the case of the **C9-Gd** tag, the presence of maltose resulted in a more substantial narrowing of the distribution than for any of the double-arm tags. While it is difficult to pinpoint the molecular mechanism underlying this observation, the flexibility of the tether of a single-arm tag evidently makes it more prone to finding a preferential binding site on the protein. Such effects are known[27] and would change the distance distributions.

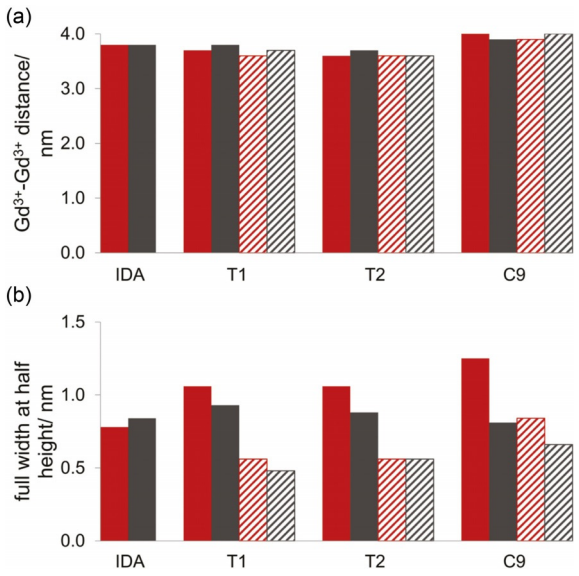


Figure 7. Bar graphs of the experimental and modeled distance and width of the distance distributions in MBP in the absence and presence of maltose. The distance was read from the maximum of a distance distribution and the width is reported as the full width at half-height. Experimental and modeled data are represented by filled and hatched bars, respectively. Data are shown for the mutants MBP-A (S233C/T237C/Y341C/T345C with **IDA-SH**, **T1-Gd** or **T2-Gd** tag) and MBP-B (T237C/T345C with **C9-Gd** tag) without (red) and with maltose (black). Distances were modeled using the crystal structures 1OMP[28] and 1ANF.[29] a) Gd^{3+}–Gd^{3+} distances for the tags indicated below the graph. b) Widths of the distance distributions at half amplitude.

Tag-specific contribution to the widths of distance distributions

The rotamer libraries calculated for the double-arm **T2-Gd** tag show that the $i,i+4$ attachment mode still leaves a wide range of coordinates accessible to the Gd^{3+} ion (Figure 8). The size and shape of the conformational space depends on the steric restrictions imposed by the protein environment. Figure 9 shows that the $i,i+8$ attachment mode restricts the metal coordinates much better, but still retains an arc of accessible coordinates, which arise from a wobbling motion of the tag around the line connecting the two cysteine residues.

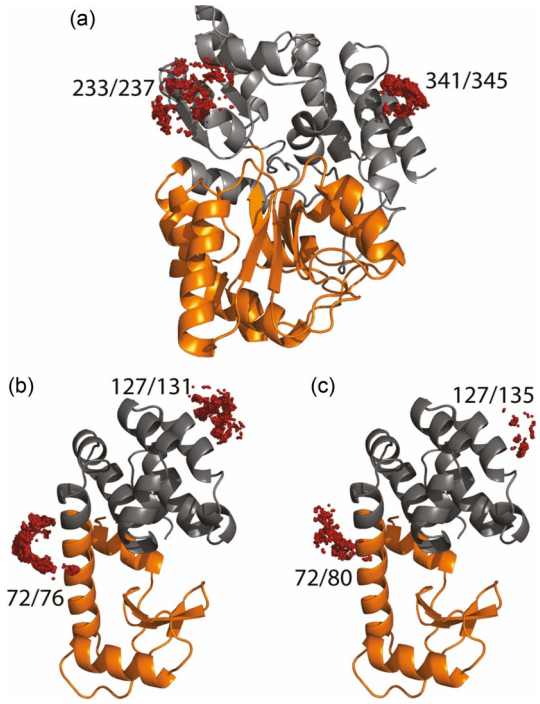


Figure 8. Coordinates found for the Gd^{3+} ion in the **T2-Gd** tag attached to MBP-A, T4L-A and T4L-C. Both proteins assume a two-domain structure, which is highlighted by coloring the N-terminal and C-terminal domains in orange and grey, respectively. The metal positions predicted by modeling are indicated by red balls. The positions of the cysteine residues ligated with the tags are marked with the residue numbers. a) Simulations for MBP-A, using the crystal structure 1OMP[28] without maltose. b) Simulations for T4L-A, where the tag is in the *i*,*i*+4 attachment mode, using the crystal structure 2LZM.[26] c) Simulations for T4L-C, where the tag is in the *i*,*i*+8 attachment mode.

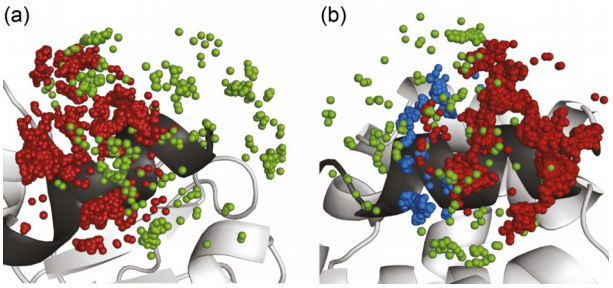


Figure 9. Comparison of Gd^{3+} ion coordinates predicted by modeling of **T2-Gd** and **C9-Gd** tags. Predicted metal positions are indicated by spheres. The helix carrying the tags is shown in dark grey. a) Metal positions predicted for the site C341/C345 in MBP-A tagged with **T2-Gd** (red) and site C345 in MBP-B tagged with **C9-Gd** (green), using the crystal structure 1OMP.[28] b) Metal positions predicted for the site C127/C131 in T4L-A tagged with **T2-Gd** (red), site C127/C135 in T4L-C tagged with **T2-Gd** (blue), and site C131 in T4L-D tagged with **C9-Gd** (green), using the crystal structure 2LZM.[26]

Discussion

The present results show that attaching a Gd^{3+} ion to a protein through covalent linkages to two rather than one cysteine residue is a successful strategy for obtaining narrow Gd^{3+}–Gd^{3+} distance distributions by EPR distance measurements. Importantly, the distance distributions can be predicted with high accuracy by a simple modeling approach that uses the crystal structure and libraries of tag conformations. The comparison of *i*,*i*+4 and *i*,*i*+8 attachment modes shows that the accuracy of the modeling increases when the conformational space available to the tags is reduced. Even in the less restricted *i*,*i*+4 attachment mode, the conformational space of the double-arm **T1/T2-Gd** tags is smaller than that of the single-arm **C9-Gd** tag, for which the tether between the sulfur and the metal ion is shorter by two bonds (Figure 1e). As an additional advantage of the double-arm tags, their ZFS is sufficiently large to justify data analysis using the weak coupling approximation for distances in the vicinity of 4 nm. This is of great practical importance, although it comes at the expense of sensitivity. The sensitivity could be improved by delivering the pump pulse as a broad-band chirp pulse.[30]

The use of double-arm attachments for improved immobilization of paramagnetic centers has been demonstrated before, with a nitroxide tag attached to two cysteine residues for DEER measurements,[15] and a cyclen-based lanthanide tag, **CLaNP-5**, which reacts with two cysteine residues to form the same chemical linkages as the **T1-Gd** and **T2-Gd** tags for NMR experiments.[19] However, the full flexibility of these tags after attachment to a protein has never been explored. The **CLaNP-5** tag was designed to generate pseudocontact shifts (PCS) in proteins for measurement by NMR spectroscopy and it was shown that the PCSs can be predicted quite reliably from the positions of the cysteine residues in the protein structure.[20] Measurement of PCS data is even more sensitive to tag motions than distance measurements, as re-orientational movements of the lanthanide relative to the protein average positive and negative PCSs, greatly decreasing the size of observable PCSs. The **T1-Ln** and **T2-Ln** tags have likewise been shown to produce large PCSs.[16] Due to the chemical nature of the linkers to the cysteine residues, the conformational space of the **CLaNP-5** tag is probably as large as that of the **T1-Gd** and **T2-Gd** tags. Evidence for residual mobility in **CLaNP-5** tags bound to a protein has indeed been obtained by relaxation measurements.[31]

Considering that the **IDA-SH** tag has only two chemical bonds between the sulfur atom and the IDA moiety, two **IDA-SH** tags coordinated to a single Gd^{3+} ion would be expected to yield even better immobilization of the metal ion than the **T1-Gd** and **T2-Gd** tags. It is thus surprising that the widths of the distance distributions obtained with the **IDA-SH** tags were much larger than expected. A possible explanation for this effect may be multiple coordination modes for the metal ion. The modulation depths in the DEER traces obtained with the **IDA-SH** tags, about 60% of those obtained for the **T1-Gd** and **T2-Gd** tag (Figures 3a,b, and 6a), are somewhat shallower than expected from the differences in ED-EPR line widths, suggesting a lower binding affinity of the metal ion. To improve the binding affinity, we also attempted the attachment of **NTA-SH** tags,[32] but consistently failed to install more than three tag molecules in the same protein sample (data not shown), possibly due to electrostatic repulsion between two negatively charged NTA moieties in close proximity. Quite generally, the **IDA-SH** tags are more difficult to use than the cyclen-based

tags, because they need to be titrated with Gd^{3+} ions after their ligation to the target protein. In view of the small sample volumes associated with DEER measurements at W-band, it is easy to over- or under-titrate the samples. In the present work, the **IDA-SH** tags did not deliver distribution widths much smaller than 0.6 nm, whereas we achieved distribution widths below 0.3 nm with the **T1-Gd** and **T2-Gd** tags. These widths approach those reported previously for the nitroxide double-arm **RX** tag attached to two different proteins (0.15–0.35 nm), in which no attempts were made to model the distance distributions obtained.[15] They are narrower even than those reported from DEER measurements of chemically synthesized rigid Gd^{3+}–Gd^{3+} rulers (which ranged from 0.32–0.48 nm for distances of 3.1–4.8 nm)[23] and close to the width of 0.28 nm measured by RIDME for another synthetic Gd^{3+}–Gd^{3+} ruler, in which the Gd^{3+} ions are separated by only 2.35 nm.[24]

Distribution widths are governed both by protein dynamics and tag flexibility. While the tag's contribution to the distribution width can readily be assessed by modeling of a rotamer library, assessing the range of protein conformations is less straightforward as molecular dynamics runs are sensitive to the accuracy of the force fields used. In the present work, we positioned the tags in single domains of two-domain proteins (Figure 8). In the case of T4 lysozyme, many crystal structures have been determined, which indicate that the C-terminal domain is structurally conserved even when the protein undergoes a hinge-bending motion that changes the relative position of the N-terminal relative to the C-terminal domain. Indeed, exceptionally narrow distance distributions were obtained with the double-arm tags. In the present work, we used the crystal structure 2LZM of the wild-type protein to model the tag conformations. Using instead the crystal structures 150L and 1SSY of multi-site mutants of T4 lysozyme,[33,34] which have previously been proposed to present better representations of the global conformation of the enzyme in solution,[35,36] the distances simulated with the **C9-Gd** tag varied by ± 0.1 nm. Crystal packing effects could thus explain small discrepancies between the experimental and predicted distances.

Similar to T4 lysozyme, MBP is composed of two distinct globular domains, the N-domain and C-domain, which are connected by a flexible hinge region. The base of the groove between the two domains forms the maltodextrin binding site. Crystal structures indicate that MBP undergoes a conformational transition, in which the domains move from an open to a closed state upon ligand binding.[28] Apo-MBP has been reported to exist 95% in an open state, with 5% populating a partially closed state.[37] Although we placed both tagging sites in the C-terminal domain, the measured distance distributions were significantly broader compared to those obtained for T4 lysozyme (Figures 3 and 5) and broader than predicted from the tag conformations modeled on the crystal structure. This indicates conformational flexibility of MBP and suggests that the distance distribution widths obtained with the **T1/T2-Gd** tags can be used to probe such minor intradomain conformation changes. This conclusion is further supported by the observation that addition of maltose led to slightly narrower distance distributions, indicating reduced intradomain mobility in the closed compared to the open state.

Conclusion

The attachment of the cyclen-based Gd^{3+} tags **T1-Gd** and **T2-Gd** to two cysteine residues improves the localization of the metal ion and delivers exceptionally narrow Gd^{3+}–Gd^{3+} distance distribution widths in DEER experiments. Both the distances and the widths of the distance distributions are readily predicted by modeling rotamer libraries of the tags on crystal structures, especially when the tags are attached to cysteines in positions i and $i+8$ of α-helices and an aspartic acid residue is present in position $i+4$. This $i,i+8$ attachment mode delivered the narrowest distributions of Gd^{3+}–Gd^{3+} distances reported to date for a protein. Limiting the contribution of tag mobility to the distance distributions opens the door to probing the conformational variability of proteins by DEER and auxiliary RIDME experiments with unprecedented accuracy at the high magnetic field strengths required for measurements on small amounts of sample.

Experimental Section

Protein expression and purification: The quadruple-mutant S233C/T237C/Y341C/T345C (MBP-A) and the double-mutant T237C/T345C (MBP-B) of MBP were cloned into pETMCSIII vectors[38] with a N-terminal His_6-tag followed by a tobacco etch virus (TEV) protease recognition site. The protein samples were produced from PCR-amplified DNA by cell-free protein synthesis (CFPS).[39] The CFPS reaction was carried out at 30 °C for 16 hours in dialysis mode according to a published protocol.[40] The proteins were purified using a 1 mL Co-NTA gravity column (GE Healthcare, USA), following the manufacturer's protocol, and the His_6 tag was removed by incubation with His_6-TEV protease in buffer A (50 mM MES-KOH, pH 6.7, 2 mM β-mercaptoethanol, 500 mM NaCl) at 4 °C for 16 h. MBP was separated from TEV protease by passing again through the Co-NTA gravity column. Finally, the proteins were dialysed against buffer B (50 mM MES-KOH, pH 6.7, 1 mM DTT (1,4-dithiothreitol)) at 4 °C and concentrated using an Amicon ultracentrifugation centrifugal tube with a molecular weight cutoff (MWCO) of 10 kDa. The protein yields were 1.5 mg of the quadruple-cysteine mutant and 1 mg of the double-cysteine mutant per mL of cell-free reaction mixture.

The cysteine mutants C54T/C97A/D72C/R76C/D127C/V131C (T4L-A), C54T/C97A/D72C/R76C/V131C/K135C (T4L-B), C54T/C97A/D72C/R76D/R80C/D127C/V131D/K135C (T4L-C), C54T/C97A/D72C/V131C (T4L-D), C54T/C97A/D72C/R80C/D127C/K135C (T4L-E), and C54T/C97A/D72C/R76D/R80C/D127C/K135C (T4L-F) of T4 lysozyme were cloned into the pETMCSIII vector with a N-terminal His_6-tag, expressed by CFPS from PCR-amplified DNA, purified with a Co-NTA column and dialysed against buffer C (50 mM MES-KOH, pH 6.7, 1 mM DTT) and concentrated under the same conditions as described for the MBP mutants. The protein yields ranged between 1 mg and 1.5 mg per mL of cell-free reaction mixture.

Synthesis of the tags: The **C9-Gd**, **T1-Gd**, **T2-Gd** and activated **IDA-SH** tags were synthesized as described previously.[14,16,18]

Protein ligation with C9-Gd, T1-Gd, T2-Gd and IDA Tags: To ensure that all cysteine residues were reduced, DTT was added to a 0.1 mM solution of protein in buffer C to a final concentration of

5 mM DTT. Excess DTT was removed using Amicon ultracentrifugation tubes (MWCO 10 kDa). The reduced protein was added slowly into 5 equivalents of **C9-Gd**, **T1-Gd**, **T2-Gd** or 10 equivalents of **IDA-SH** tag in the same buffer. Reaction mixtures were shaken overnight at room temperature. The completion of the ligation reactions was confirmed by mass spectrometry.

DEER sample preparation: The tagged MBP mutants were concentrated and exchanged into an EPR buffer (50 mM MES-KOH in D_2O, pD 6.7, uncorrected pH meter reading) using an Amicon ultracentrifugation tube (MWCO 10 kDa). Similarly, the tagged T4L mutants were concentrated and exchanged into a buffer containing 50 mM Tris-HCl in D_2O, pD 7.5 (uncorrected pH meter reading). Perdeuterated glycerol was added to reach a 20% (v/v) final composition. $GdCl_3$ dissolved in D_2O was added to the samples tagged with the **IDA-SH** tag in 1:2 molar ratio of $GdCl_3$ to conjugated **IDA-SH** tag.

DEER and RIDME measurements: All EPR measurements were carried out on a home-built W-band spectrometer[41,42] at 10 K. Echo-detected EPR (ED-EPR) spectra were recorded with $\pi/2$ and π pulse durations of 15 and 30 ns (adjusted on the maximum of the spectrum), respectively, with an echo delay of 550 ns and a repetition time of 1 ms.

DEER measurements were recorded using the standard four-pulse DEER sequence.[43] The frequency for pump pulses was set to the maximum of the Gd^{3+} EPR spectrum, and the observer pulses 100 MHz higher. The pump pulse duration was 15 ns, and the observer pulses were 15 and 30 ns, respectively. The delay time, τ, was 400 ns, the pump pulse timing, t, was stepped with an increment of 30 ns, and the repetition time was 800 μs. An eight-step phase cycle was employed to remove instrumental artifacts and to compensate for DC offset. The accumulation time ranged from 2 to 10 h.

RIDME traces were recorded using the 5-pulse dead-time free sequence $\pi/2-\tau_1-\pi-\tau_1+t-\pi/2-T_{mix}-\pi/2-\tau_2-t-\pi-\tau_2$ echo.[44] The receiver was optimized to detect a maximal positive signal of the echo generated by the pulses on channel 1, while the 90° phase shift of the $\pi/2$ pulse on channel 2 was determined by producing a symmetric dispersion-like echo shape. The magnetic field was set to the maximum of the Gd^{3+} EPR spectrum. The delays τ_1 and τ_2 were set to 400 ns and 2.5 μs, respectively, and the variable delay t was stepped with an increment of 10 ns starting from −300 ns. T_{mix} was 50 μs. Each trace was accumulated in about 1–2 h.

For all measurements, transients were collected for each t value and echo-integration was carried out post-measurement, usually integrating the echo region at half height. The DEER and RIDME data were analyzed using the program DeerAnalysis 2015.[45] Distance distributions were obtained using Tikhonov regularization. The regularization parameter α was chosen according to the L curve criterion, resulting in values of $\alpha=1$ for the DEER trace of all the **T1**- and **T2**-labeled mutants, $\alpha=10$ for the DEER trace of all the **IDA-SH**- and **C9**-labeled mutants, and $\alpha=10$ for all the RIDME traces.

Modeling: The crystal structures of apo-MBP (PDB ID: 1OMP),[28] holo-MBP (PDB ID: 1ANF),[29] and T4 lysozyme (PDB ID: 2LZM)[26] were used to model the distance distributions.

The coordinates of the **C9-Gd** tag were crafted onto each of the cysteine residues at positions 237 and 345 of MBP and 72 and 131 of T4 lysozyme and the Gd-Gd distance distributions modeled using PyParaTools[46] as described previously,[14] except that the widths of the distance distributions were broadened by allowing a greater range of dihedral angles for the C–S bonds (namely ±10° around the rotamer states at −60°, +60°, −90°, +90°) and defining steric clashes as interatomic distances less than 0.9 times the sum of van der Waals radii.

To model the distance distributions obtained with the **T1-Gd** and **T2-Gd** tags, tags were crafted onto each cysteine residue by a single arm and rotamer libraries of the tags were generated as described above for the **C9-Gd** tag, taking into account additional rotatable bonds in the tether to the cysteine sulfur. To establish valid conformations of double-arm attachment, the coordinates of the metal ion and cyclen nitrogen atoms were extracted from the libraries of two neighboring tag molecules, compared in a pairwise manner and accepted as a valid representation of the double-arm tag if the coordinates of the Gd^{3+} ion and cyclen nitrogen atoms superimposed within 1 Å.

No distance distributions were simulated for the proteins with **IDA-SH** tags, because a lanthanide ion has up to 12 coordination sites and NMR experiments indicate that a number of different bis-IDA coordination species are formed rather than a single complex.[18] The structures of these complexes are currently unknown.

Acknowledgements

Financial support by the Australian Research Council is gratefully acknowledged. D. G. acknowledges the support of the Israel Science Foundation (ISF grant No. 334/14). This research was made possible in part by the historic generosity of the Harold Perlman Family (D.G.). D.G. holds the Erich Klieger Professorial Chair in Chemical Physics.

Conflict of interest

The authors declare no conflict of interest.

Keywords: double electron-electron resonance · EPR spectroscopy · gadolinium · protein structures

[1] G. Jeschke, *Annu. Rev. Phys. Chem.* **2012**, *63*, 419–446.
[2] G. Jeschke, Y. Polyhach, *Phys. Chem. Chem. Phys.* **2007**, *9*, 1895–1910.
[3] A. M. Raitsimring, C. Gunanathan, A. Potapov, I. Efremenko, J. M. Martin, D. Milstein, D. Goldfarb, *J. Am. Chem. Soc.* **2007**, *129*, 14138–14139.
[4] A. Feintuch, G. Otting, D. Goldfarb, *Methods Enzymol.* **2015**, *563*, 415–457.
[5] D. Goldfarb, *Phys. Chem. Chem. Phys.* **2014**, *16*, 9685–9699.
[6] A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting, D. Goldfarb, *J. Am. Chem. Soc.* **2010**, *132*, 9040–9048.
[7] A. Martorana, G. Bellapadrona, A. Feintuch, E. Di Gregorio, S. Aime, D. Goldfarb, *J. Am. Chem. Soc.* **2014**, *136*, 13458–13465.
[8] M. Qi, A. Gross, G. Jeschke, A. Godt, M. Drescher, *J. Am. Chem. Soc.* **2014**, *136*, 15366–15378.
[9] F. C. Mascali, H. Y. V. Ching, R. M. Rasia, S. Un, L. C. Tabares, *Angew. Chem. Int. Ed.* **2016**, *55*, 11041–11043; *Angew. Chem.* **2016**, *128*, 11207–11209.
[10] F. X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb, P. Selenko, *Nature* **2016**, *530*, 45–50.
[11] Y. Yang, F. Yang, Y. J. Gong, J. L. Chen, D. Goldfarb, X.-C. Su, *Angew. Chem. Int. Ed.* **2017**, *56*, 2914–2918; *Angew. Chem.* **2017**, *129*, 2960–2964.
[12] B. Graham, C. T. Loh, J. D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, G. Pintacuda, T. Huber, G. Otting, *Bioconjugate Chem.* **2011**, *22*, 2118–2125.
[13] H. Yagi, D. Banerjee, B. Graham, T. Huber, D. Goldfarb, G. Otting, *J. Am. Chem. Soc.* **2011**, *133*, 10418–10421.

[14] E. H. Abdelkader, M. D. Lee, A. Feintuch, M. R. Cohen, J. D. Swarbrick, G. Otting, B. Graham, D. Goldfarb, *J. Phys. Chem. Lett.* **2015**, *6*, 5016 – 5021.
[15] M. R. Fleissner, M. D. Bridges, E. K. Brooks, D. Cascio, T. Kálai, K. Hideg, W. L. Hubbell, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16241 – 16246.
[16] M. D. Lee, M. L. Dennis, J. D. Swarbrick, B. Graham, *Chem. Commun.* **2016**, *52*, 7954 – 7957.
[17] J. D. Swarbrick, P. Ung, X.-C. Su, A. Maleckis, S. Chhabra, T. Huber, G. Otting, B. Graham, *Chem. Commun.* **2011**, *47*, 7368 – 7370.
[18] J. D. Swarbrick, P. Ung, M. L. Dennis, M. D. Lee, S. Chhabra, B. Graham, *Chem. Eur. J.* **2016**, *22*, 1228 – 1232.
[19] P. H. J. Keizers, J. F. Desreux, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.* **2007**, *129*, 9292 – 9293.
[20] P. H. J. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.* **2008**, *130*, 14802 – 14812.
[21] M. Matsumura, B. W. Matthews, *Science* **1989**, *243*, 792 – 794.
[22] M. D. Lee, C.-T. Loh, J. Shin, S. Chhabra, M. L. Dennis, G. Otting, J. D. Swarbrick, B. Graham, *Chem. Sci.* **2015**, *6*, 2614 – 2624.
[23] A. Dalaloyan, M. Qi, S. Ruthstein, S. Vega, A. Godt, A. Feintuch, D. Goldfarb, *Phys. Chem. Chem. Phys.* **2015**, *17*, 18464 – 18476.
[24] A. Collauto, V. Frydman, M. D. Lee, E. H. Abdelkader, A. Feintuch, J. D. Swarbrick, B. Graham, G. Otting, D. Goldfarb, *Phys. Chem. Chem. Phys.* **2016**, *18*, 19037 – 19049.
[25] S. Razzaghi, M. Qi, A. I. Nalepa, A. Godt, G. Jeschke, A. Savitsky, M. Yulikov, *J. Phys. Chem. Lett.* **2014**, *5*, 3970 – 3975.
[26] L. H. Weaver, B. W. Matthews, *J. Mol. Biol.* **1987**, *193*, 189 – 199.
[27] E. H. Abdelkader, A. Feintuch, X. Yao, L. A. Adams, L. Aurelio, B. Graham, D. Goldfarb, G. Otting, *Chem. Commun.* **2015**, *51*, 15898 – 15901.
[28] A. J. Sharff, L. E. Rodseth, J. C. Spurlino, F. A. Quiocho, *Biochemistry* **1992**, *31*, 10657 – 10663.
[29] F. A. Quiocho, J. C. Spurlino, L. E. Rodseth, *Structure* **1997**, *5*, 997 – 1015.
[30] A. Doll, M. Qi, N. Wili, S. Pribitzer, A. Godt, G. Jeschke, *J. Magn. Reson.* **2015**, *259*, 153 – 162.
[31] M. A. S. Hass, P. H. J. Keizers, A. Blok, Y. Hiruma, M. Ubbink, *J. Am. Chem. Soc.* **2010**, *132*, 9952 – 9953.
[32] H. Yagi, A. Maleckis, G. Otting, *J. Biomol. NMR* **2013**, *55*, 157 – 166.
[33] M. M. He, Z. A. Wood, W. A. Baase, H. Xiao, B. W. Matthews, *Protein Sci.* **2009**, *13*, 2716 – 2724.
[34] S. Daopin, T. Alber, W. A. Baase, J. A. Wozniak, B. W. Matthews, *J. Mol. Biol.* **1991**, *221*, 647 – 667.
[35] N. K. Goto, N. R. Skrynnikov, F. W. Dahlquist, L. E. Kay, *J. Mol. Biol.* **2001**, *308*, 745 – 764.
[36] J.-L. Chen, Y. Yang, L.-L. Zhang, H. Liang, T. Huber, X.-C. Su, G. Otting, *Phys. Chem. Chem. Phys.* **2016**, *18*, 5850 – 5859.
[37] C. Tang, C. D. Schwieters, G. M. Clore, *Nature* **2007**, *449*, 1078 – 1082.
[38] C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love, N. E. Dixon, *Biochemistry* **2000**, *39*, 11989 – 11999.
[39] P. S. C. Wu, K. Ozawa, S. P. Lim, S. G. Vasudevan, N. E. Dixon, G. Otting, *Angew. Chem. Int. Ed.* **2007**, *46*, 3356 – 3358; *Angew. Chem.* **2007**, *119*, 3420 – 3422.
[40] M. A. Apponyi, K. Ozawa, N. E. Dixon, G. Otting in *Methods in Molecular Biology Vol. 426—Structural Proteomics: High-Throughput Methods* (Eds.: B. Kobe, M. Guss, T. Huber), Humana Press, Totowa (USA), **2008**, pp. 257 – 268.
[41] D. Goldfarb, Y. Lipkin, A. Potapov, Y. Gorodetsky, B. Epel, A. M. Raitsimring, M. Radoul, I. Kaminker, *J. Magn. Reson.* **2008**, *194*, 8 – 15.
[42] F. Mentink-Vigier, A. Collauto, A. Feintuch, I. Kaminker, V. Tarle, D. Goldfarb, *J. Magn. Reson.* **2013**, *236*, 117 – 125.
[43] M. Pannier, S. Veit, A. Godt, G. Jeschke, H. W. Spiess, *J. Magn. Reson.* **2000**, *142*, 331 – 340.
[44] S. Milikisyants, F. Scarpelli, M. G. Finiguerra, M. Ubbink, M. Huber, *J. Magn. Reson.* **2009**, *201*, 48 – 56.
[45] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger, H. Jung, *Appl. Magn. Reson.* **2006**, *30*, 473 – 498.
[46] M. Stanton-Cook, X.-C. Su, G. Otting, T. Huber, http://compbio.anu.edu.au/mscook/PPT.

Manuscript received: June 1, 2017
Accepted manuscript online: July 9, 2017
Version of record online: August 3, 2017

CHEMISTRY
A European Journal

Supporting Information

Double-Arm Lanthanide Tags Deliver Narrow Gd^{3+}–Gd^{3+} Distance Distributions in Double Electron–Electron Resonance (DEER) Measurements

Adarshi P. Welegedara[+],[a] Yin Yang[+],[b] Michael D. Lee[+],[c] James D. Swarbrick,[c] Thomas Huber,[a] Bim Graham,*[c] Daniella Goldfarb,*[b] and Gottfried Otting*[a]

chem_201702521_sm_miscellaneous_information.pdf

Table S1. Gd^{3+}–Gd^{3+} distances in MBP and T4 lysozyme mutants

		Gd^{3+}-Gd^{3+} distance/nm				
		MBP-A[a]		T4L-A[c]	T4L-B[d]	T4L-C[e]
		- maltose	+ maltose			
IDA-SH	experimental	3.8	3.8	3.9	4.2	
T1-Gd	experimental	3.7	3.8	4.2	4.2[g] 4.1[h]	3.9
	predicted	3.6	3.7	3.9	3.7	3.8
T2-Gd	experimental	3.6	3.7	4.2	4.2	3.9[g] 3.9[h]
	predicted	3.6	3.6	3.9	3.8	3.8
		MBP-B[b]			T4L-D[f]	
C9	experimental	4.0	3.9		4.0[g] 3.9[h]	
	predicted	3.9	4.0		4.3	

[a] MBP-A: quadruple mutant S233C/T237C/Y341C/T345C

[b] MBP-B: double mutant T237C/T345C

[c] T4L-A: mutant C54T/C97A/D72C/R76C/D127C/V131C

[d] T4L-B: mutant C54T/C97A/D72C/R76C/V131C/K135C

[e] T4L-C: mutant C54T/C97A/D72C/R76D/R80C/D127C/V131D/K135C

[f] T4L-D: mutant C54T/C97A/D72C/V131C

[g] from DEER experiment

[h] from RIDME experiment

Table S2. Experimental and predicted widths of DEER distance distributions produced by the Gd^{3+} tags attached to T4 lysozyme and MBP mutants

		full width at half height/nm				
		MBP-A[a]		T4L-A[c]	T4L-B[d]	T4L-C[e]
		- maltose	+ maltose			
IDA-SH	experimental	0.78	0.84	0.56	0.56	
T1-Gd	experimental	1.06	0.93	0.28	0.38[g] 0.38[h]	0.27
	predicted	0.56	0.48	0.30	0.30	0.22
T2-Gd	experimental	1.06	0.88	0.38	0.28	0.27[g] 0.25[h]
	predicted	0.56	0.56	0.30	0.28	0.24
		MBP-B[b]			T4L-D[f]	
C9	experimental	1.25	0.81		0.79[g] 0.49[h]	
	predicted	0.84	0.66		0.52	

[a] MBP-A: quadruple mutant S233C/T237C/Y341C/T345C

[b] MBP-B: double mutant T237C/T345C

[c] T4L-A: mutant C54T/C97A/D72C/R76C/D127C/V131C

[d] T4L-B: mutant C54T/C97A/D72C/R76C/V131C/K135C

[e] T4L-C: mutant C54T/C97A/D72C/R76D/R80C/D127C/V131D/K135C

[f] T4L-D: mutant C54T/C97A/D72C/V131C

[g] from DEER experiment

[h] from RIDME experiment

Table S3. Gd^{3+}–Gd^{3+} distances and widths of distance distributions produced by the **T1-Gd** and **T2-Gd** tags in the *i*,*i*+8 attachment mode

	T4 lysozyme mutant					
	T4L-C[a]		T4L-E[b]		T4L-F[c]	
	T1-Gd	**T2-Gd**	**T1-Gd**	**T2-Gd**	**T1-Gd**	**T2-Gd**
Gd^{3+}–Gd^{3+} distance/nm	3.9[d]	3.9[d]/3.9[e]	3.9[d]/3.8[e]	4.0[d]	4.1[d]/4.0[e]	4.1[d]
full width at half amplitude/nm	0.27[d]	0.27[d]/0.25[e]	0.60[d]/0.43[e]	0.70[d]	0.64[d]/0.44[e]	0.72[d]

[a] T4L-C: mutant C54T/C97A/D72C/R76D/R80C/D127C/V131D/K135C

[b] T4L-E: mutant C54T/C97A/D72C/R80C/D127C/K135C

[c] T4L-F: mutant C54T/C97A/D72C/R76D/R80C/D127C/K135C

[d] from DEER experiment

[e] from RIDME experiment

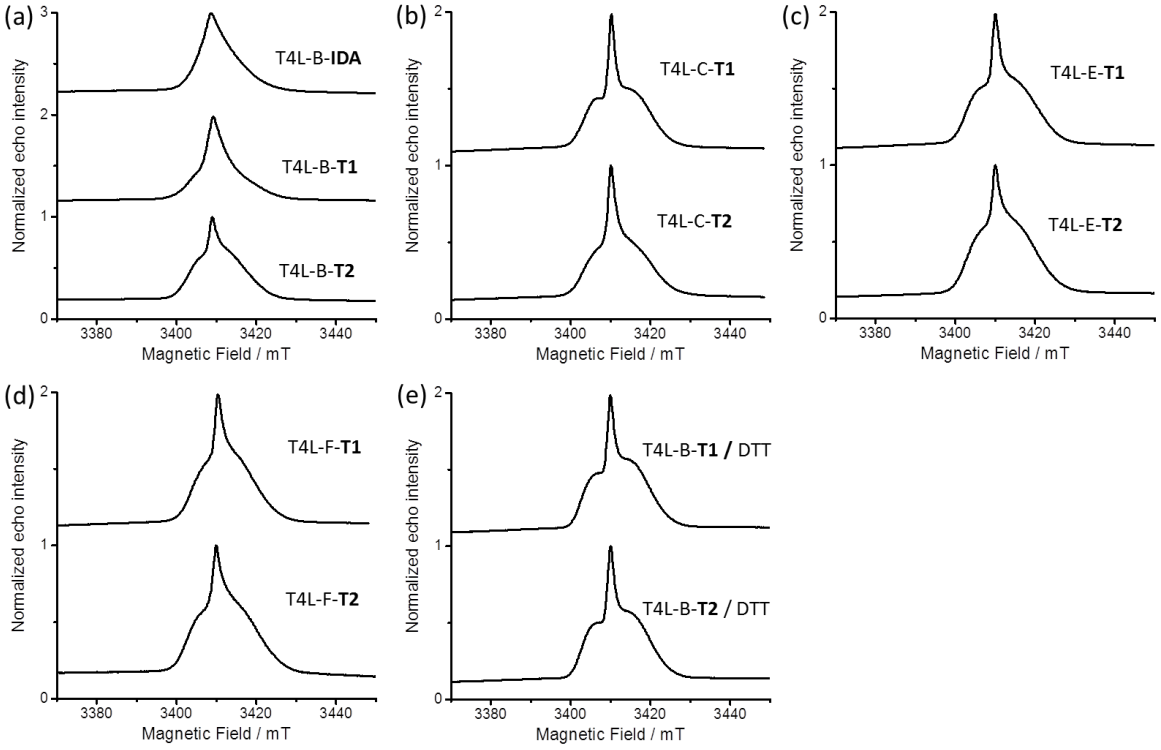


Figure S1. ED-EPR spectra of the region of the central transition region obtained for the **T1-Gd** and **T2-Gd** tags associated with the T4 lysozyme mutants (a) T4L-B, (b) T4L-C, (c) T4L-E, and (d) T4L-F. Panel (e) presents the spectra of the T4L-B mutants after treatment with 20 equivalents DTT at room temperature for 30 minutes before freezing and measurement, in order to reduce the S-S bond conjugation to the protein and release the tag. These spectra thus represent free tag and, as expected, are practically the same for the **T1-Gd** and **T2-Gd** tag. Comparison with the other panels shows that the spectral line shape of the tags is more or less conserved between free and bound tag, except for small differences that depend on the specific mutant and chirality of the tag. All spectra were normalized to unity and are shifted by 1 for improved visualization and comparison.

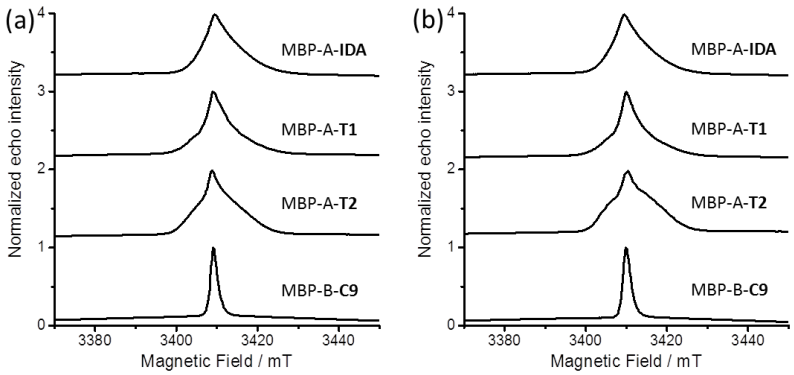


Figure S2. ED-EPR spectra of the region of the central transition obtained for the Gd^{3+} ions in different tags attached to the MBP mutants.

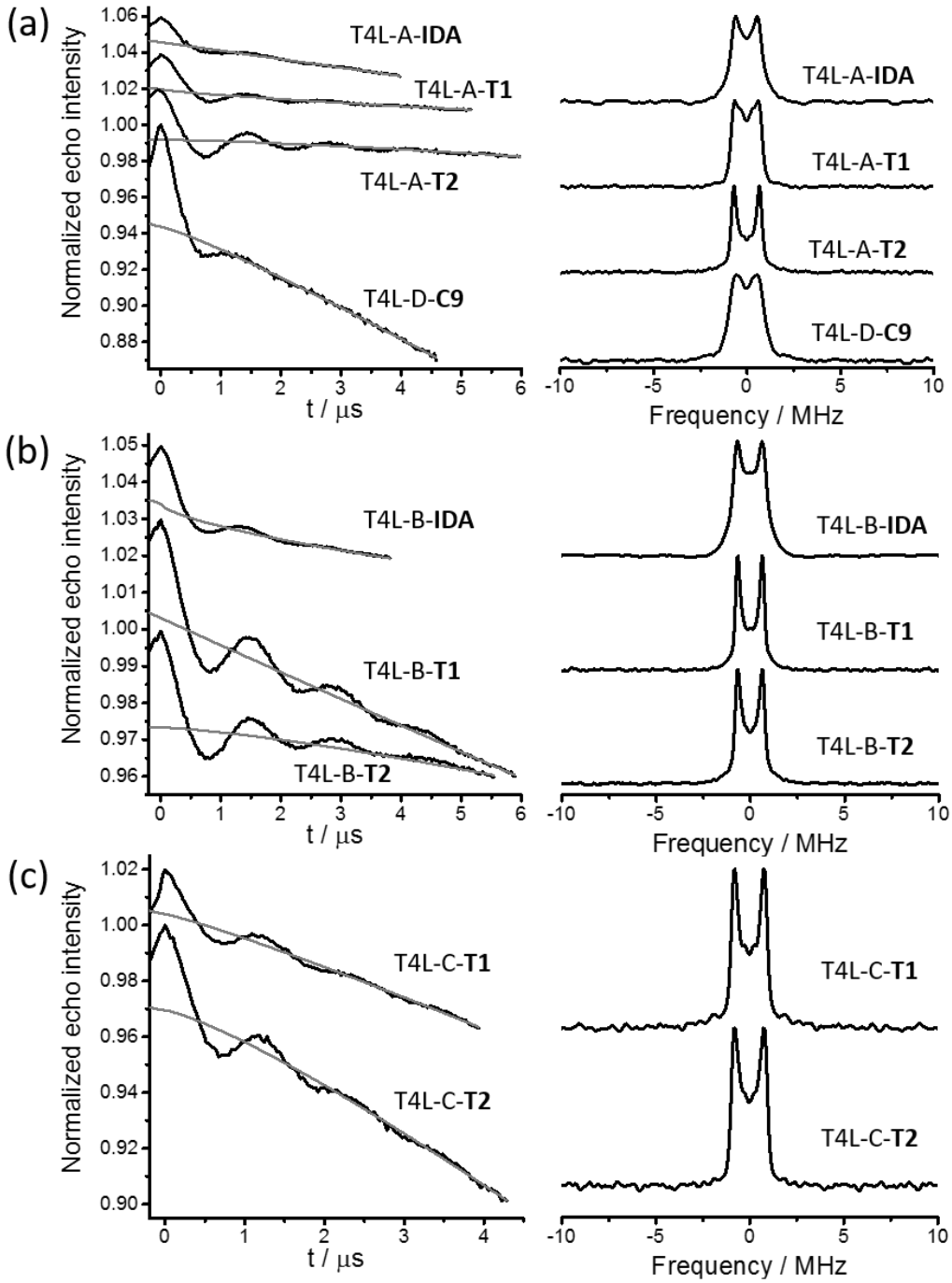


Figure S3. Primary DEER traces and Fourier transformed spectra obtained with different Gd^{3+} tags on T4 lysozyme mutants. (a) T4L-A and T4L-D. (b) T4L-B. (c) T4L-C. (Left) Primary DEER traces (black) and the background function used (grey). (Right) Corresponding data after background subtraction and Fourier transformation.

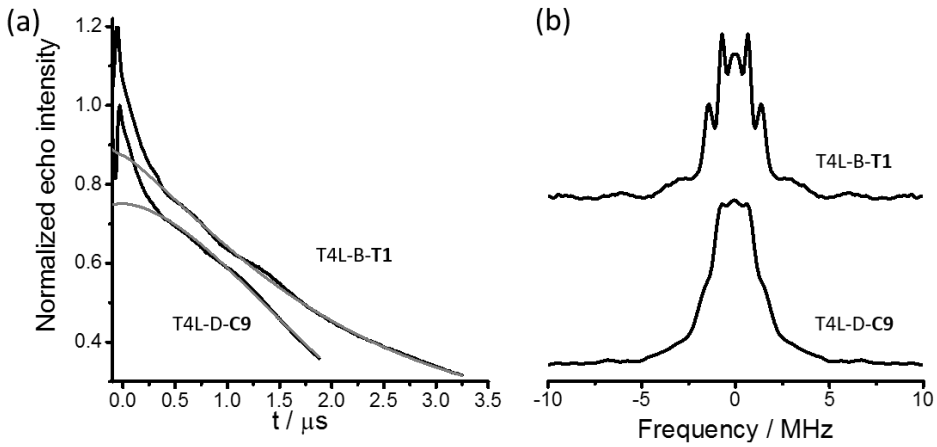


Figure S4. RIDME results of T4L-B-**T1** and T4L-D-**C9**. (a) Primary RIDME traces. The fitted background decays are indicated by grey lines. (b) Corresponding data after background subtraction and Fourier transformation.

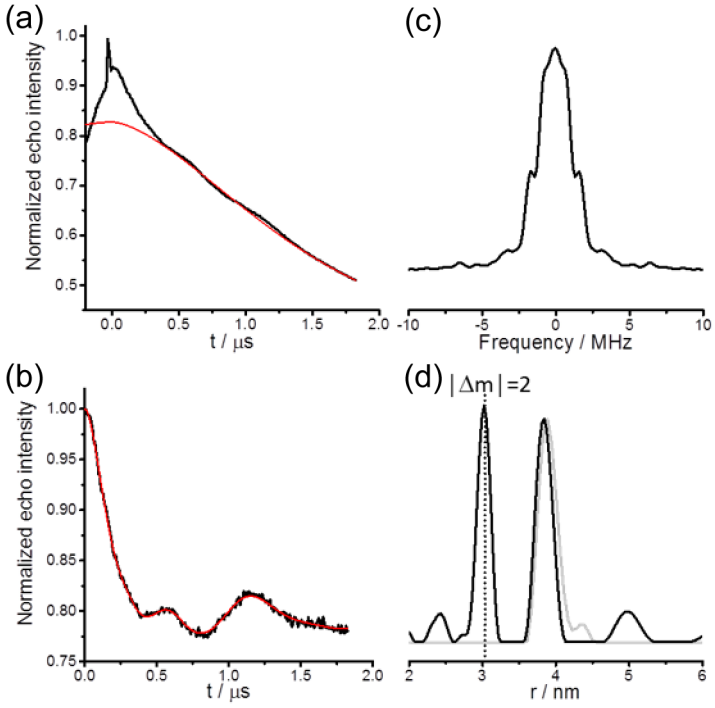


Figure S5. RIDME results of T4L-C-**T2**. (a) Primary data along with the background correction function used in red. (b) Background corrected RIDME trace (black) along with the fit obtained with the distance distributions shown in (d) (red). (c) Fourier transform of the experimental trace in (b). (d) Comparison of the distance distributions obtained from the analysis of RIDME (black) and DEER (grey) data. The dotted line shows the position expected for the second harmonic calculated from the main peak.

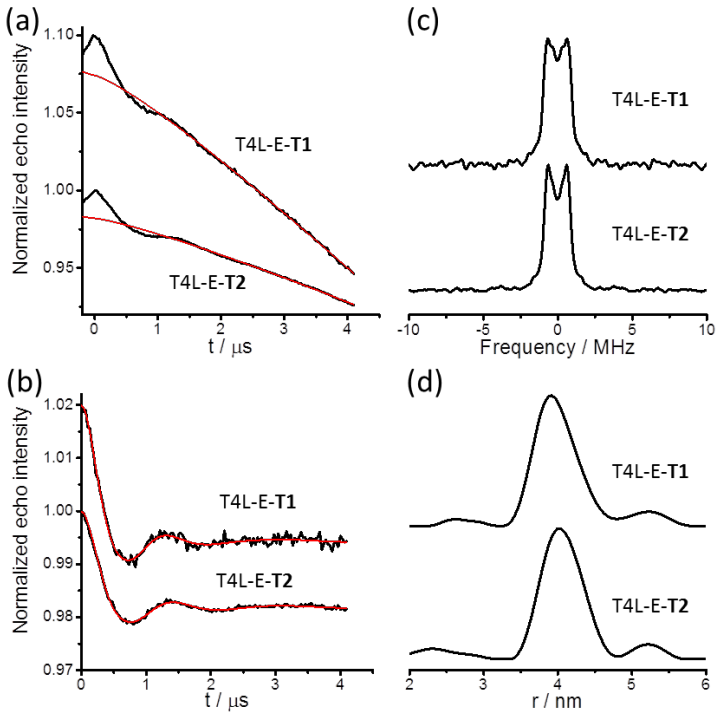


Figure S6. DEER results of T4L-E-**T1** and T4L-E-**T2**. (a) Primary data along with the background correction function used in red. (b) Background corrected DEER traces (black) along with the fit obtained with the distance distributions shown in (d) (red). (c) Fourier transform of the experimental traces in (b). (d) Distance distributions obtained from the analysis shown in (b).

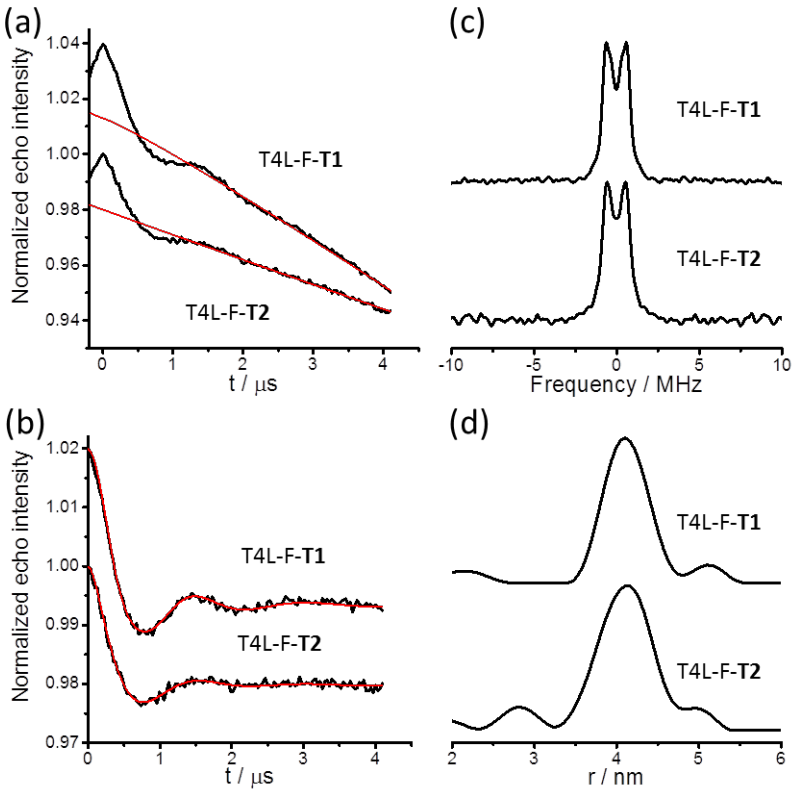


Figure S7. DEER results of T4L-F-**T1** and T4L-F-**T2**. (a) Primary data along with the background correction function used in red. (b) Background corrected DEER traces (black) along with the fit obtained with the distance distributions shown in (d) (red). (c) Fourier transform of the experimental traces in (b). (d) Distance distributions obtained from the analysis shown in (b).

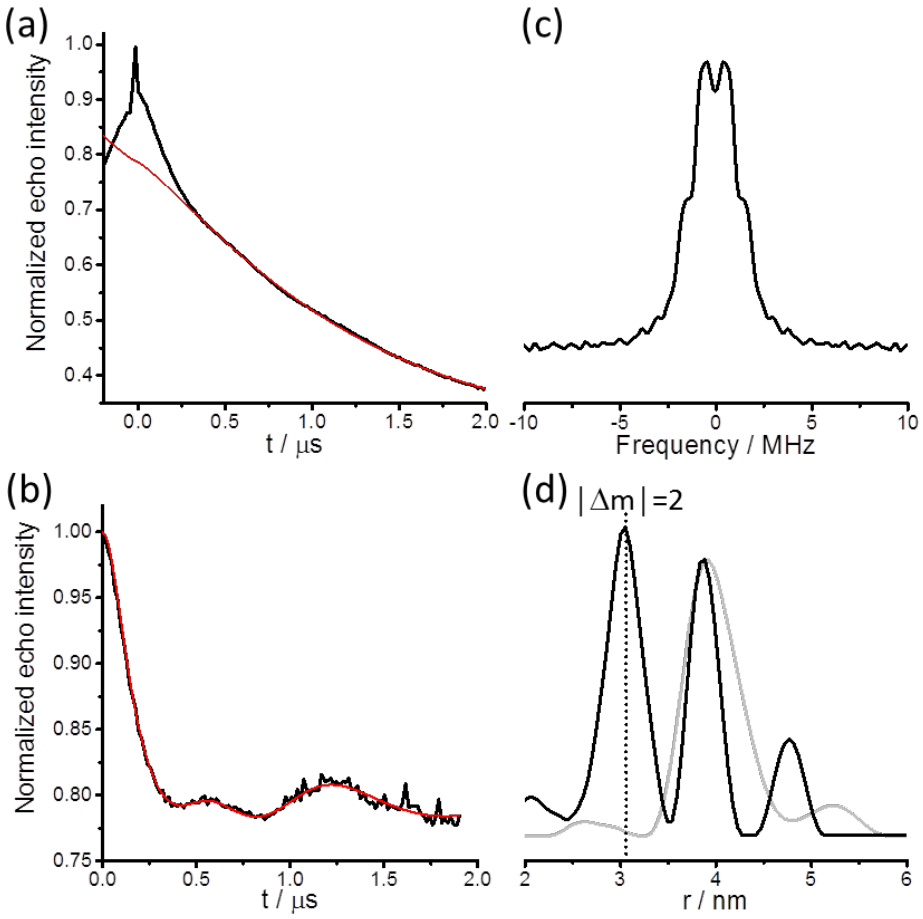


Figure S8. RIDME results of T4L-E-**T1**. (a) Primary data along with the background correction function used in red. (b) Background corrected RIDME trace (black) along with the fit obtained with the distance distribution shown in (d) (red). (c) Fourier transform of the experimental trace in (b). (d) Comparison of the distance distributions obtained from the analysis of RIDME (black) and DEER (grey) data. The dotted line shows the position expected for the second harmonic calculated from the main peak.

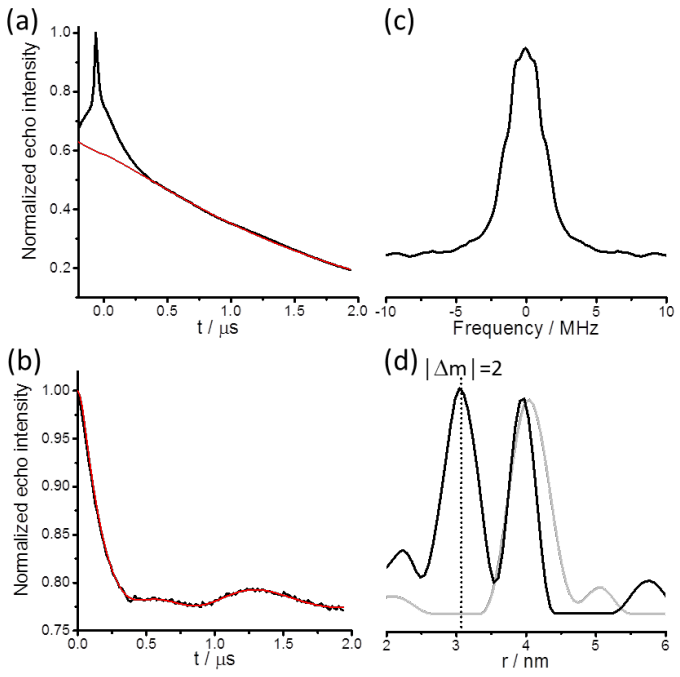


Figure S9. RIDME results of T4L-F-**T1**. (a) Primary data along with the background correction function used in red. (b) Background corrected RIDME trace (black) along with the fit obtained with the distance distribution shown on the right (red). (c) Fourier transform of the experimental trace in (b). (d) Comparison of the distance distributions obtained from the analysis of RIDME (black) and DEER (grey) data. The dotted line shows the position expected for the second harmonic calculated from the main peak.

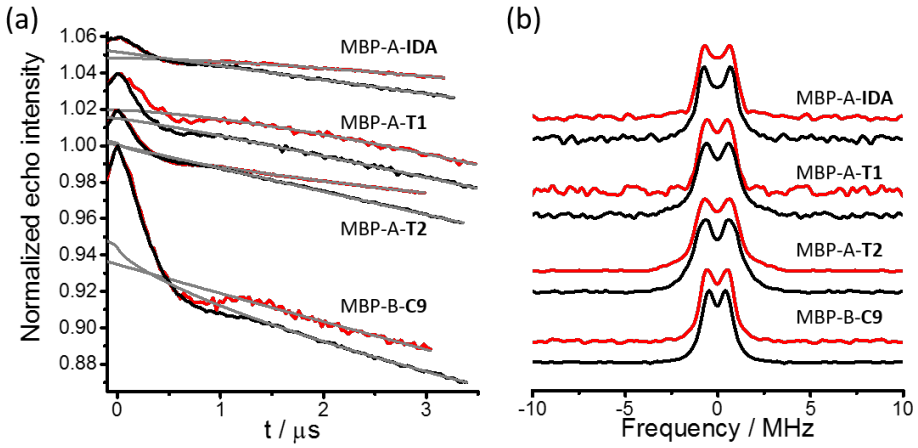


Figure S10. DEER data obtained with Gd^{3+} ions in different tags attached to MBP mutants in the presence (red) and absence (black) of one equivalent of maltose. (a) Primary DEER traces with background functions (grey). (b) Corresponding spectra after background subtraction and Fourier transformation.

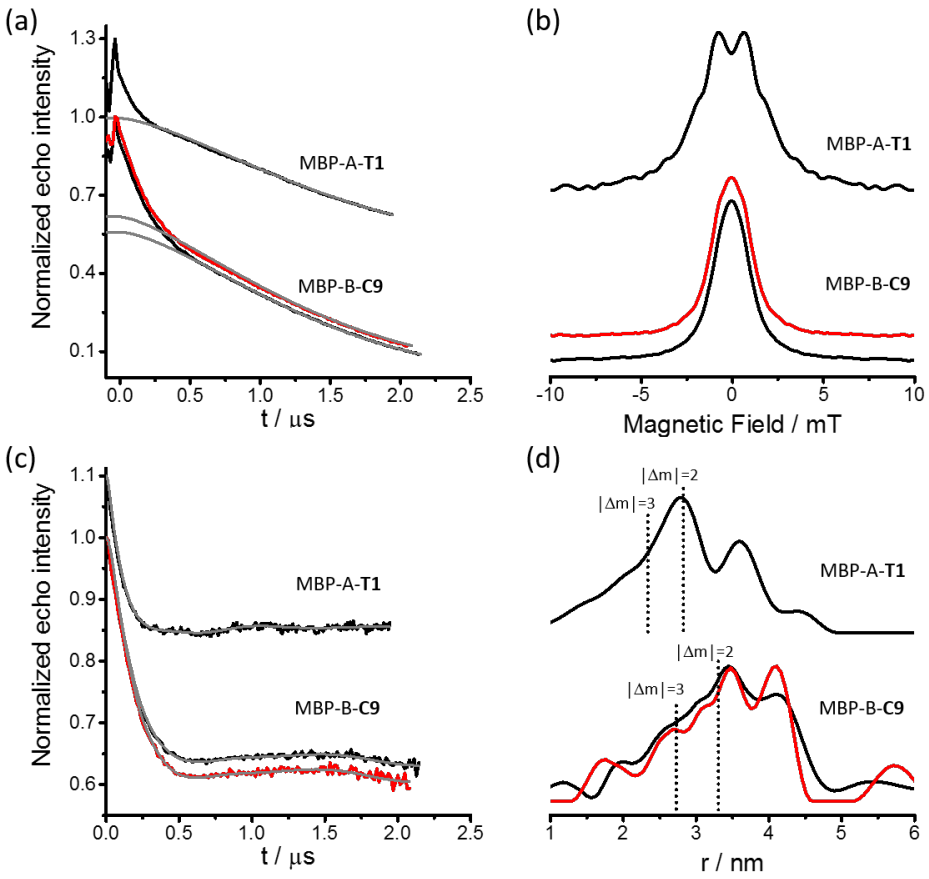


Figure S11. RIDME results obtained with MBP-A-**T1** and MBP-B-**C9** with (red) and without (black) one equivalent of maltose. (a) Primary RIDME traces. Grey lines indicate the background decay. (b) Corresponding spectra after background subtraction and Fourier transformation. (c) Background corrected RIDME traces. The grey lines correspond to the fits obtained with the distance distributions in (d). (d) Distance distributions obtained from data analysis. The dotted lines show the positions expected for the second and third harmonics calculated from the main peak.

CHAPTER 3

SHORT-ARM Gd^{3+} TAGS TO OBTAIN NARROW DEER DISTANCE DISTRIBUTIONS

This chapter discusses two short-arm Gd^{3+} SLs, DO3MA-3BrPy and 3BrPy-MTA, that deliver narrow distance distributions. DEER experiments with both tags have been published by Yang et al. (2017a, b). The following experiments were carried out in parallel to the published experiments to compare their performances with the C9 tag.

Contribution

I prepared all protein samples for EPR analysis, including protein synthesis, purification and ligation with the DO3MA-3BrPy, 3BrPy-MTA and C9 tags, and prepared the EPR samples. The tags were synthesized in the group of Prof. Xun-Cheng Su at the State Key Labroatory of Elemento-Orgranic Chemistry, Nankai University, Tianjin, China. Distance measurements and data analysis were performed by Drs Y. Yan and A. Feintuch in the group of Prof. D. Goldfarb at the Department of Chemical Physics, Weizmann Institute of Science, Israel. I interpreted the distance measurements, modelled the tagged protein and calculated theoretical distance distributions.

Introduction

Many Gd^{3+}-based SLs have been developed to measure nanometer-scale distances on proteins by the DEER experiment. Deriving accurate information on protein structure and conformational changes from DEER measurements, however, necessitates SLs with outstanding chemical and spectroscopic properties.

With regard to the chemical properties of the SLs, they should be chemically inert and stable under near-physiological conditions and in in-cell environments. In contrast to the conventional nitroxide SLs, using a Gd^{3+} ion as the paramagnetic centre provides full resistance towards reduction inside cells and hence much better performance in in-cell EPR experiments (Igarashi et al., 2010; Martorana et al., 2014). Many Gd^{3+}-based SLs, however, are attached to the protein via disulfide bonds, which are not stable at high pH or under reducing conditions. This has driven the development of Gd^{3+} SLs, which are attached via stable thioether bonds, specifically the maleimide-DO3A-Gd^{3+} tag (Martorana et al., 2014; Theillet et al., 2016) and the 4-vinyl-PyMTA tag loaded with Gd^{3+} (Qi et al., 2014; Yang et al., 2013). Alternatively, the DOTA-based C3 tag can be ligated onto a pAzF residue, which can be site-specifically incorporated into the target protein by genetic code expansion (Abdelkader et al., 2015a; Mahawaththa et al., 2018).

Reaction of 4-phenylsulfonated pyridine (4PS-Py) with cysteine thiols has recently been identified as an outstanding strategy for tagging proteins with a short and rigid C-S tether, which is resistant towards reduction (Yang et al., 2015; Yang et al., 2016). The phenylsulfonyl group at the 4-position of pyridine forms the leaving group for thioether bond formation between the cysteine residue and the pyridine ring. To provide sufficient affinity towards paramagnetic lanthanides, the pyridine moiety must functionalized with other, metal-chelating groups or metal complexes such as DOTA-derived frameworks. DO3MA-Py, where Py stands for 4PS-Py, and 4PS-PyMTA are two such tags. They were first introduced as SLs for paramagnetic NMR to produce large PCSs in proteins (Yang et al., 2015; Yang et al., 2016). In DO3MA-Py, the lanthanide ion is chelated by the DOTA framework, whereas the PyMTA moiety in 4PS-PyMTA forms a high-affinity heptadentate binding motif with paramagnetic metal ions in its own right.

The initial SLs derived from 4PS-Py suffered from low reactivity and quantitative tagging efficiency could be achieved only at high pH, using long incubation times or elevated temperatures during the ligation reactions. Since efficient tagging reactions are important for many biological applications of these tags, a bromide group was introduced

into the pyridine moiety to increase the reactivity of the DO3MA-Py tag. Indeed, the 3-bromo-4-(phenylsulfonyl)pyridine-DO3MA tag, DO3MA-3BrPy (Figure 1A), shows superior reactivity compared to the DO3MA-Py tag (Yang et al., 2017b). It would be expected that the 3-bromo-4-(phenylsulfonyl)pyridine-MTA tag, 3BrPy-MTA (Figure 1B), also displays increased reactivity compared with the parent 4PS-PyMTA tag. Indeed, the 4PS-PyMTA tag, which is a heptadentate metal chelator (Figure 1C), performs well in DEER experiments with proteins (Yang et al., 2017a) and the DO3MA-3BrPy tag has been shown to be stable in in-cell DEER measurements (Yang et al., 2017b).

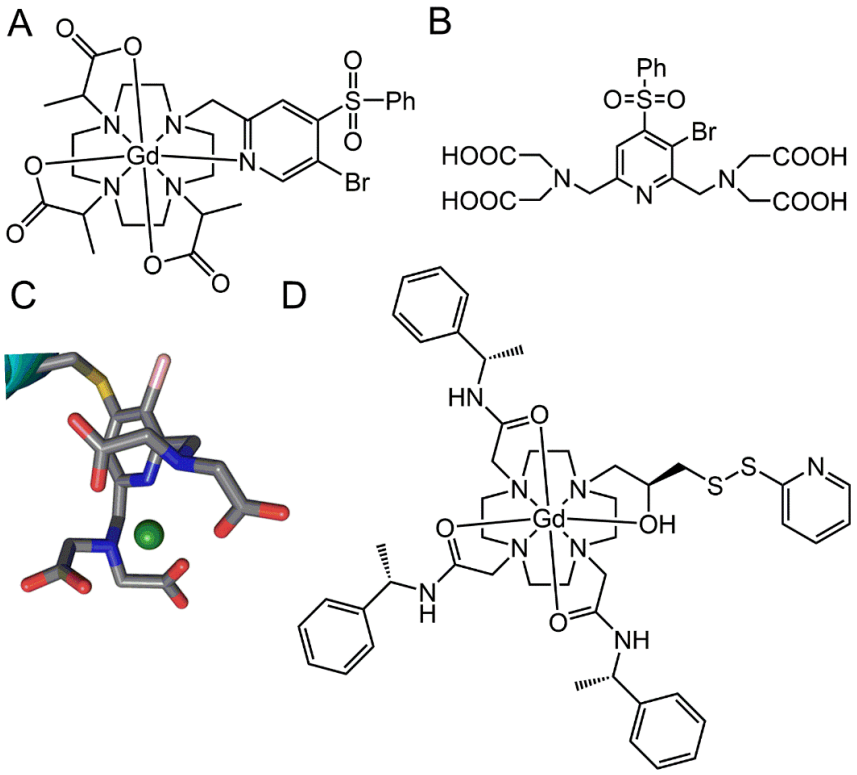


Figure 1. Chemical structures of tags used in this study. A) DO3MA-3BrPy-Gd^{3+} tag. B) 3BrPy-MTA tag. C) Paramagnetic metal-bound 3BrPy-MTA tag modelled onto a solvent-exposed cysteine residue in a protein. D) C9-Gd^{3+} tag.

Movement of the paramagnetic centre relative to the protein inevitably broadens the widths of distance distributions and has been identified as a major reason for limiting the sensitivity and information content of DEER measurements. This inspired the development of SLs with short and rigid tethers. 4PS-Py generates the shortest possible tether to a cysteine sulfur atom. In addition, the coordination of the pyridine nitrogen with the paramagnetic metal ion in the 3BrPy-DO3MA and 4PS-PyMTA tags affords excellent immobilization of the paramagnetic centre on the protein. These two tags are thus expected to deliver the narrowest DEER distributions achievable with single-arm cysteine tags. Although the double-arm tags T1 and T2 loaded with Gd^{3+} can, in principle, deliver

even narrower DEER distance distributions when attached to cysteine residues in positions *i* and *i*+8 of an α-helix with an aspartate residue in position *i*+4, single-arm tags have a significant advantage, if the structure of the protein is unknown or when introduction of several mutations in the protein is undesirable.

Although the DO3MA-3BrPy and 4PS-PyMTA tags have been identified as highly suitable tags for EPR experiments, their performance in DEER experiments has not been compared directly with previously published tags. One of the objectives of this study was thus to compare their DEER performance by comparison with the C9 tag (Figure 1D), which is one of the best performing single-arm Gd^{3+} tags (Abdelkader et al., 2015b). As described in Chapter 1, the ZFS of a Gd^{3+} SL is an important EPR spectroscopic property, which determines its utility. The magnitude of the ZFS parameter is directly proportional to the width of the central transition in the ED-EPR spectrum of the spin-labelled protein. Therefore, in this chapter, the C9, DO3MA-3BrPy and 3BrPy-MTA tags were also compared by their ZFS characteristics as reflected in the width of the central transition in their ED-EPR spectra and their performance was assessed in RIDME and DEER experiments.

Data were recorded for the DO3MA-3BrPy tag using two different single-cysteine mutants of the protein ERp29, which forms a homodimer. The data recorded with the DO3MA-3BrPy tag were compared with the DEER and RIDME data that had been published with the C9 tag attached to the same ERp29 mutants (Abdelkader et al., 2015b). The 3BrPy-MTA tag was used in the experiment instead of the previously reported 4PS-PyMTA tag because of its greater reactivity. Furthermore, the characteristics of the 3BrPy-MTA tag were evaluated on a double-cysteine mutant of T4 lysozyme (T4L) and were compared with the DEER and RIDME data of the C9 tag on the same T4L mutant (data from Chapter 2). Finally, to assess the tag-specific contributions to the widths of the distance distributions, the experimental distance distributions were compared with the theoretical distance distributions calculated from rotamer libraries of the tags modelled on the protein crystal structure.

Experimental section

Protein expression and purification

The single-cysteine mutants ERp29 S114C/C157S and ERp29 G147C/C157S of ERp29 and the double-cysteine mutant T4L C54T/C97A/D72C/V131C (T4L-D, see Chapter 2) were cloned into pETMCSI vectors (Neylon et al., 2000) with C-terminal His_6-tag and

N-terminal His_6-tag respectively. The protein samples were produced from PCR-amplified DNA by CFPS (Wu et al., 2007). CFPS reactions were carried out at 30 ^oC for 16 hours in dialysis mode according to a published protocol. The proteins were purified using 1 mL Co-NTA gravity columns (GE Healthcare, USA), following the manufacturer's protocol. The purified ERp29 mutants and T4L-D were dialysed against buffer A (50 mM phosphate buffer, pH 7.5, 1 mM dithiothreitol (DTT)) at 4 ^oC and concentrated using an Amicon centrifugal filter with a molecular weight cutoff (MWCO) of 10 kDa. The protein yields were 1 mg and 1.5 mg per mL of cell-free reaction mixture.

Synthesis of the tags

The DO3MA-3BrPy (Yang et al., 2017b) and 3BrPy-MTA tags were provided by Prof. Xun-Cheng Su.

Protein ligation with DO3MA-3BrPy and 3BrPy-MTA tags

To 0.1 mM solution of ERp 29 S114C/C157S and ERp29 G147C/C157S in buffer A, DTT was added to a final concentration of 5 mM to ensure that all cysteine residues were reduced. Excess DTT was removed and the buffer exchanged for buffer B (50 mM Tris-HCl, pH 8.5) using an Amicon centrifugation filter (MWCO 10 kDa). The protein solutions were subsequently added slowly into 15 equivalents of DO3MA-3BrPy tag in the same buffer. Reaction mixtures were shaken overnight at room temperature. The completion of the ligation reactions was confirmed by mass spectrometry. The same procedure was followed in buffer C (50 mM MES-KOH, pH 6.7) to ligate T4L-D with the 3BrPy-MTA tag.

DEER sample preparation

The tagged ERp29 mutants were concentrated and exchanged into EPR buffer (50 mM Tris-HCl in D_2O, pD 8.5, uncorrected pH meter reading) using Amicon centrifugal filters (MWCO 10 kDa). Similarly, the tagged T4L-D was concentrated and exchanged into a buffer containing 50 mM MES-KOH in D_2O, pD 6.5, uncorrected pH meter reading). Perdeuterated glycerol was added to reach 20% (v/v) final composition. $GdCl_3$ dissolved in D_2O was added to the samples tagged with 3BrPy-MTA tag in 2:1 molar ratio of $GdCl_3$ to protein.

DEER measurements

All EPR measurements were carried out in the group of Prof. D. Goldfarb at the Department of Chemical Physics, Weizmann Institute of Science (Israel), using a home-built W-band EPR spectrometer.

Modelling

The crystal structures of ERp29 (PDB ID: 2QC7; Barak et al., 2009) and T4 lysozyme (PDB ID: 2LZM; Weaver and Matthews, 1987) were used to model the distance distributions. The coordinates of the DO3MA-3BrPy, 3BrPy-MTA and C9 tags loaded with Gd^{3+} were crafted onto each of the cysteine residues in positions 72 and 131 of T4L and positions 114 and 147 of ERp29, and the distance distributions were modelled using PyParaTools as described previously (Abdelkader et al., 2015b), except that broader distance distributions were generated by allowing a greater range of dihedral angles for the C-S bonds ($\pm 10^{o}$ around the rotamer states at -60^{o}, $+60^{o}$, -90^{o}, $+90^{o}$) and defining steric clashes as interatomic distances less than 0.9 times the sum of the van der Waals radii.

Results

Protein samples and tagging strategies

Two mutants of ERp29 were used to assess the performance of the DO3MA-3BrPy tag. In both mutants, the wild-type cysteine residue at position 157 was mutated to serine. An additional mutation introduced a cysteine residue at a solvent-exposed position (position 114 or 147), resulting in the constructs ERp29 S114C/C157S and ERp29 G147C/C157S. For simplicity, the two mutants will be referred to in the following as ERp29 S114C and ERp29 G147C respectively. After overnight reaction at room temperature and pH 8.5, completion of the reaction between the DO3MA-3BrPy tag and ERp29 mutants was confirmed by mass spectrometry. ERp29 forms a 51 kDa homodimer, so that the DEER experiment of the tagged protein measures an inter-monomer distance.

The 3BrPy-MTA tag was tested with the T4L-D, where the natural cysteine residues at positions 54 and 97 were mutated to threonine and alanine, respectively. These mutations are known to maintain the structure of the protein (Matsumura and Matthews, 1989). After overnight reaction at room temperature and pH 6.5, completion of the reaction

between the 3BrPy-MTA tag and T4L-D was confirmed by mass spectrometry. $GdCl_3$ was added in 2:1 molar ratio of $GdCl_3$ to protein for subsequent EPR measurements.

Echo detected-EPR spectra and Gd^{3+}-Gd^{3+} distance measurements: DO3MA-3BrPy tag

W-band echo-detected EPR (ED-EPR) and DEER measurements were carried out with the two ERp29 mutants tagged with the DO3MA-3BrPy tag. The region of the Gd^{3+} central transition of the ED-EPR spectrum is shown in Figure 2A. The spectrum shows the characteristic single peak for Gd^{3+} and the width of the central transition is 8.3 mT.

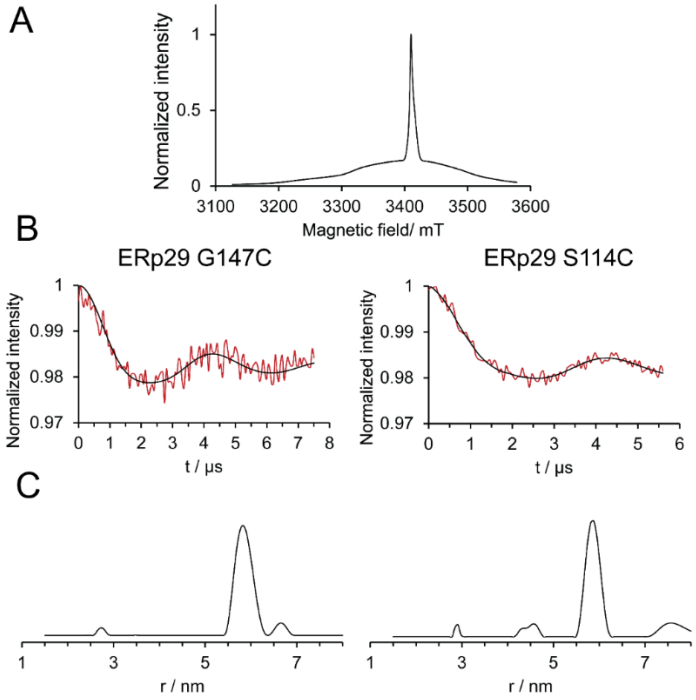


Figure 2. Echo-detected (ED) EPR and DEER results obtained with ERp29 mutants tagged with the DO3MA-3BrPy tag. A) ED-EPR spectrum of the central transition region of the Gd^{3+} spins. B) Background-corrected DEER traces (red) along with the fits obtained with the distance distributions shown in C (black). Left: ERp29 G147C mutant. Right: ERp29 S114C mutant. C) Analyzed distance distributions. Left: ERp29 G147C mutant. Right: ERp29 S114C mutant.

Figure 2B and C shows the DEER results obtained with the same two ERp29 samples. The numerical values for mean distances and widths of distance distributions are listed in Table 1 together with the DEER and RIDME data that have previously been published for the C9 tag.

Table 1. Experimental and predicted Gd^{3+}–Gd^{3+} distances and widths of distance distributions of the ERp29 mutants tagged with the 3BrPy-DO3MA tag

		Gd^{3+}–Gd^{3+} distance/nm		full width at half height/nm	
		ERp29 S114C	ERp29 G147C	ERp29 S114C	ERp29 G147C
DO3MA-3BrPy	experimental	5.8[D]	5.8[D]	0.4[D]	0.6[D]
	predicted	5.7	5.7	0.3	0.4
C9	experimental	5.8[D] 5.7[R]	5.8[D] 5.7[R]	0.4[D] 0.3[R]	1.2[D] 0.6[R]
	predicted	5.8	5.8	0.5	0.6

[D] from DEER experiments

[R] from RIDME experiments

Both mutants show an inter-spin DEER distance of 5.8 nm. Modelling of the DO3MA-3BrPy tag on the cysteine residues of the ERp29 samples predicts mean distances that are within 0.1 nm than the observed distances. The DO3MA-3BrPy tag attached to the ERp29 S114C mutant delivered a 0.2 nm narrower distance distribution than the same tag attached to the ERp29 G147C mutant. The C9 tag showed a much larger difference in distance distribution width between these two mutants (0.8 nm in the DEER experiment) and the DEER experiment yielded a particularly broad distance distribution for the G147C mutant.

As the ZFS of the C9 tag is small, the Gd^{3+} central transition line is significantly narrower in the C9 (2.3 mT; Abdelkader et al., 2015b) than in the DO3MA-3BrPy tag (8.3 mT). As a result, the DEER distance distributions obtained with the C9 tag are subject to artificial broadening arising from pseudo-secular terms of the dipolar interaction neglected in the data analysis. To overcome this problem, Collauto et al. (2016) performed RIDME experiments with the two ERp29 mutants tagged with the C9 tag. The DEER and RIDME results obtained with the C9 tag are summarized in Table 1. For both mutants, the RIDME experiments reported a 0.1 nm shorter mean distance than the DEER experiments and the widths of distance distributions was always narrower. As a drawback, additional peaks appeared at shorter distances (4.4 nm and 4.7 nm), which were assigned to harmonics of the dipolar modulation frequency, and in one instance such a peak was taller than the "real" distance distribution (Collauto et al., 2016).

To assess the tag-specific contribution to the widths of the distance distributions, the range of coordinates accessible to the Gd^{3+} ion of the DO3MA-3BrPy tag tied to the

cysteine side-chain was determined by calculating a library of permissible rotamers of the residue composed of cysteine plus tag. The distance distributions calculated in this way were 0.1 nm and 0.2 nm narrower than the experimental DEER values for the S114C and G147C mutant, respectively (Table 1). The much too large distribution width obtained in the DEER experiment with the G147C mutant labelled with the C9 tag remains unexplained. For a very similar Gd^{3+}–Gd^{3+} distance, the S114C mutant did not show this effect.

ED-EPR spectra and Gd^{3+}–Gd^{3+} distance measurements: 3BrPy-MTA tag

W-band ED-EPR and DEER measurements were carried out with the doubly 3BrPy-MTA-tagged T4L-D titrated with $GdCl_3$ in a protein:Gd^{3+} ratio of 1:2. The region of the Gd^{3+} central transition of the ED-EPR spectrum is shown in Figure 3A. The width of the central transition is very narrow, 4.8 mT.

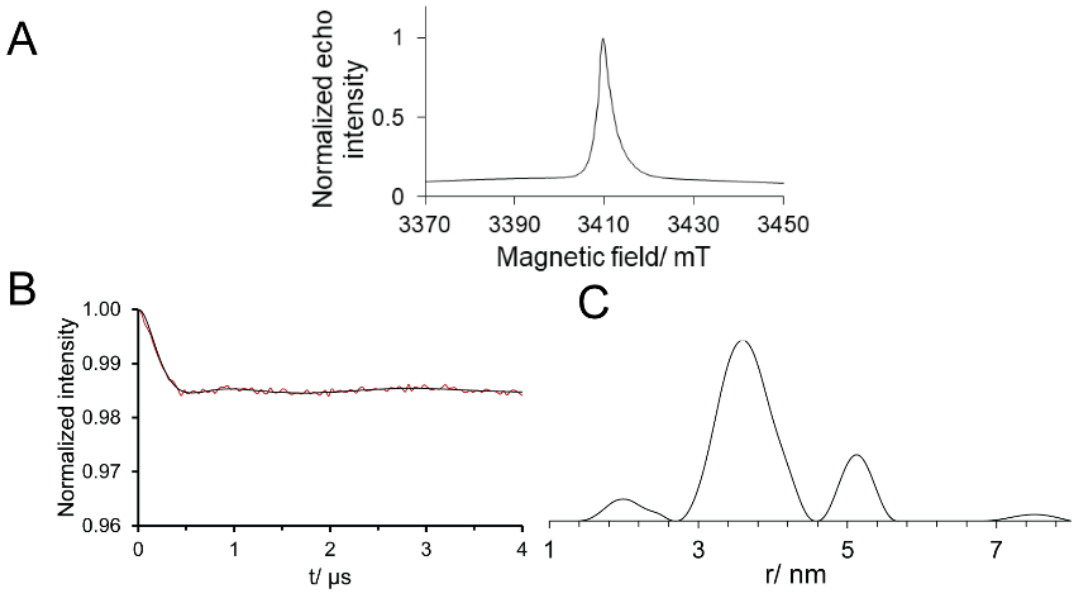


Figure 3. ED-EPR and DEER results obtained with T4L-D tagged with 3BrPy-MTA loaded with Gd^{3+}. A) ED-EPR spectrum of the central transition region. B) Background-corrected DEER trace (red) along with the fit obtained with the distance distribution shown in C (black). C) Analyzed distance distribution.

Figure 3B and C shows the DEER results obtained with the 3BrPy-MTA tag on the T4L-D. The numerical values for the mean distance and width of distance distribution are tabulated in Table 2 together with the data recorded with the C9-Gd^{3+} tag on the same T4L mutant.

Table 2. Experimental and predicted Gd^{3+}–Gd^{3+} distances and widths of the distance distributions of the T4L-D tagged with the 3BrPy-MTA tag

		Gd^{3+}–Gd^{3+} distance/nm		full width at half height/nm	
		DEER	RIDME	DEER	RIDME
3BrPy-MTA	experimental	3.6	3.4	0.8	0.4
	predicted	3.7		0.5	
C9	experimental	4.0	3.9	0.8	0.5
	predicted	4.3		0.5	

As predicted for a tag with a shorter tether, the 3BrPy-MTA tag delivered a DEER distance 0.4 nm shorter than the C9 tag. The experimental mean distance was only 0.1 nm shorter than the calculated mean distance for the 3BrPy-MTA tag, while there is a 0.3 nm discrepancy between the experimental and calculated values for the C9 tag. Despite the shorter tether in the 3BrPy-MTA tag, both tags delivered DEER distance distributions with the same widths.

Similar to the C9 tag (2.1 mT; see Chapter 2), the narrow central transition of the 3BrPy-MTA tag (4.8 mT) is narrow and therefore prone to artificial broadening arising from pseudo-secular terms of the dipolar interaction neglected in the data analysis. To test this hypothesis, RIDME data were recorded with the T4L-D tagged with the 3BrPy-MTA tag (Figure 4A and B) and compared with RIDME data of the T4L-D tagged with the C9 tag (data from Chapter 2). Indeed, the RIDME experiments showed narrower distance distributions than the DEER experiments. The mean distances observed with the RIDME experiments were slightly shorter than with the DEER experiments.

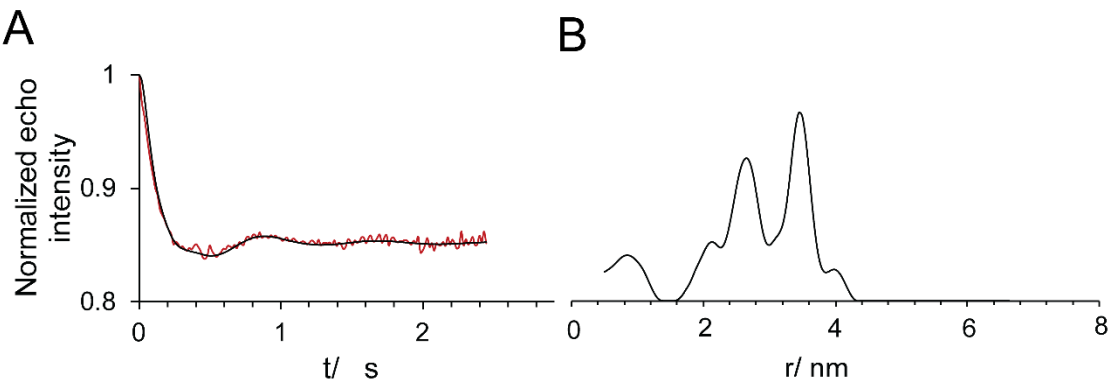


Figure 4. RIDME results obtained with the T4L-D tagged with the 3BrPy-MTA tag. A) Background-corrected RIDME trace (red) along with the fit obtained with the distance distribution shown in C. B) Distance distribution obtained from the analysis of the RIDME data.

Tag-specific contributions to the widths of distance distributions were determined by modelling the tags onto the two cysteine residues and generating rotamer libraries. The range of coordinates accessible to the Gd^{3+} ion of each tag was then used to derive theoretical distance distributions. For both tags, the theoretical widths of distance distributions were 0.3 nm narrower than measured by the DEER experiments, while the corresponding widths obtained by RIDME were within 0.1 nm of the predictions.

Discussion

The DO3MA-3BrPy tag performed well in delivering complete ligation reaction yields. The tag coordinates the Gd^{3+} ion in a highly unsymmetrical geometry, which may explain the large ZFS associated with the tag and the broad width of the central transition in the EPR spectrum. DEER data obtained with this Gd^{3+}-tag can thus be fitted using the weak coupling approximation without resulting in artificial broadening of the distance distributions. This is illustrated by the excellent fits between the experimental and predicted widths of the distance distributions, as well as the mean distances (see Table 1).

Both the C9 tag and the C7 tag introduced by Prokopiou et al. (2018) delivered significantly broader DEER distance distributions with the ERp29 G147C mutant. Interestingly, the DO3MA-3BrPy tag delivered the narrowest distance distribution ever recorded with the ERp29 G147C mutant. Red spheres in Figure 5A mark the range of coordinates accessible to the Gd^{3+} ion of the DO3MA-3BrPy tag attached to the two

cysteine mutants of ERp29. The spread in coordinates shows that residue 147 of ERp29 is more solvent-exposed than residue 114, which is indeed manifested in the different widths of DEER distance distributions observed with the two ERp29 mutants.

One of the objectives of this work was to compare the performances of the DO3MA-3BrPy and C9 tags. In spite of the shorter tether in the DO3MA-3BrPy tag, the mean distance measured by the DO3MA-3BrPy tag was very similar to the distance reported for the C9 tag. In addition, the ERp29 S114C mutant gave similar DEER distance distributions and distribution widths with both tags.

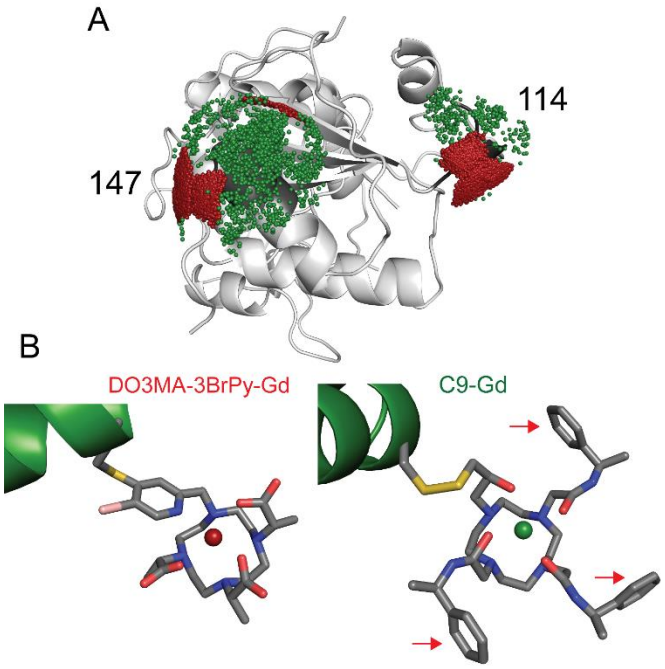


Figure 5. Modelling of the DO3MA-3BrPy and C9 tag onto ERp29 using the crystal structure 2QC7 (Barak et al., 2009). A) Comparison of the range of coordinates accessible to the Gd^{3+}ion in the DO3MA-3BrPy and C9 tag on the mutants S114C and G147C of ERp29. The metal positions predicted by the modelling are indicated by red balls for the DO3MA-3BrPy tag and green balls for the C9 tag. The helices carrying the tags are highlighted in dark grey. B) The DO3MA-3BrPy (left) and C9 tags (right) modelled onto a solvent-exposed cysteine residue of an α-helix showing the difference in tether length and overall size of the two tags. The arrows identify the phenyl rings of the C9 tag.

Consistent with the theoretical distance distribution calculations, the ERp29 G147C mutant delivered a broader distance distribution with the C9 than the DO3MA-

3BrPy tag, and the distance distribution widths were larger than for the S114C mutant. This reflects the difference in solvent exposure between the two tagging sites (Figure 5A). The fact that the DEER distance distribution width observed for the ERp29 G147C mutant with C9 tag appeared 0.6 nm broader than with the DO3MA-3BrPy tag and wider than modelled is explained by the small ZFS of the C9 tag, resulting in artificial broadening that was eliminated by RIDME experiments of the same sample (Collauto et al., 2016). This interpretation, however, is contradicted by the DEER distance distribution measured with the ERp29 S114C mutant, where the C9 tag did not yield any significant broadening of the distance distribution and DEER and RIDME experiments delivered similar results, despite a very similar Gd^{3+}–Gd^{3+} distance as in the G147C mutant (5.8 nm). At present, there is no good explanation for this observation, especially as artificial broadening of DEER distance distributions is not expected for long distances. In light of the inconsistencies observed with the C9 tag, the present work suggests that the DO3MA-3BrPy tag is a superior tag to measure nanometer scale distances by DEER experiments.

On a practical level, the bromine atom in the 3BrPy-MTA tag seems to solve the problems with incomplete ligation yields observed with the earlier 4PS-MTA tag, which were very sensitive to the local environment of cysteine residue the protein. In fact, attempts to attach the 4PS-PyMTA tag to the ERp29 S114C mutant repeatedly failed. This problem was solved by the modification of the pyridine with a bromide. Remarkably, this increased the reactivity of the tag sufficiently to allow complete ligation of the T4L-D also at a lower pH value.

The ED-EPR spectrum of T4L-D tagged with 3BrPy-MTA showed a narrow central EPR line for the Gd^{3+} ion. Therefore, artificial broadening of the DEER distance distributions was expected due to the weak coupling approximation during data analysis. Indeed, the significant discrepancy in width (0.3 nm) between the theoretical and experimental DEER distance distributions confirmed the artificial broadening situation. As expected, RIDME experiments delivered a narrower distance distribution in better agreement with the theoretical prediction. Another important observation was that the new 3BrPy-MTA tag delivers shorter mean distances than the C9 tag because of its shorter tether and the mean distance reported with the 3BrPy-MTA tag more closely matches the predictions too (within 0.1 nm). Nonetheless, the distance distributions produced by the single-arm 3BrPy-MTA tag are still somewhat broader than those produced by the double-arm T2 and T1 tags attached to the same sites of T4L.

The widths of the DEER distance distributions observed delivered by the DO3MA-3BrPy and 3BrPy-MTA tag matched those of the RIDME distance distributions

delivered by the C9 tag, although the shorter tethers of the DO3MA-3BrPy and 3BrPy-MTA tags should produce narrower distance distribution widths. Notably, however, the DO3MA-3BrPy and 3BrPy-MTA tags lack the bulky phenylethylamide pendant groups of the C9 tag (Figure 5B). This raises the question, whether the bulky pendants of the C9 tag assist in confining the space accessible to the Gd^{3+} ion. Recent studies by Mahawaththa et al. (2018) and Prokopiou et al. (2018) compared the DEER distance distribution widths of the C9 tag with those of the C7, C8 and C11 tags, which also lack the phenylethylamide pendant groups of the C9 tag. These authors concluded that the hydrophobic pendants of the C9 tag do not significantly contribute to narrowing the distance distribution widths by virtue of their size, but rather due to undesired hydrophobic interactions with neighbouring residues in the protein, which the potential to both perturb the protein structure and result in unpredictable distance distributions. This conclusion was drawn from DEER distance measurements. RIDME distance measurements could yield a more accurate comparison because, in contrast to the C9 tag, the C7, C8 and C11 tags are characterized by a large ZFS and hence are not subject to artificial broadening during data analysis.

Conclusion

The DO3MA-3BrPy tag is an excellent tag for measuring DEER distance distributions, which can be very accurately predicted irrespective of the tagging site based on rotamer libraries of the tag modelled onto a crystal structure of the protein. For the C9 tag, RIDME experiments are required to obtain faithful distance distributions that match the DEER distance distributions obtained with the DO3MA-3BrPy tag. In contrast to the DO3MA-3BrPy tag, the 3BrPy-MTA tag displays a narrow central transition in the ED-EPR spectrum, requiring RIDME experiments for accurate distance distribution measurements.

References

Abdelkader, E. H., Feintuch, A., Yao, X., Adams, L. A., Aurelio, L., Graham, B., Goldfarb, D. and Otting, G. (2015a) Protein conformation by EPR spectroscopy using gadolinium tags clicked to genetically encoded *p*-azido-L-phenylalanine. *Chem. Commun*. **51**, 15898–15901.

Abdelkader, E. H., Lee, M. D., Feintuch, A., Ramirez-Cohen, M., Swarbrick, J. D., Otting, G., Graham, B. and Goldfarb, D. (2015b) A new Gd^{3+} spin label for Gd^{3+}-Gd^{3+}

distance measurements in proteins produces narrow distance distributions. *J. Phys. Chem. Lett.* **6**, 5016–5021.

Barak, N. N., Neumann, P., Sevvana, M., Schutkowski, M., Naumann, K., Malesevic, M., Reichardt, H., Fischer, G., Stubbs, M. T. and Ferrari, D. M. (2009) Crystal structure and functional analysis of the protein disulfide isomerase-related protein ERp29. *J. Mol. Biol*. **385**, 1630–1642.

Ching, H. Y. V., Demay-Drouhard, P., Bertrand, H. C., Policar, C., Tabares, L. C. and Un, S. (2015) Nanometric distance measurements between Mn(II) DOTA centers. *Phys. Chem. Chem. Phys*. **17**, 23368–23377.

Collauto, A., Frydman, V., Lee, M. D., Abdelkader, E. H., Feintuch, A., Swarbrick, J. D., Graham, B., Otting, G. and Goldfarb, D. (2016) RIDME distance measurements using Gd(iii) tags with a narrow central transition. *Phys. Chem. Chem. Phys.* **18**, 19037–19049.

Dalaloyan, A., Qi, M., Ruthstein, S., Vega, S., Godt, A., Feintuch, A. and Goldfarb, D. (2015) Gd(iii)-Gd(iii) EPR distance measurements-the range of accessible distances and the impact of zero field splitting. *Phys. Chem. Chem. Phys*. **17**, 18464–18476.

Igarashi, R., Sakai, T., Hara, H., Tenno, T., Tanaka, T., Tochio, H. and Shirakawa, M. (2010) Distance determination in proteins inside *Xenopus laevis* oocytes by double electron−electron resonance experiments. *J. Am. Chem. Soc.* **132**, 8228–8229.

Mahawaththa, M. C., Lee, M. D., Giannoulis, A., Adams, L. A., Feintuch, A., Swarbrick, J. D., Graham, B., Nitsche, C., Goldfarb, D. and Otting, G. (2018) Small neutral Gd(III) tags for distance measurements in proteins by double electron–electron resonance experiments. *Phys. Chem. Chem. Phys*., submitted.

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.

Martorana, A., Bellapadrona, G., Feintuch, A., Di Gregorio, E., Aime, S. and Goldfarb, D. (2014) Probing protein conformation in Cells by EPR distance measurements using Gd^{3+} spin labeling. *J. Am. Chem. Soc.* **136**, 13458–13465.

Matsumura, M. and Matthews, B. W. (1989) Control of enzyme activity by an engineered disulfide bond. *Science* **243**, 792–794.

Prokopiou, G., Lee, M. D., Collauto, A., Abdelkader, E. H., Bahrenberg, T., Feintuch, A., Ramirez-Cohen, M., Clayton, J., Swarbrick, J. D., Graham, B., Otting, G. and Goldfarb, D. (2018) Small Gd(III) tags for Gd(III)-Gd(III) distance measurements in proteins by EPR spectroscopy. *Inorg. Chem.* **57**, 5048–5059.

Qi, M., Gross, A., Jeschke, G., Godt, A. and Drescher, M. (2014) Gd(III)-PyMTA label is suitable for in-cell EPR. *J. Am. Chem. Soc.* **136**, 15366–15378.

Theillet, F. X., Binolfi, A., Bekei, B., Martorana, A., Rose, H. M., Stuiver, M., Verzini, S., Lorenz, D., van Rossum, M., Goldfarb, D. and Selenko, P. (2016) Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature* **530**, 45–50.

Weaver, L. H. and Matthews, B. W. (1987) Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol*. **193**, 189–199.

Wu, P. S. C., Ozawa, K., Lim, S. P., Vasudevan, S., Dixon, N. E. and Otting, G. (2007) Cell-free transcription/translation from PCR amplified DNA for high-throughput NMR studies. *Angew. Chem. Int. Ed.* **46**, 3356–3358.

Yang, Y., Li, Q. F., Cao, C., Huang, F. and Su, X. –C. (2013) Site-specific labeling of proteins with a chemically stable, high-affinity tag for protein study. *Chem. Eur. J*. **19**, 1097–1103.

Yang, F., Wang, X., Pan, B. –B. and Su, X. –C. (2016) Single-armed phenylsulphonated pyridine derivative of DOTA is rigid and stable paramagnetic tag in protein analysis. *Chem. Commun.* **52**, 11535–11538.

Yang, Y., Gong, Y. –J., Litvinov, A., Liu, H. –K., Yang, F., Su, X. –C. and Goldfarb, D. (2017a) Generic tags for Mn(II) and Gd(III) spin labels for distance measurements in proteins. *Phys. Chem. Chem. Phys*. **19**, 26944–26956.

Yang, Y., Wang, J. –T., Pei, Y. –Y. and Su, X. –C. (2015) Site-specific tagging proteins *via* a rigid, stable and short thiolether tether for paramagnetic spectroscopic analysis. *Chem. Commun.* **51**, 2824–2827.

Yang, Y., Yang, F., Gong, Y. –J., Chen, J. –L., Goldfarb, D. and Su, X. –C. (2017b) A reactive, rigid Gd-III labeling tag for in-cell EPR distance measurements in proteins. *Angew. Chem. Int. Ed*. **56**, 2914–2918.

CHAPTER 4

SITE-SPECIFIC INCORPORATION OF SELENOCYSTEINE BY GENETIC ENCODING AS A PHOTOCAGED UNNATURAL AMINO ACID

This chapter is reproduced from the following published article:
Welegedara, A. P., Adams, L. A., Huber, T., Graham, B. and Otting, G. (2018) Site-specific incorporation of selenocysteine by genetic encoding as a photocaged unnatural amino acid. *Bioconjugate Chem.* **19**, 2257–2264.

Contribution:

The photocaged unnatural amino acid was synthesized by Dr Luke A. Adams in the group of A/Prof. Bim Graham at the Monash Institute of Pharmaceutical Sciences, Monash University, Australia. I expressed all the protein samples used in the study, prepared the protein samples for mass spectrometry and NMR analysis, and carried out all the NMR experiments. I also analyzed the mass spectrometry and NMR data and made a significant contribution to writing not only the first but also subsequent versions of the manuscript.

Bioconjugate Chemistry

Article

Cite This: *Bioconjugate Chem.* 2018, 29, 2257−2264

pubs.acs.org/bc

Site-Specific Incorporation of Selenocysteine by Genetic Encoding as a Photocaged Unnatural Amino Acid

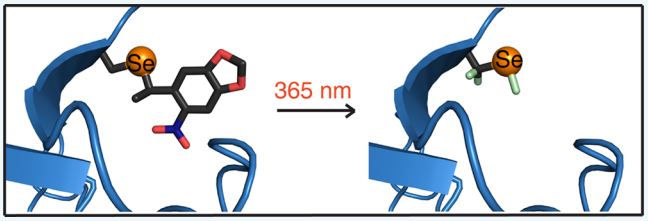
Adarshi P. Welegedara,[†] Luke A. Adams,[‡] Thomas Huber,[†] Bim Graham,[‡] and Gottfried Otting*[,†]

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

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

Supporting Information

ABSTRACT: Selenocysteine (Sec) is a naturally occurring amino acid that is also referred to as the 21st 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 pK_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 comprising 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

Received: April 10, 2018
Revised: May 16, 2018
Published: June 6, 2018

DOI: 10.1021/acs.bioconjchem.8b00254
Bioconjugate Chem. 2018, 29, 2257−2264

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 is site-selectively 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

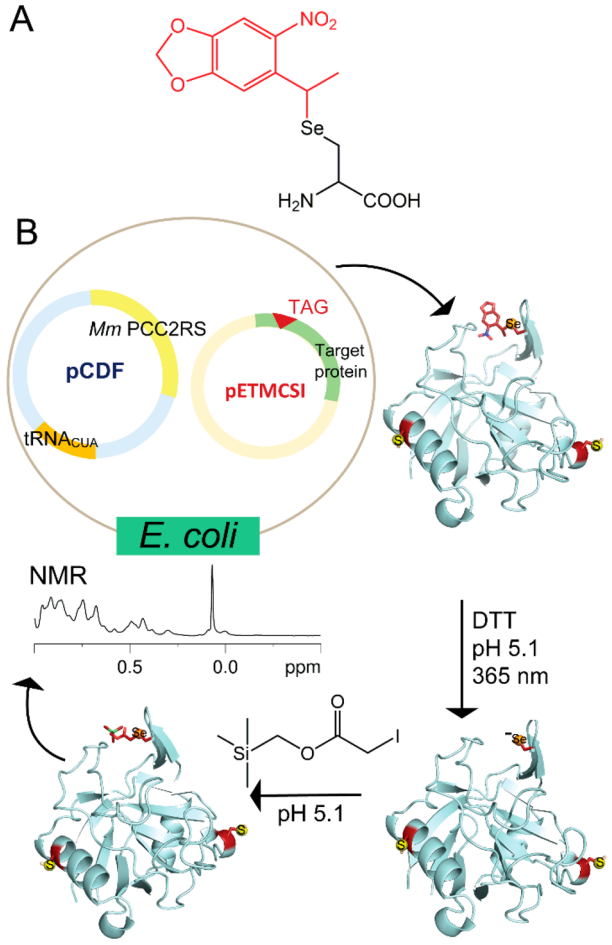


Figure 1. Genetically encoded incorporation of PSc into proteins in *E. coli*. (A) Chemical structure of photocaged selenocysteine (PSc). The red moiety is the caging group that is released upon UV irradiation. (B) Strategy for the production and site-specific alkylation of selenoproteins followed in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

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 under our reaction conditions.

DOI: 10.1021/acs.bioconjchem.8b00254
Bioconjugate Chem. 2018, 29, 2257−2264

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).

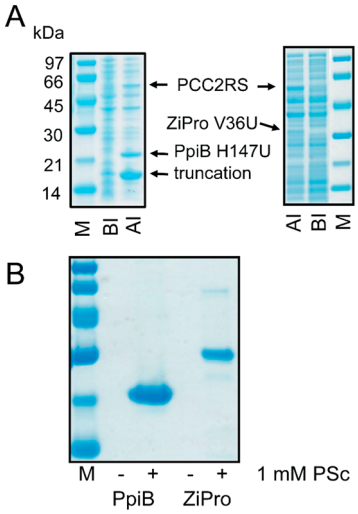


Figure 2. SDS-PAGE analysis of the expression and purification of PpiB H147U and ZiPro V36U, where U stands for PSc. (A) Expression of PpiB H147U (left panel) and ZiPro V36U (right panel). M: marker lane; BI and AI: before and after induction with IPTG, showing the total cellular protein of *E. coli*. 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 Co-NTA column.

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 Co-NTA column (Figure 2B).

High-resolution mass spectrometry of full-length proteins confirmed the incorporation of PSc (Figure 3A and C) 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.

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 3B,D). Importantly, dehydroalanine or serine byproducts 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 min 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 toward 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 disfavored. The pH is very important, however, as 5 min 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.

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.

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.

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

DOI: 10.1021/acs.bioconjchem.8b00254
Bioconjugate Chem. 2018, 29, 2257–2264

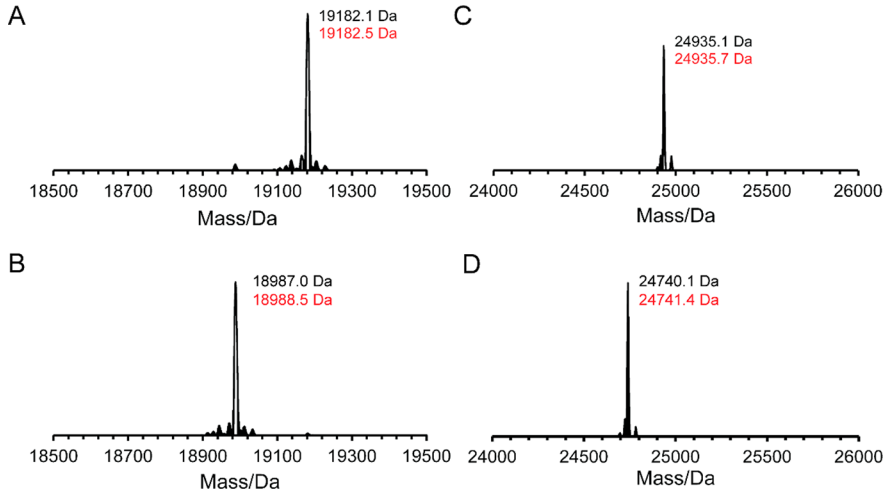


Figure 3. Mass spectra of intact proteins containing PSc and, after decaging, Sec. Observed and expected masses are indicated in black and red, respectively. The expected masses were calculated based on the masses observed for wild-type PpiB and ZiPro (Figure S3). (A) PpiB H147PSc. (B) PpiB H147Sec.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

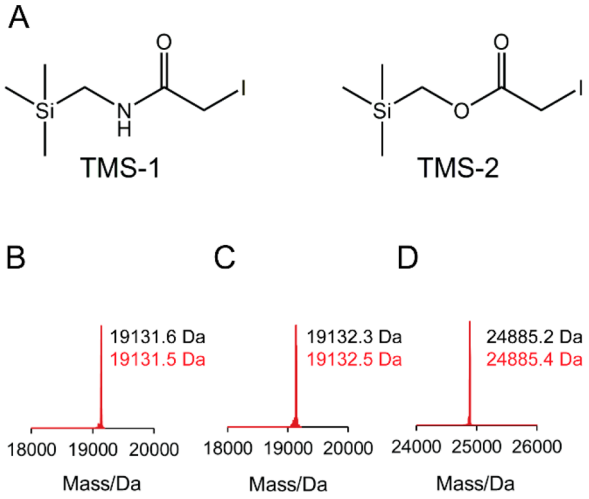


Figure 4. Selective Sec labeling with alkylating TMS tags. Observed and expected masses are indicated in black and red, respectively. The expected mass increases with TMS-1 and TMS-2 are 143 and 144 Da, respectively. Compare with Figure 3B and D for the masses of the untagged selenoproteins. (A) Structures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

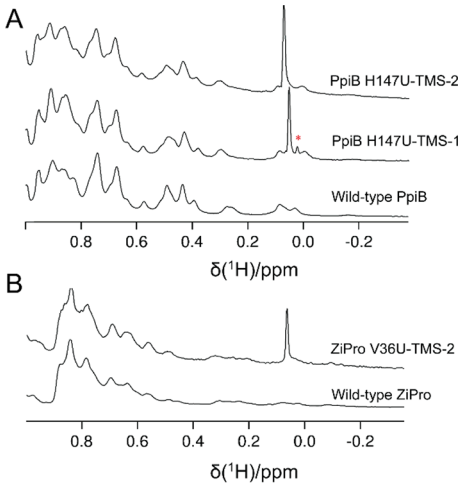


Figure 5. 1D ^{1}H NMR spectra of PpiB and ZiPro labeled with TMS tags. All spectra were measured at 25 °C. (A) ^{1}H NMR spectrum of wild-type PpiB and PpiB H147U tagged with TMS-1 or TMS-2. Protein concentrations ranged between 10 and 160 μM in 50 mM HEPES, pH 8.0. A red star marks the position of 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 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).

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

DOI: 10.1021/acs.bioconjchem.8b00254
Bioconjugate Chem. 2018, 29, 2257−2264

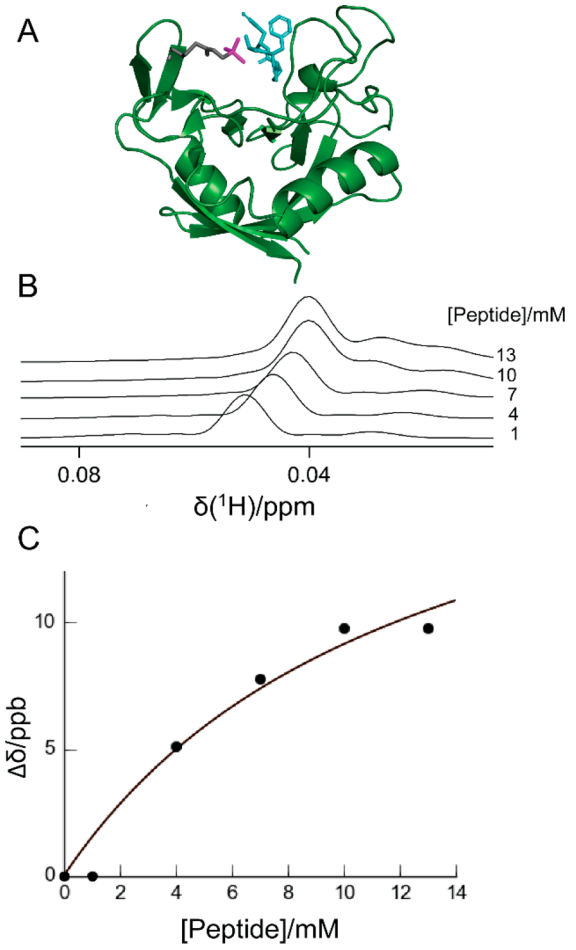


Figure 6. Peptide binding study with PpiB H147U-His_6 tagged with TMS-1. (A) Substrate binding site of PpiB with a Sec-TMS-1 residue modeled at position 147. The protein is displayed in a ribbon representation (PDB ID: 1LOP),[33] showing the suc-AAPF-pNA peptide in cyan and the TMS group in magenta. (B) 1D ^{1}H NMR spectra of the TMS signal of PpiB H147U-His_6 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 eq 1.

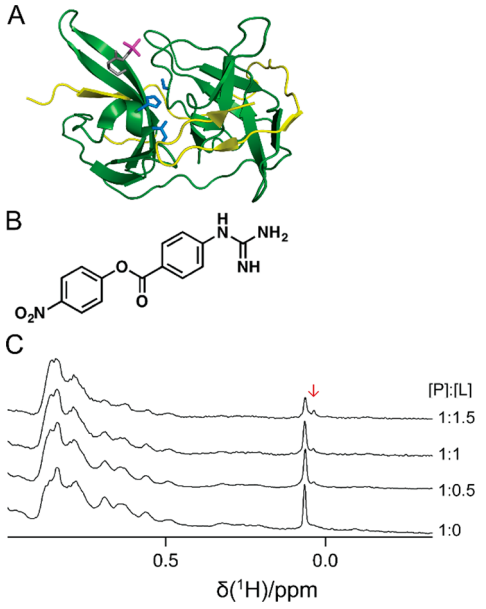


Figure 7. Inhibitor binding studies with ZiPro V36U-His_6 tagged with TMS-2. (A) Ribbon representation of ZiPro (PDB ID: 5LC0),[36] showing the NS3 protease domain in green and the NS2B cofactor in yellow. The TMS group of the TMS-2 tag attached to the Sec residue at position 36 is shown in magenta and 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.

residues[14−17] and invites many other applications,[11] including the use of Sec for expressed protein ligation.[17−19]

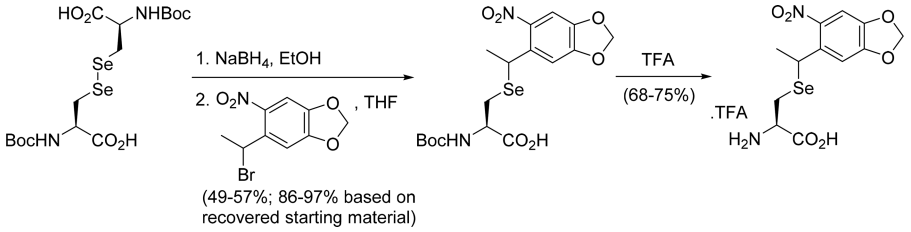
EXPERIMENTAL SECTION

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 toward autocleavage. 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 (17 300 *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

DOI: 10.1021/acs.bioconjchem.8b00254
Bioconjugate Chem. 2018, 29, 2257−2264

Scheme 1. Synthesis of Photocaged Selenocysteine (PSc)



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 glovebox for 10 min at room temperature, sealed, and shaken for another 20 min. 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 5-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 glovebox for 10 min at room temperature, sealing and shaking for another 20 min, 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 h. Excess DTT was removed by passing the samples through PD10 columns (GE Healthcare, USA) before adding a 5-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 °C, 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 analyzer 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 eq 1 to the data points of Figure 6C. In eq 1,

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

$\Delta\delta$ is the observed change in chemical shift, $\Delta\delta_{max}$ 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 $\Delta\delta$ replaced by the signal intensity of the complex and $\Delta\delta_{max}$ replaced by the total signal intensity of free and bound protein.[32]

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 precoated silica gel aluminum-backed plates (Merck Kieselgel 60 F_{254}). Visualization was by examination under UV light at 254 nm, with subsequent staining by $KMnO_4$. Dry-flash column chromatography[44] was run using Davisil P60 silica gel (40−63 μm). 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 ($CDCl_3$, δ_H = 7.26 and δ_C = 77.16; DMSO, δ_H = 2.50 and δ_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 ionization (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 H_2O; buffer B: 0.1% formic acid in MeCN. The following gradient was used with a Poroshell 120 EC-C18 50 × 3.0 mm 2.7 μm 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

DOI: 10.1021/acs.bioconjchem.8b00254
Bioconjugate Chem. 2018, 29, 2257−2264

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 (∼3 mL) 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.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.bioconjchem.8b00254.

Figure displaying locations of cysteine residues, LC-MS/MS analysis of ZiPro V36PSc, mass spectra of wild-type PpiB and ZiPro, analysis of the reactions of cysteine mutants of PpiB and ZiPro with TMS-1, ^{1}H NMR spectra of PSc and the synthetic intermediates (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gottfried.otting@anu.edu.au. Phone: +61 2 61256507. Fax: +61 2 61250750.

ORCID

Gottfried Otting: 0000-0002-0563-0146

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Mr Mithun Chamikara Mahawaththage for analyses by mass spectrometry, Dr Heinz Neumann for the original plasmid pCDF Duet_4-components_lac(P1), Ms Zuyan Wu for substitution of the *Mb* PylRS for *Mm* PylRS, 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.

■ 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, Gen. Subj. 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. 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.*, 1.
(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*, 7215.
(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 15*, 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 Spectrom. 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 70*, 211.
(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., Ser. 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.

DOI: 10.1021/acs.bioconjchem.8b00254
Bioconjugate Chem. 2018, 29, 2257−2264

Supporting information

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

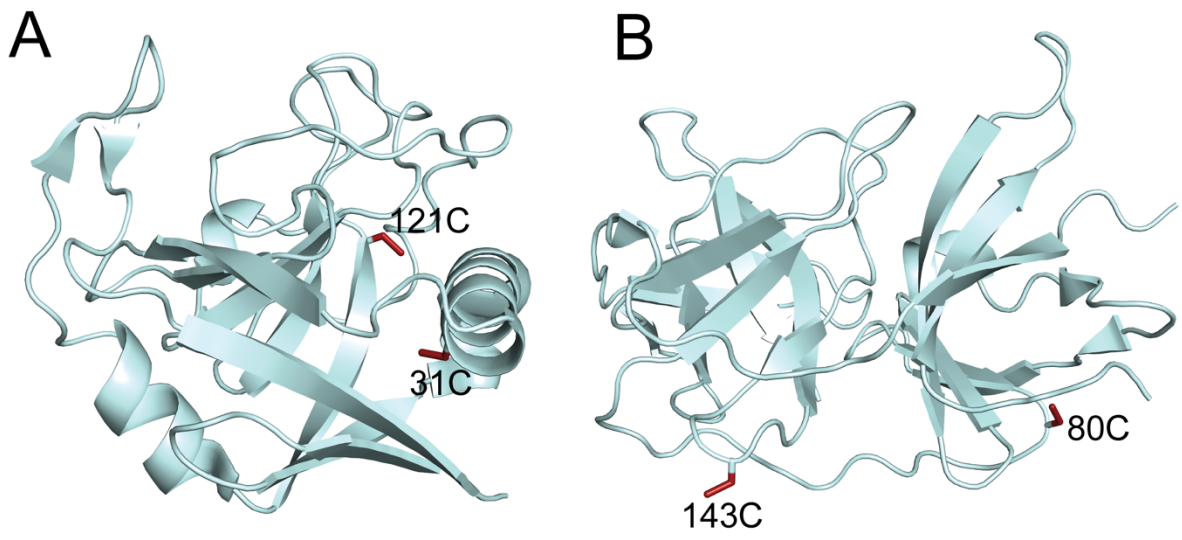


Figure S1. Cysteine residues in PpiB are buried inside the protein, while cysteine residues in ZiPro are fully solvent exposed. (A) Ribbon representation of the crystal structure of PpiB (PDB ID 2NUL[1]), highlighting the side chains of cysteine residues in red. (B) Same as (A), but for ZiPro (PDB ID 5LC0[2]).

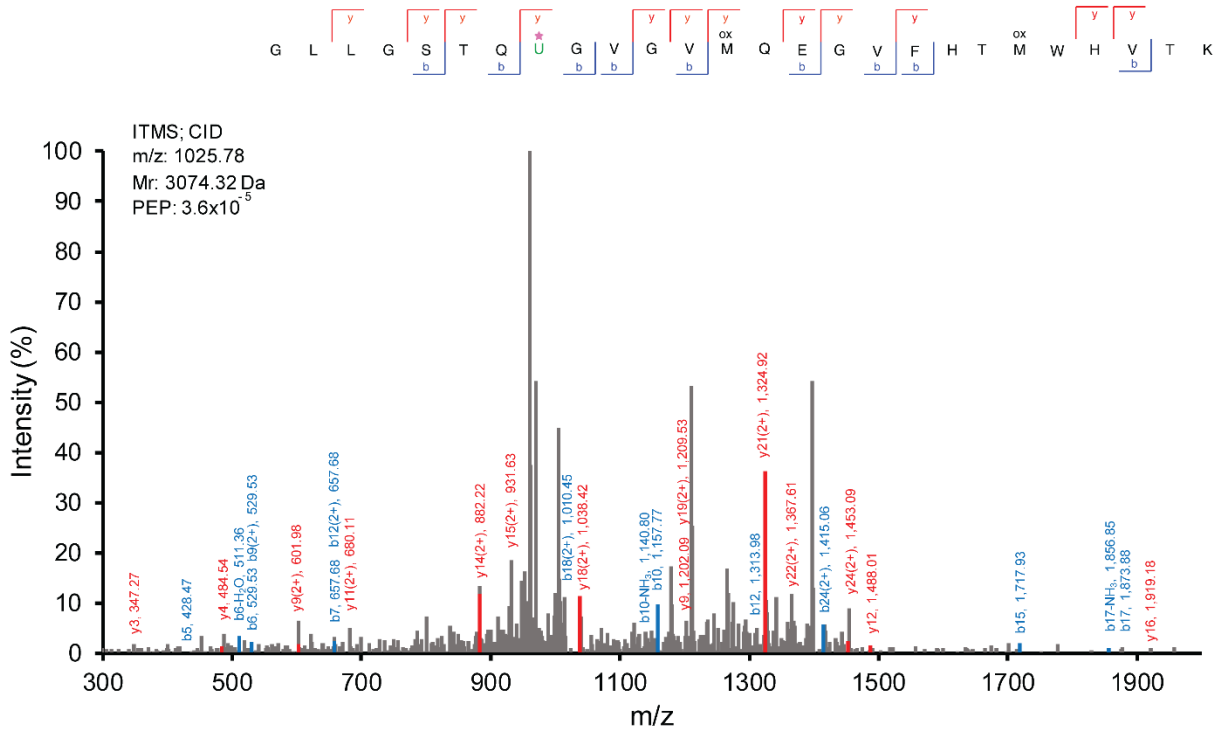


Figure S2. LC-MS/MS analysis of ZiPro V36PSc showing PSc incorporation at the amber codon. The triple-charged (+3) ion with m/z of 1025.78 was fragmented to produce the MS/MS spectrum. Annotated peaks in the MS/MS spectrum correspond to the series of b ions (blue) and y ions (red) that resulted from fragmentation of the peptide from ZiPro V36PSc shown at the top. b- and y-ion fragmentation is also indicated in the amino acid sequence, where the position of the PSc residue is marked by a star.

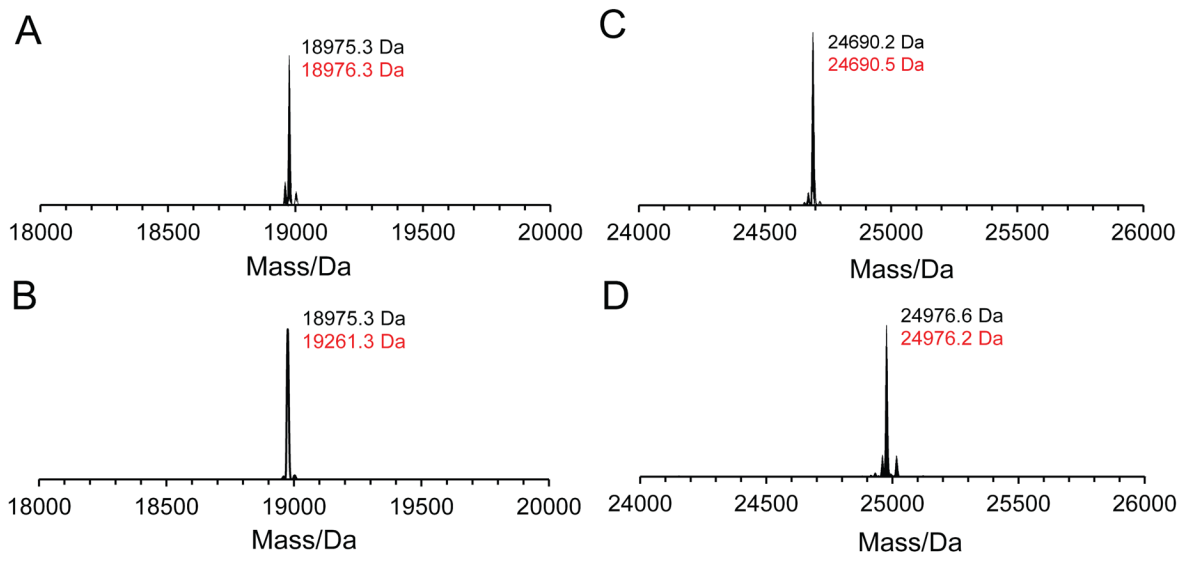


Figure S3. Mass spectra of wild-type PpiB and ZiPro before and after overnight incubation with TMS-1 at pH 7.5 and room temperature. Observed and expected masses are indicated in black and red, respectively. (A) Wild-type PpiB. (B) Wild-type PpiB after incubation with TMS-1. The unchanged mass highlights the inaccessibility of the cysteine residues in the interior of the three-dimensional structure of PpiB. (C) Wild-type ZiPro. The observed mass corresponds to the protein without N-terminal methionine. (D) Wild-type ZiPro tagged with TMS-1. The protein contains two solvent-exposed cysteine residues. Ligation with two TMS-1 tags leads to an expected increase in mass of 286 Daltons.

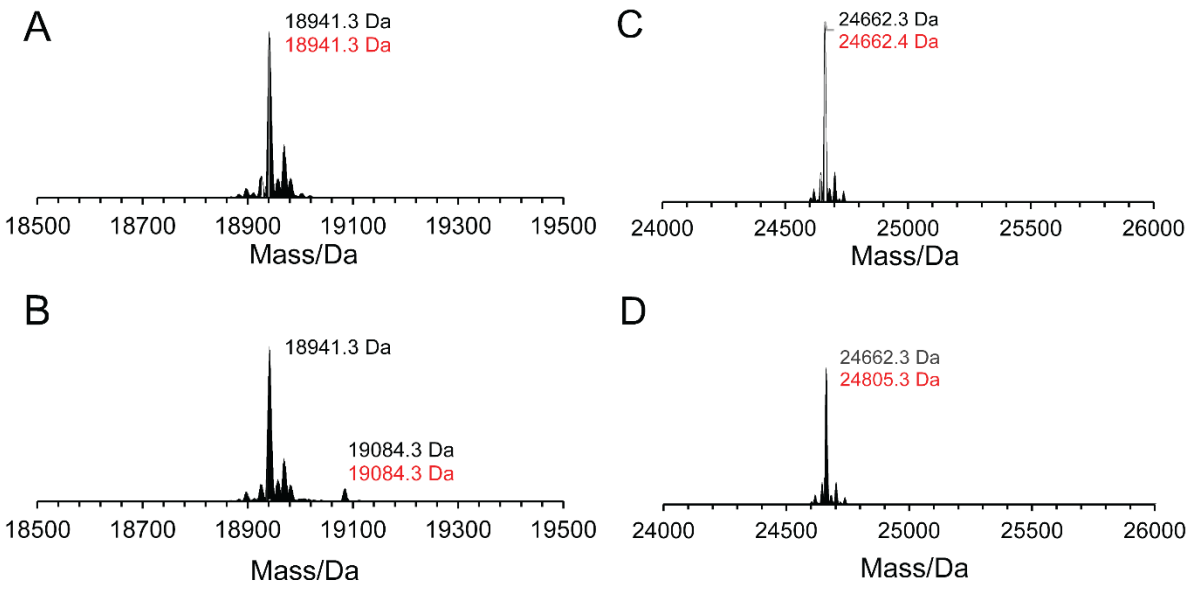


Figure S4. Mass spectra illustrating the poor reactivity of solvent-exposed cysteine residues of PpiB H147C and ZiPro V36C/C80S/C143S at pH 5.1 and room temperature towards overnight reaction with TMS-1. Observed and expected masses are indicated in black and red, respectively. (A) PpiB H147C mutant. (B) PpiB H147C after reaction with TMS-1. The small peak at 19084.3 Da indicates a tagging yield of less than 10%. (C) ZiPro V36C/C80S/C143S mutant. (D) Mass spectrum of ZiPro V36C/C80S/C143S after reaction with TMS-1. No evidence for reaction product was observed.

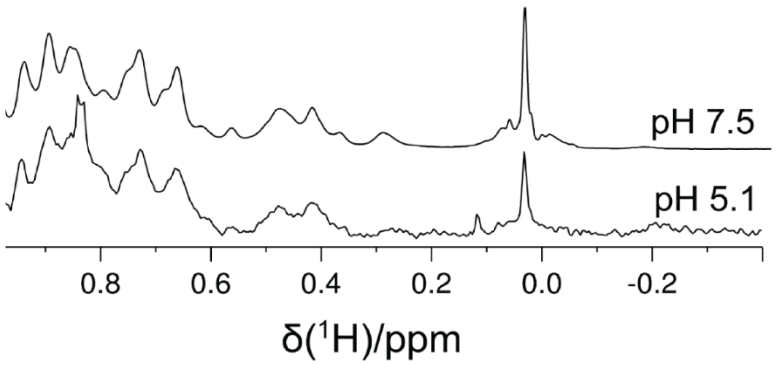
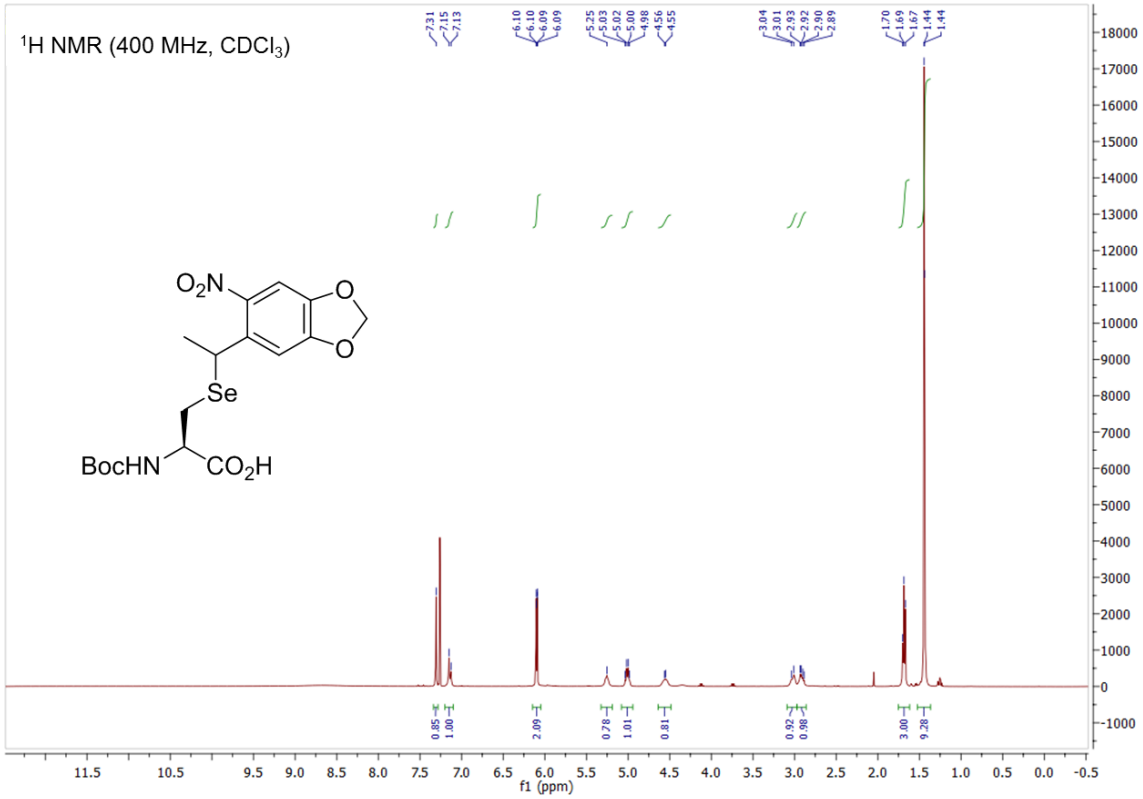
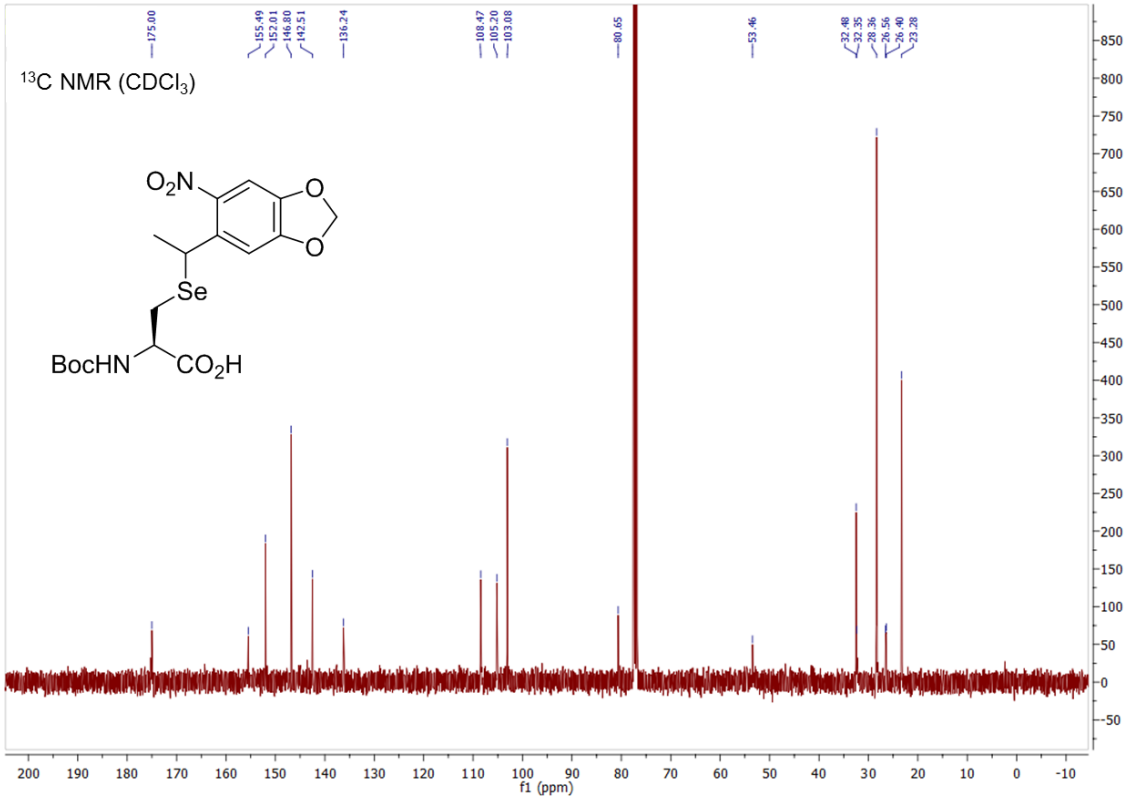
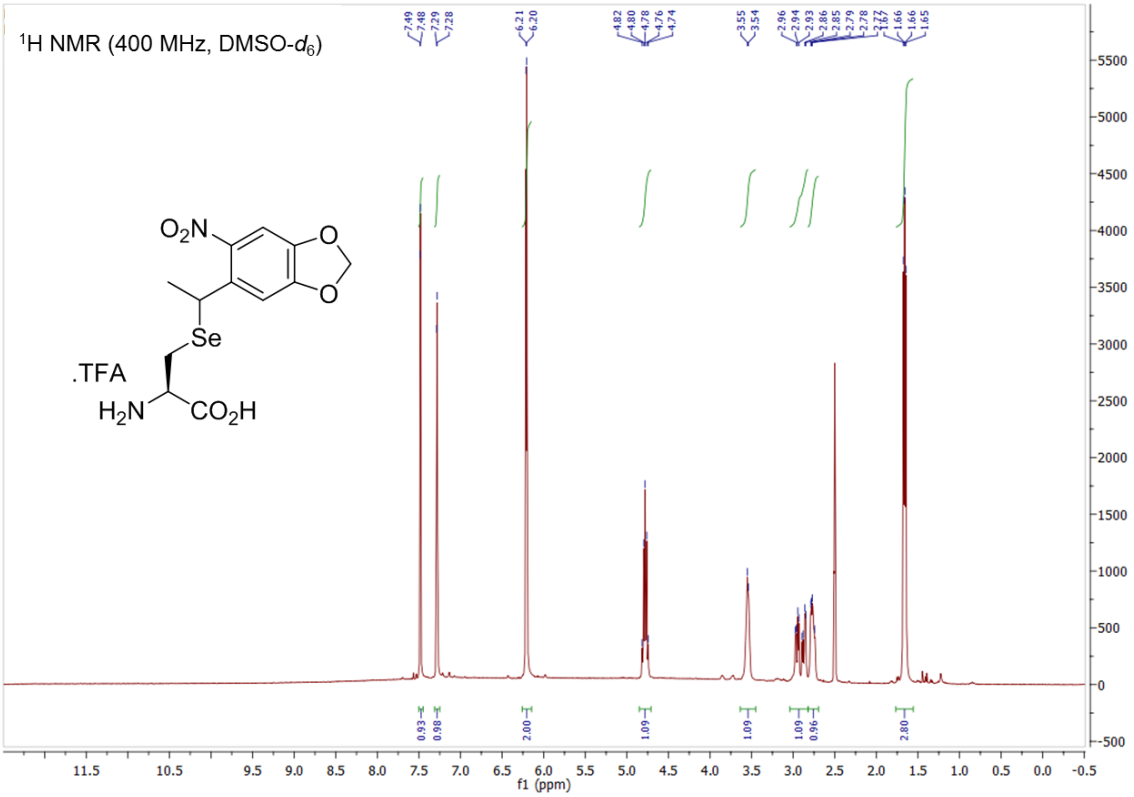


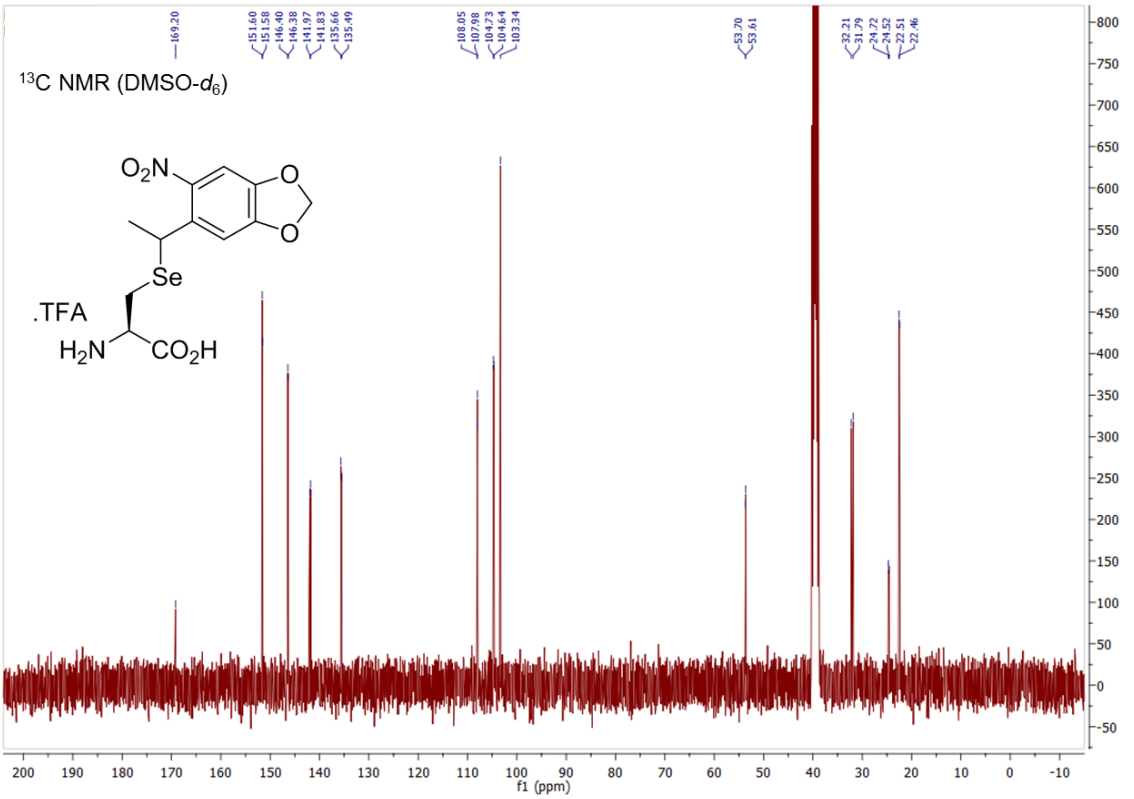
Figure S5. 1D ^{1}H NMR spectra comparing the reactivity of TMS-1 at pH 7.5 and 5.1. PpiB H147 was incubated with TMS-1 overnight at room temperature. The signal of the TMS moiety (at about 0.03 ppm) was less intense when the reaction was performed at pH 5.1 rather than 7.5. The NMR spectra were recorded of 10 and 50 µM protein solutions in 50 mM HEPES, pH 8.0.

NMR spectra of the synthetic intermediate and PSc






^{1}H NMR (400 MHz, DMSO-d_6)
O_2N
O
O
Se
.TFA
H_2N
CO_2H
f1 (ppm)


^{13}C NMR (DMSO-d_6)
O_2N
O
O
Se
.TFA
H_2N
CO_2H
f1 (ppm)

References

(1) 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.

(2) 2) 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.

CHAPTER 5

INCORPORATION OF ISOTOPE-LABELLED PHOTOCAGED SELENOCYSTEINE FOR SITE-SELECTIVE ISOTOPE LABELLING OF ALANINE

Contribution

Ms Natalie Murnane expressed and purified PpiB H147PSc and performed the deselenization reaction according to procedure 1 under my supervision. I performed all the other experiments in this Chapter.

Introduction

Synthesis of isotope-labelled proteins enables structure determinations of proteins and the study of ligand binding to proteins using heteronuclear NMR techniques. Isotope labelling is attractive as it causes minimal perturbations to the structure and activity of a protein. It is most common to use *in vivo* systems for the production of uniformly isotope-labelled protein. To reduce the complexity of 2D NMR spectra, selective isotope labelling can be achieved by providing individual amino acids in isotope-labelled form. *In vivo* protein expression with isotope-labelled amino acids, however, is expensive and metabolic conversions make it difficult to avoid the appearance of isotopes in other amino acids ("isotope scrambling"), which is a problem particularly for ^{15}N-labelled amino acids (LeMaster, 1990; McIntosh and Dahlquist, 1990; Shortle, 1994; Takeuchi et al., 2007). Isotope scrambling can be partially eliminated by using auxotrophic bacterial strains for protein expression *in vivo* (LeMaster, 1990; McIntosh and Dahlquist, 1990; Waugh, 1996; Whittaker, 2007). Selective isotope-labelling is most economically achieved by CFPS, where metabolic enzymes are far less active than *in vivo* (Kainosho and Güntert, 2010; Kigawa et al., 1995; Ozawa et al., 2004; Ozawa et al., 2006; Shi et al., 2004; Sobhanifar et al., 2010; Su et al., 2011; Yabuki et al., 1998). While residue-type selective isotope-labelling can readily be achieved in this way (Jia et al., 2009), isotope-labelling of a single individual amino-acid residue is much more difficult. In principle, suppressor-tRNA can be chemically loaded with an isotope-labelled amino acid for incorporation in response to a stop codon (Ellman et al., 1992; Yabuki et al., 1998). In practice, the large quantities required for NMR make this approach not only laborious but also very expensive.

An alternative method uses a genetically encoded unnatural amino acid that can be converted into one of the canonical 20 amino acids after incorporation into the target protein. Site-specific isotope-labelling of proteins at a specific tyrosine residue was achieved in this way by Cellitti et al. (2008). There, the isotope-labelled amino acid was synthesized with a photo-sensitive protection group for site-specific incorporation as a UAA and the labelled natural amino-acid residue was released by photolysis. In principal, this strategy can be applied to achieve site-specific incorporation of any of the 20 canonical amino acids in isotope-labelled form.

Orthogonal aaRS/tRNA pairs for the incorporation of natural amino acids have been reported for photocaged tyrosine, serine, cysteine and lysine (Deiters et al., 2006; Gautier et al., 2010; Lemke et al., 2007; Nguyen et al., 2014; Wu et al., 2004).

The incorporation of alanine as a photocaged UAA has not been reported. Notably, however, the PSc UAA introduced in Chapter 4 could indirectly enable the site-selective incorporation of isotope-labelled alanine. This is based on the fact that PSc can efficiently be converted into Sec, which can be converted into alanine by a radical deselenization mechanism using tris(2-carboxyethyl)-phosphine (TCEP) at slightly acidic pH and under anaerobic conditions (Figure 1; Derry et al., 2015; Metanis et al., 2010). In fact, conditions for converting Sec into alanine have been optimized as this is one of the key steps in Sec-based expressed protein ligation (EPL) and native chemical ligation (NCL) at alanine residues (Derry et al., 2015; Metanis et al., 2010).

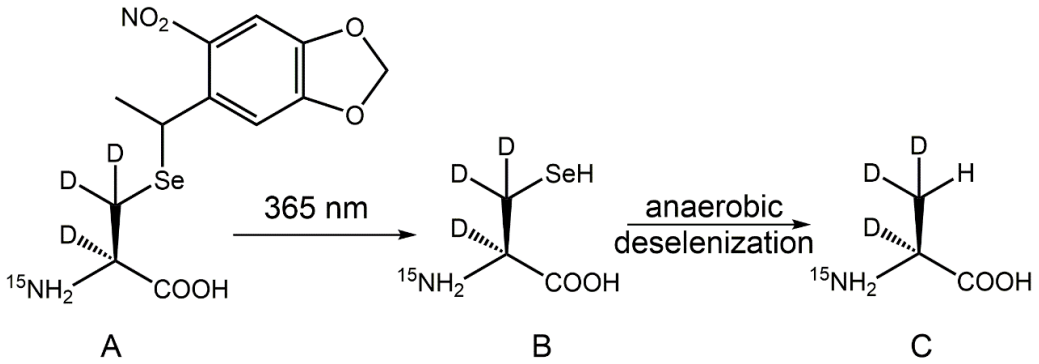


Figure 9. Reaction scheme for the conversion of isotope-labelled PSc to isotope-labelled alanine via photolysis and anaerobic deselenization. Chemical structures of A) isotope-labelled PSc, B) isotope-labelled Sec, C) isotope-labelled alanine.

The objective of the present study was to incorporate isotope-labelled PSc (Figure 1A) in response to an amber stop codon. A model protein, peptidyl-prolyl *cis-trans* isomerase (PpiB H147U) was used to produce a sample where an alanine residue at position 147 is site-specifically ^{15}N-labelled (PpiB H147(^{15}N)A), using anaerobic deselenization following photolysis. In this study, optimizations of the reaction conditions were carried out with unlabelled PSc to save costs.

Experimental procedure

Expression and purification of PpiB H147PSc

PpiB H147PSc was expressed and purified as described in Chapter 4.

Photolysis and deselenization

Procedure 1:

0.1 mM PpiB H147PSc in buffer A (50 mM phosphate buffer pH 7.5, 150 mM NaCl, 1 mM DTT) was illuminated with UV light as described in Chapter 4. Oxygen in buffer B (100 mM acetate buffer, pH 5.1) was removed by bubbling nitrogen gas through the solution for 16 h. The following steps were performed inside a low-oxygen glove box using buffer B (also degassed by bubbling nitrogen through the solution) which will in the following be referred to as buffer B'. Buffer A in the decaged protein was exchanged to buffer B', using PD-10 columns equilibrated with buffer B'. The eluate was concentrated to a protein concentration of 0.1 mM, 200 equivalents TCEP were added and the mixture was kept at 37 °C for 1 min.

Procedure 2:

0.1 mM PpiB H147PSc in buffer A was illuminated with UV light as described in the Chapter 4. Oxygen in buffer C (50 mM phosphate buffer, pH 5.1) was removed by bubbling nitrogen gas through the buffer as described in Procedure 1. The following steps also followed Procedure 1, using buffer C instead of buffer B, where buffer C' stands for degassed buffer C. Buffer A in the decaged protein was exchanged to buffer C' using a centrifugal filter unit (MWCO 10 kDa; Amicon Ultra, Millipore, Billerica, USA). At a concentration of 0.1 mM in buffer C', 200 equivalents TCEP were added and the protein was kept at 37 °C for 1 min.

Procedure 3:

Procedure 3 was similar to Procedure 2, except that 0.1 mM PpiB H147Sec in buffer C' was treated with 15 mM Na_2SO_3 and incubated for 15 minutes at room temperature before heating to 37 °C for one minute in the presence of 200 equivalents TCEP.

Procedure 4:

0.1 mM PpiB H147PSc in buffer D (50 mM phosphate buffer pH 5.1, 1 mM DTT) was illuminated with UV light as described in Chapter 4. The protein was treated with 50 equivalents of TCEP, 5 equivalents of DTT and 10 equivalents of glutathione. Nitrogen gas was bubbled for 15 minutes through the reaction mixture and finally, the reaction mixture was incubated at 37 ^{o}C for 5 minutes.

Mass spectrometry

Mass spectrometry analysis 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.

Results

Deselenization of Sec to alanine needs to be carried out at acidic pH and under rigorously anaerobic conditions. In the initial attempts to convert Sec into alanine (Procedure 1), we selected 100 mM acetate buffer, pH 5.1, as the deselenization buffer, removing dissolved oxygen by bubbling nitrogen through the solution. Unfortunately, as shown in Figure 2A, these reaction conditions resulted in serine as the major product, presumably due to the presence of traces of oxygen in the reaction buffers. In addition, deselenization was not 100% complete, in contrast to the observations reported by Derry et al. (2015) for selenopeptides.

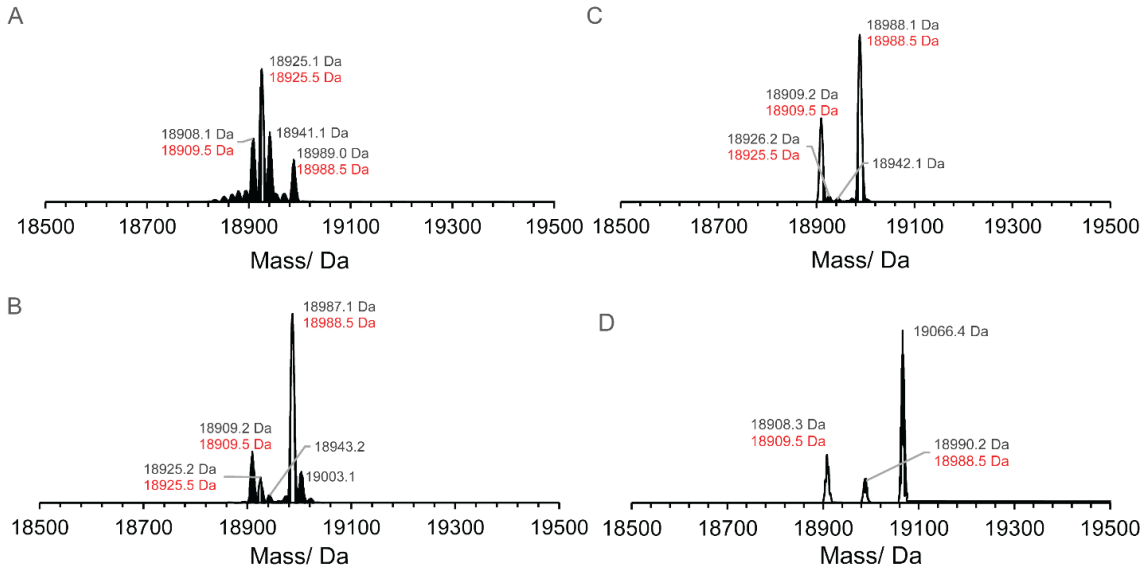


Figure 2. Mass spectra illustrating the efficiencies of different deselenization procedures. A) Procedure 1. B) Procedure 2. C) Procedure 3. D) Procedure 4. Theoretical masses are indicated in red. Theoretical masses of the possible products of PpiB H147Sec deselenization are as follows. PpiB H147Sec:18988.5 Da, PpiB H147Ser: 18925.5 Da, PpiB H147Ala: 18909.5 Da.

Assuming that buffer exchange with a PD-10 column introduces oxygen, procedure 2 employed a centrifugal filter unit in the buffer exchange steps. 50 mM phosphate buffer at pH 5.1 was selected as the deselenization buffer and care was taken to keep oxygen away from the protein sample. Indeed, more alanine than serine was produced under these reaction conditions (Figure 2B) and the relative amount of alanine was further increased with the help of Na_2SO_3 as oxygen scavenger (Procedure 3, Figure 2C). These results show that the presence of oxygen, even in trace amounts, yields the undesired product serine. The conversion to serine could be suppressed completely by careful removal of oxygen from the protein sample in 50 mM phosphate buffer at pH 5.1 (Procedure 4, Figure 2D). While Procedure 4 succeeded in preventing hydroxylation by introducing glutathione as a hydrogen atom donor and DTT, the bubbling of nitrogen through the protein solution resulted a substantial loss of protein concentration. In addition, deselenization was incomplete and an additional mass peak was obtained, which is difficult to interpret.

Discussion

Reduction of selenopeptides to their corresponding alanine analogs is a common practice in NCL and EPL protocols (Derry et al., 2015; Metanis et al., 2010). Historically, deselenization was preceded by desulfurization of cysteine (Yan and Dawson, 2001). The attraction of selenocysteine is the high propensity of selenols to form radicals, which allows selective deselenization in the presence of native and unprotected cysteine residues (Metanis et al., 2010). Deselenization thus has emerged as the most preferable method in peptide or protein ligation protocols, which originally were based on cysteine. Moreover, deselenization conditions can be controlled to convert Sec to serine (Dery et al., 2015; Malins et al., 2015; Mitchell et al., 2017), allowing greater variation of the ligation site. Importantly, the radical-mediated deselenization mechanism by TCEP has proven to yield highly chemo- and enantioselective products (Dery et al., 2015; Metanis et al., 2010). Finally, as described in Chapter 4, site-specific incorporation of Sec by genetic encoding as a photocaged UAA enables synthesis of selenoproteins, which may also contain cysteine. Notably, photolysis of PSc does not yield dehydroalanine (see Chapter 4). Dehydroalanine is not very useful, as conversion of dehydroalanine to serine does not retain stereoselectivity (Bernardes et al., 2008; Metanis et al., 2010).

Therefore, in principle, incorporation of isotope-labelled PSc and its subsequent photolysis and deselenization would be ideal to achieve site-selective isotope-labelling of

alanine and serine. Notably, however, site-selective incorporation of isotope-labelled serine can be achieved in a more straightforward way using a photocaged serine UAA (Lemke et al., 2007). The main attraction of the project thus remains the site-selective isotope-labelling of alanine.

It was disappointing to find that reduction of Sec to alanine is so susceptible to oxygen. Presence of Na_2SO_3 as oxygen scavenger significantly enhanced the relative amount of alanine obtained. Bubbling nitrogen through the protein solution completely eliminated the conversion to serine, but this resulted in a severe loss of protein, limiting any further applications and produced a new mass peak the origin of which remains to be determined. Besides incomplete deselenization, the mass spectra indicated the buildup of adducts. Clearly, the deselenization reaction conditions need to be optimized further to obtain complete and pure conversion of Sec to alanine.

Conclusion

Reduction of Sec to alanine was achieved by the reaction conditions we selected, but further optimization is needed to obtain complete and pure conversion without substantial loss of protein.

References

Bernardes, G. J. L., Grayson, E. J., Thompson, S., Chalker, J. M., Errey, J. C., El Oualid, F., Claridge, T. D. W. and Davis, B. G. (2008) From disulfide-to thioether-linked glycoproteins. *Angew. Chem. Int. Ed*. **120**, 2276–2279.

Cellitti, S. E., Jones, D. H., Lagpacan, L., Hao, X., Zhang, Q., Hu, H., Brittain, S. M., Brinker, A., Caldwell, J., Bursulaya, B., Spraggon, G., Brock, A., Ryu, Y., Uno, T., Schultz, P. G. and Geierstanger, B. H. (2008**)** *In vivo* incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **130**, 9268−9281.

Deiters, A., Groff, D., Ryu, Y., Xie, J. and Schultz, P. G. (2006) A genetically encoded photocaged tyrosine. *Angew. Chem. Int. Ed*. **45**, 2728–2731.

Dery, S., Reddy, P. S., Dery, L., Mousa, R., Dardashti, R. N. and Metanis, N. (2015) Insights into the deselenization of selenocysteine into alanine and serine. *Chem. Sci*. **6**, 6207–6212.

Ellman, J. A., Volkman, B. F., Mendel, D., Schulz, P. G. and Wemmer, D. E. (1992) Site-specific isotopic labeling of proteins for NMR studies. *J. Am. Chem. Soc*. **114**, 7959–7961.

Gautier, A., Nguyen, D. P., Lusic, H., An, W., Deiters, A. and Chin, J. W. (2010) Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* **132**, 4086–4088.

Jia, X., Ozawa, K., Loscha, K. and Otting, G. (2009) Glutarate and N-acetyl-L-glutamate buffers for cell-free synthesis of selectively ^{15}N-labelled proteins. *J. Biomol. NMR* **44**, 59–67.

Kainosho, M. and Güntert, P. (2010) SAIL–stereo-array isotope labeling. *Q. Rev. Biophys*. **42**, 247–300.

Kigawa, T., Muto, Y. and Yokoyama, S. (1995) Cell-free synthesis and amino acid-selective stable-isotope labelling of proteins for NMR analysis. *J. Biomol. NMR* **6**, 129–134.

LeMaster, D. M. (1990) Deuterium labeling in NMR structural analysis of larger proteins. *Quart. Rev. Biophys.* **23**, 133–174.

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.

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.

McIntosh, L. P. and Dahlquist, F. W. (1990) Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear-magnetic resonance spectra of proteins. *Quart. Rev. Biophys.* **23**, 1–38.

Metanis, N., Keinan, E. and Dawson, P. E. (2010) Traceless ligation of cysteine peptides using selective deselenization. *Angew. Chem. Int. Ed*. **49**, 7049−7053.

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.

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.

Ozawa, K., Headlam, M. J., Schaeffer, P. M., Henderson, B. R., Dixon, N. E. and Otting, G. (2004) Optimization of an *Escherichia coli* system for cell-free synthesis of selectively ^{15}N-labelled proteins for rapid analysis by NMR spectroscopy. *Eur. J. Biochem*. **271**, 4084–4093.

Ozawa, K., Wu, P. S. C., Dixon, N. E. and Otting, G. (2006) ^{15}N-Labelled proteins by cell-free protein synthesis: strategies for high-throughput NMR studies of proteins and protein-ligand complexes. *FEBS J.* **273**, 4154–4159.

Shi, J., Pelton, J. G., Cho, H. S. and Wemmer, D. E. (2004) Protein signal assignments using specific labeling and cell-free synthesis. *J. Biomol. NMR* **28**, 235–247.

Shortle, D. (1994) Assignment of amino-acid type in $^{1}H–^{15}N$ correlation spectra by labeling with ^{14}N-amino acids. *J. Magn. Reson. B.* **105**, 88–90.

Sobhanifar, S., Reckel, S., Junge, F., Schwarz, D., Kai, L., Karbyshev, M., Löhr, F., Bernhard, F. and Dötsch, V. (2010) Cell-free expression and stable isotope labelling strategies for membrane proteins. *J. Biomol. NMR* **46**, 33–43.

Su, X. –C., Loh, C. –T., Qi, R. and Otting, G. (2011) Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enzymes for selective ^{15}N-labelling and production of perdeuterated proteins in H_2O. *J. Biomol. NMR* **50**, 35–42.

Takeuchi, K., Ng, E., Malia, T. J. and Wagner, G. (2007) 1–13C amino acid selective labeling in a $^{2}H^{15}N$ background for NMR studies of large proteins. *J. Biomol. NMR* **38**, 89–98.

Waugh, D. S. (1996) Genetic tools for selective labeling of proteins with alpha-^{15}N-amino acids. *J. Biomol. NMR* **8**, 184–192.

Whittaker, J. W. (2007) Selective isotopic labeling of recombinant proteins using amino acid auxotroph strains. *Methods Mol. Biol*. **389**, 175–188.

Wu, N., Deiters, A., Cropp, T. A., King, D. and Schultz, P. G. (2004) A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 4306–14307.

Yabuki, T., Kigawa, T., Dohmae, N., Takio, K., Terada, T., Ito, Y., Laue, E. D., Cooper, J. A., Kainosho, M. and Yokoyama, S. (1998) Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. *J. Biomol. NMR* **11**, 295–306.

Yan, L. Z. and Dawson, P. E. (2001) Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **123**, 526−533.

CHAPTER 6

CELL-FREE PROTEIN SYNTHESIS WITH PYRROLYSYL-tRNA SYNTHETASE/ $tRNA_{CUA}$ PAIR

Contribution

I carried out all the experiments discussed in this chapter including construction of all the protein and the suppressor tRNA expression vectors and constructs, expression and purification of all the synthetases, expression and extraction of suppressor tRNA and execution of the CFPS experiments. Dr Heinz Neumann kindly provided the original plasmid pCDF Duet_4-components_lac(P1) which was modified and used to extract *M. barkeri suppressor* tRNA. Dr Christoph Nitsche provided the expression plasmid for wild-type ZiPro. CFPS was carried out using commercial UAAs.

Introduction

Pyrrolysyl-tRNA synthetase (PylRS) is an outstanding aaRS enzyme for the incorporation of a range of structurally different UAAs. Unfortunately, its inherent enzymatic activity is lower than that of other aaRSs (Suzuki et al., 2017). As a consequence, the PylRS variants do not give high yields of proteins with unnatural amino acids *in vivo*. Since CFPS allows optimizing the reaction conditions for high protein yields, a CFPS system with PylRS/$tRNA_{CUA}$ is of great interest, especially for expensive UAAs and UAAs that are laborious to synthesize. A CFPS system with purified PylRS has indeed been reported by Seki et al. (2018), but repeated attempts by several group members of our laboratory failed to reproduce these results. Overexpression of the wild-type *M. mazei* and *M. barkeri* synthetases produces only low yields and the purified enzymes do not seem to be active. This has been attributed to the hydrophobic property of the N-terminal domain (NTD) of these synthetases (Jiang and Krzycki, 2012; Suzuki et al., 2018; Yanagisawa et al., 2006). To overcome this situation, one might attempt to stabilize the NTD of these synthetases or find synthetase variants that are active without the NTD.

Genes encoding $tRNA^{Pyl}$ (*pylT*) and PylRS (*pylS*) are found in the archaeal methanogens (*Methanosarcinaceae* family) and in the bacterium *Desulfitobacterium hafniense (Dha)*. Both the *pylT* and the *pylS* genes are highly conserved within the *Methanosarcinaceae* family, but the bacterial genes are markedly different from their archaeal counterparts. In the archaeal PylRS, the structure of the C-terminal catalytic core domain resembles that of class II aaRSs and the C-terminal domain is connected to the N-terminal tRNA binding domain via a linker (Herring et al., 2007). In contrast, the PylRS in *Dha* is encoded by a split gene, where the C-terminal catalytic domain (*Dha* PylRSc) and N-terminal domain (*Dha* PylRSn) are encoded separately by *pylSc* and *pylSn* genes, respectively (Nozawa et al., 2009). *Dha* PylRSc comprises the class II aaRS catalytic core, a tRNA recognition domain, a bulge domain and a carboxy-terminal tail. Both *Dha* PylRSc and the C-terminal domain (CTD) of *M. mazei* PylRS are structurally similar with 39% sequence similarity (Nozawa et al., 2009).

Figure 1. Crystal structure of Dha PylRSc complexed with Dha tRNA$_{CUA}$ (PDB ID: 2ZNI; Nozawa et al., 2009). Pink: catalytic core; red: tRNA binding domain; blue: bulge domain; orange: C-terminal tail; black: Dha tRNA$_{CUA}$.

It has been shown that both *Dha* PylRSc and an N-terminally truncated version of *M. barkeri* PylRS aminoacylate tRNA^{Pyl} *in vitro* (Nozowa et al., 2009; Herring et al., 2007). The K_d value of the complex between full-length *M. barkeri* PylRS and *M. barkeri* tRNA^{Pyl} was reported to be 0.15 μM, i.e. 46-fold lower than the K_d value of *Dha* PylRSc binding to *Dha* tRNA^{Pyl} (6.9 μM). The weaker affinity of *Dha* PylRSc towards *Dha* RNA^{Pyl} is attributed to the absence of N-terminal tRNA binding domain (Herring et al., 2007). It has also been shown in *in vitro* experiments, that *Dha* PylRSc is able to bind tRNA^{Pyl} even in the presence of competing total yeast tRNAs, that *Dha* PylRSc activates *N*-*ε*-cyclopentyloxycarbonyl-L-lysine (Cyc) and aminoacylates *Dha* tRNA^{Pyl} with Cyc (Nozawa et al., 2009; Herring et al., 2007).

Two contradictory observations have been reported regarding the *in vivo* activity of *Dha* PylRSc. Herring et al. (2007) reported that the naturally truncated enzyme, *Dha* PylRSc does not make favourable interactions with its cognate tRNA in the presence of competition by non-cognate cellular tRNA, whereas Nozawa et al. (2009) pointed out that unfractionated *E. coli* tRNA is not a substrate for *Dha* PylRS and, more importantly, that Cyc-tRNA^{Pyl} formation is not perturbed by competition with the tRNA pool in *E. coli*. In conclusion, Herring et al. (2007) argued that the amino terminus, the function of which is to recognize and bind tRNA, is required for *Dha* PylRSc enzymes *in vivo* in the presence of competitive non-cognate tRNA. In contrast, Jiang (2013) pointed out that the concentration of tRNA^{Pyl} is the limiting factor for the *in vivo* activity of *Dha* PylRSc and this can be overcome by transcribing tRNA^{Pyl} from a high-copy-number plasmid under a strong T7 promoter to outcompete the *E. coli* tRNA pool. This observation can be

rationalized by the finding that *Dha* PylRSc shows 85% of the aminoacylation activity of full-length archaeal PylRS and aminoacylates 80% of $tRNA^{Pyl}$ with Cyc (Nozawa et al., 2009). Under the same conditions, N-terminally truncated *M. mazei* PylRS acylates only 53% of $tRNA^{Pyl}$ with Cyc. This difference can be attributed to the well-ordered and more stable tRNA binding domain of *Dha* PylRSc compared to the C-terminal domain of *M. mazei* PylRS. Furthermore, the C-terminal tail of *Dha* PylRSc also participates in $tRNA^{Pyl}$ recognition (Nozawa et al., 2009).

Considering that *Dha* PylRSc is soluble and stable (Jiang, 2013) and its activity is not improved by the additional presence of separately purified *Dha* PylRSn (Nozawa et al., 2009), *Dha* PylRSc appears like a good candidate for developing a CFPS system with the PylRS/$tRNA_{CUA}$ pair. Notably, however, as pointed out by Jiang (2013), it would be necessary to maintain a high $tRNA^{Pyl}$ concentration in the CFPS reaction mixture.

In conventional CFPS systems, orthogonal amber suppressor tRNA (o-tRNA) is provided as a mixture of total tRNA purified from *E. coli*. Therefore, the actual amount of o-tRNA may not be sufficient to outcompete non-cognate *E. coli* tRNA. To selectively increase the amount of suppressor tRNA in the cell-free reaction mixture, it could be transcribed *in vitro*, but *in vitro* transcription is expensive. A more attractive method would be *in situ* transcription of the o-tRNA in the cell-free reaction mixture as reported previously (Albayrak and Swartz., 2013; Hong et al. 2014). This can conveniently be performed using PCR-amplified linear DNA, where the o-tRNA is transcribed under a T7 promoter in a fusion with a hammerhead ribozyme. The hammerhead ribozyme is a transzyme that spontaneously cleaves itself during the CFPS to liberate the o-tRNA (Figure 2). In addition, the construct contains a 500 nucleotide overhang upstream of the T7 promoter to protect the DNA template from enzymatic degradation (Hong et al., 2014). As the concentration of transcribed o-tRNA is proportional to the template concentration, this can be used to optimize the *Dha* $tRNA_{CUA}$ concentration in the CFPS and hence boost the *in vitro* activity of *Dha* PylRSc.

Figure 2. Schematic diagram showing the hammerhead ribozyme construct with Dha $tRNA_{CUA}$ used. The hammerhead ribozyme (HH ribozyme) was inserted between the T7 promoter and Dha $tRNA_{CUA}$. The 500 base pair (bp) overhang protects the DNA template from enzymatic degradation.

Chimeric PylRS (ch-PylRS) is a fusion of the *M. barkeri* PylRS NTD (residues 1–149) and the *M. mazei* PylRS CTD (residues 185–454). It has been reported to possess similar kinetic parameters as the wild-type archaeal PylRS enzymes while showing much better solubility (>15 mg/mL) than pure *M. mazei* PylRS or *M. barkeri* PylRS (~2 mg/mL) (Suzuki et al., 2017).

Figure 3. Schematic diagram of the chimeric PylRS construct.

The ch-PylRS construct would thus be expected to allow making synthetase stock solutions of sufficient concentration for use in CFPS. Furthermore, the CFPS can be performed in the presence of zinc to stabilize the NTD, which contains a Zn^{2+} binding site (Suzuki et al., 2017). In addition, the CFPS can be carried out with total *E. coli* tRNA without specifically increasing the concentration of suppressor tRNA, as the intact tRNA binding domain of *M. barkeri* PylRS present in ch-PylRS makes the suppressor tRNA concentration less critical.

The objective of this work was to develop a CFPS system with the orthogonal PylRS/$tRNA_{CUA}$ pair to incorporate UAAs into proteins via amber codon suppression. The following PylRS/$tRNA_{CUA}$ combinations were tested:

a) Purified *Dha* PylRSc, with *Dha* $tRNA_{CUA}$ co-produced in the cell-free reaction mixture,
b) Purified *Dha* PylRSc and total *E. coli* tRNA containing *Dha* $tRNA_{CUA}$,
c) Purified ch-PylRS and total *E. coli* tRNA containing *M. barkeri* $tRNA_{CUA}$.

The activity of the *Dha* PylRSc/$tRNA_{CUA}$ pair was tested for its ability to incorporate the UAA N^{ε}-trifluoroacetyl-lysine (CF_3-AcK; Figure 4) in response to an amber mutant of the Zika virus NS2B-NS3 protease, ZiPro V36U. The activity of the ch-PylRS/$tRNA_{CUA}$ pair was tested for its ability to incorporate the UAA Boc-lysine (BocK; Figure 4) into ZiPro V36U.

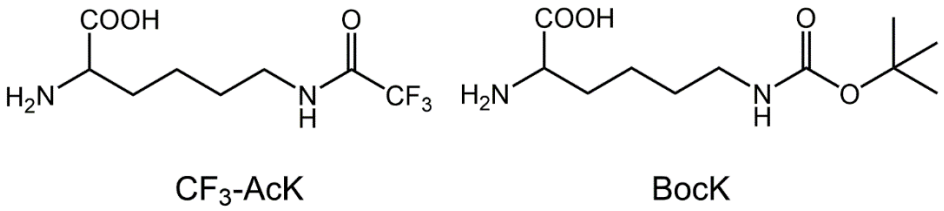


Figure 4. Chemical structures of the unnatural amino acids used in this study.

Experimental procedure

Plasmid construction

The wild-type genes of *M. barkeri* PylRS and *Dha* PylRSc were cloned into the T7 expression vector pETMCSI (Neylon et al., 2000) between the *Nde*1 and *Eco*R1 sites. Both constructs contained an N-terminal His_6-tag followed by a tobacco etch virus (TEV) protease recognition site. The *Dha pylT* gene was flanked by the *lpp* promoter at the 5' end and the *rrnC* terminator at the 3' end and inserted at the GAATTC site of the pETMCSI vector (Neylon et al., 2000).

The DNA fragment encoding the N-terminal 149 amino acid residues of *M. barkeri* PylRS (*M. barkeri* NTD) and the DNA fragment encoding the amino acid residues 185–454 of *M. mazei* PylRS (*M. mazei* CTD) including a C-terminal His_6 tag were cloned as a fusion into the pETMCSI vector between the *Nde*1 and *Eco*R1 sites. The *M. mazei* CTD also contained the mutation Y384F for improved incorporation of BocK (Yanagisawa et al., 2008). A pCDF plasmid containing *M. mazei* PCC2RS and *M. barkeri* $tRNA_{CUA}$ was used to produce total *E. coli* tRNA containing *M. barkeri* $tRNA_{CUA}$.

The ZiPro mutant contained a C-terminal His_6-tag and was cloned into the *Nde*I and *Eco*RI sites of the T7 expression vector pETMSCI (Neylon et al., 2000). The wild-type ZiPro construct was described previously by Mahawaththa et al. (2017). 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.

Expression and purification of *Dha* PylRSc

Dha PylRSc was produced in *E. coli* BL21(DE3) pLysS cells. The cells were transformed with the plasmid pETMCSI-His_6-TEV-*Dha* PylRSc and a starter culture was grown for

16 h at 37 °C in LB medium supplemented with 100 µg/mL ampicillin and 33 µg/mL chloramphenicol. Overnight cultures were inoculated into fresh LB medium (1:100 dilution), supplemented with ampicillin and chloramphenicol and grown at 37 °C. After the cultures reached an OD_{600} value of 0.5–0.8, IPTG was added to a final concentration of 1 mM. The cells were harvested following overnight expression. The cell pellet was resuspended in buffer A (50 mM phosphate, 300 mM NaCl, 10% glycerol, pH 7.5) and lysed by French press at 12,000 psi and at 4 °C. The lysate was clarified by centrifugation (17,300 *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). Afterwards, the fraction containing His_6-TEV-*Dha* PylRSc were combined and incubated with His_6-tagged TEV protease in buffer C (50 mM phosphate, 50 mM NaCl, 2 mM 2-mercaptoethanol, pH 8) at 4 °C for 16 h. The cleaved N-terminal His_6-tagged peptide and the His_6-tagged TEV protease were separated from *Dha* PylRSc by passing the mixture again over the Co-NTA column. Finally, the purified protein was dialyzed into buffer D (50 mM phosphate, pH 7.5, 150 mM NaCl, 1 mM DTT, 10% glycerol) for storage at -80 °C.

Expression and purification of *M. barkeri* PylRS

M. barkeri PylRS was produced in *E. coli* BL21(DE3) pLysS. The cells were transformed with the plasmid pETMCSI- His_6-TEV-*M. barkeri* PylRS and the protein was expressed as described above for the expression *Dha* PylRSc. The cells were harvested following overnight expression. The cell pellet was resuspended in buffer E (50 mM HEPES, 500 mM NaCl, 10% glycerol, pH 7.2). Lysis by French press, clarification of the lysate and purification by a 5 mL Co-NTA column were performed as described above for *Dha* PylRSc, except that buffer E was used for chromatography and buffer F (buffer E containing, in addition, 300 mM imidazole) was used for elution. The fractions containing His_6-TEV-*M. barkeri* PylRS were combined and treated with TEV protease. The His_6-tagged TEV protease was separated from *M. barkeri* PylRS as described above for *Dha* PylRSc. Finally, the purified protein was dialyzed into buffer G (50 mM HEPES, pH 7.2, 150 mM NaCl, 1 mM DTT, 10% glycerol) for storage at -80 °C.

Expression and purification of ch-PylRS

Ch-PylRS was produced in *E. coli* BL21(DE3) pLysS. The cells were transformed with the plasmid pETMCSI-ch-PylRS-His_6 and a starter and overnight culture were grown to an OD_{600} value of 0.5–0.8 as described above for the expression *Dha* PylRSc. IPTG was added to final concentrations of 1 mM. Since the addition of Zn^{2+}leads to increased solubility of the NTD (Suzuki et al., 2017), the LB medium was supplemented with 20 µM $ZnCl_2$ at the time of induction with IPTG (final concentration 1 mM). The cells were harvested following 16 h expression at room temperature. The cell pellet was resuspended in buffer H (50 mM Tris-HCl, 300 mM NaCl, 5 mM $MgCl_2$, 20 µM $ZnCl_2$, pH 7.4). Lysis by French press, clarification of the lysate and purification by a 5 mL Co-NTA column were performed as described above for *Dha* PylRSc, except that buffer H was used for chromatography and buffer I (buffer H containing, in addition, 300 mM imidazole) was used for elution. Finally, the buffer H in the purified protein was exchanged into buffer J (20 mM Tris-HCl, 150 mM NaCl, 1 mM TCEP, 20 µM $ZnCl_2$, 5 mM $MgCl_2$, 20% glycerol, pH 7.4) using an Amicon centrifugal filter (MWCO 10 kDa) for storage at -80 ^oC.

Expression and purification of total tRNA containing *Dha* suppressor $tRNA_{CUA}$

Total tRNA containing *Dha* $tRNA_{CUA}$ was purified from *E. coli* BL21(DE3) cells containing the plasmid pETMCSI-*Dha pylT*. A starter culture was grown for 16 h at 37 ^oC in LB medium supplemented with 100 µg/mL ampicillin. Overnight cultures were inoculated into fresh LB medium (1:100 dilution), supplemented with 100 µg/mL ampicillin and 0.18% glucose and grown at 37 ^oC. After the cultures reached an OD_{600} value of 3.0, the cells were harvested and the cell pellet was resuspended in buffer K (0.3 M potassium acetate, pH 4.8). After adding an equal volume of water-saturated phenol, the mixture was shaken at room temperature for 60 min. Then the mixture was centrifuged at 18,000 rpm for 30 min at room temperature and the supernatant was subjected to ethanol precipitation. The pellet was air-dried and resuspended in sterile Milli-Q water. About 15 mg of tRNA was purified from 6 L of LB medium. The concentration of total tRNA was measured by its absorbance at 260 nm.

Expression and purification of total tRNA containing *M. barkeri* suppressor $tRNA_{CUA}$

Total tRNA containing *M. barkeri* $tRNA_{CUA}$ was purified from *E. coli* BL21(DE3) cells containing a pCDF plasmid with the *M. mazei* PCC2RS and *M. barkeri* $tRNA_{CUA}$ genes. A starter culture was grown for 16 h at 37 ºC in LB medium supplemented with 50 µg/mL spectinomycin. Overnight cultures were inoculated into fresh LB medium (1:100 dilution), supplemented with 50 µg/mL spectinomycin and 0.18% glucose and grown at 37 ºC. Culture growth and purification of the total tRNA proceded as described above for total tRNA containing *Dha* $tRNA_{CUA}$. About 10 mg of tRNA was purified from 4 L of LB medium. The concentration of total tRNA was measured by its absorbance at 260 nm.

Hammerhead construct of *Dha* $tRNA_{CUA}$

TTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCG
GTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGC
ATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCC
ACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGACGCTC
TCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGT
TGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAG
TCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATG
AGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGG
CGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGG
CGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGA
GACCGGCTGATGAGTCCGTGAGGACGAAACGGTACCCGGTACCGTCGGGGG
GTGGATCGAATAGATCACACGGACT**CTA**AATTCGTGCAGGCGGGTGAAACTCC
CGTACTCCCCGCCA

Figure 5. Nucleotide sequence of the hammerhead construct of Dha $tRNA_{CUA}$ used. Black: 500 nucleotide overhang; green: T7 promoter; orange: hammerhead ribozyme; blue: Dha $tRNA_{CUA}$.

The above construct was PCR-amplified using the forward primer 5'-TTCCGAATACCGCAGCGACAG-3' and the reverse primer 5'-TGGCGGGGAGTACGGGAGTTTC-3' to produce linear DNA for use in the CFPS. The PCR-amplified DNA was subjected to ethanol precipitation without PCR purification. The DNA pellet was resuspended in Tris-HCl buffer, pH 7.5.

Cell-free protein synthesis with *Dha* PylRS

The PCR-amplified gene (Wu et al., 2007) for ZiPro V36U was used as DNA template to evaluate the activity of the *Dha* PylRS/$tRNA_{CUA}$ pair. The protein was expressed by continuous exchange CFPS with the purified total tRNA and synthetase. The CFPS was carried out at 30 ^{o}C for 16 h as described previously by Apponyi et al. (2008), Loscha et al. (2012) and Ozawa et al. (2012) with minor modifications. The purified *Dha* PylRSc was added to the inner reaction mixture at a final concentration of 50 μM. *Dha* $tRNA_{CUA}$ was provided in the inner reaction mixture, either as 0.3 mg/mL total *E. coli* tRNA containing *Dha* $tRNA_{CUA}$ or as the PCR-amplified hammerhead construct of *Dha* $tRNA_{CUA}$ for *in-situ* transcription of *Dha* $tRNA_{CUA}$ in the cell-free reaction mixture. A concentration series of the DNA template of the hammerhead construct was used (10, 15, 20, 30 ng/μL final concentrations). 1 mM CF_3-AcK (Merck Schuchardt OHG, Hohenbrunn, Germany) was used in the outer reaction mixture. The volumes of the inner and the outer reaction mixtures were 200 μL and 2 mL, respectively.

Cell-free protein synthesis with ch-PylRS

In the same way, ZiPro V36U was used to evaluate the activity of the ch-PylRS/$tRNA_{CUA}$ pair. The protein was expressed by continuous exchange CFPS with the purified ch-PylRS. CFPS was carried out as described above for *Dha* PylRS, except that freshly purified ch-PylRSc was added to the inner reaction mixture at a final concentration of 20 μM and tRNA was provided in the inner reaction mixture as 0.3 mg/mL total *E. coli* tRNA containing *M. barkeri* $tRNA_{CUA}$. 1 mM BocK (CreoSalus, Advanced ChemTech, Inc.) was used in the outer reaction mixture and 20 μM $ZnCl_2$ was present in both inner and outer reaction mixtures. In addition, DTT in the cell-free reaction was replaced with 50 mM TCEP to maintain the concentration of free zinc, and DTT in the S30 extract was replaced with TCEP by dialysis prior to use in the cell-free reaction. The volumes of the inner and outer reaction mixtures were 200 μL and 2 mL, respectively.

Results

Dha PylRS was purified from *E. coli* BL21(DE3) pLysS cells in good yield (17 mg from 1 L of LB medium) and solubility (1 mM), which was sufficient to achieve a 50 μM concentration in the cell-free reaction mixture. *In vivo*, the UAA BocK is efficiently incorporated by the wild-type *M. mazei* synthetase and its Y384F mutant (Yanagisawa et

al., 2008). Since the amino-acid binding pocket of *Dha* PylRS is tighter than those of the *M. mazei* and *M. barkeri* synthetases, the activity of the synthetase was tested with CF_3-AcK, which is less bulky than BocK. Disappointingly, expression of full-length ZiPro was not observed, neither with the total *E. coli* tRNA nor with increasing concentrations of the hammerhead construct of *Dha* $tRNA_{CUA}$ (Figure 6A).

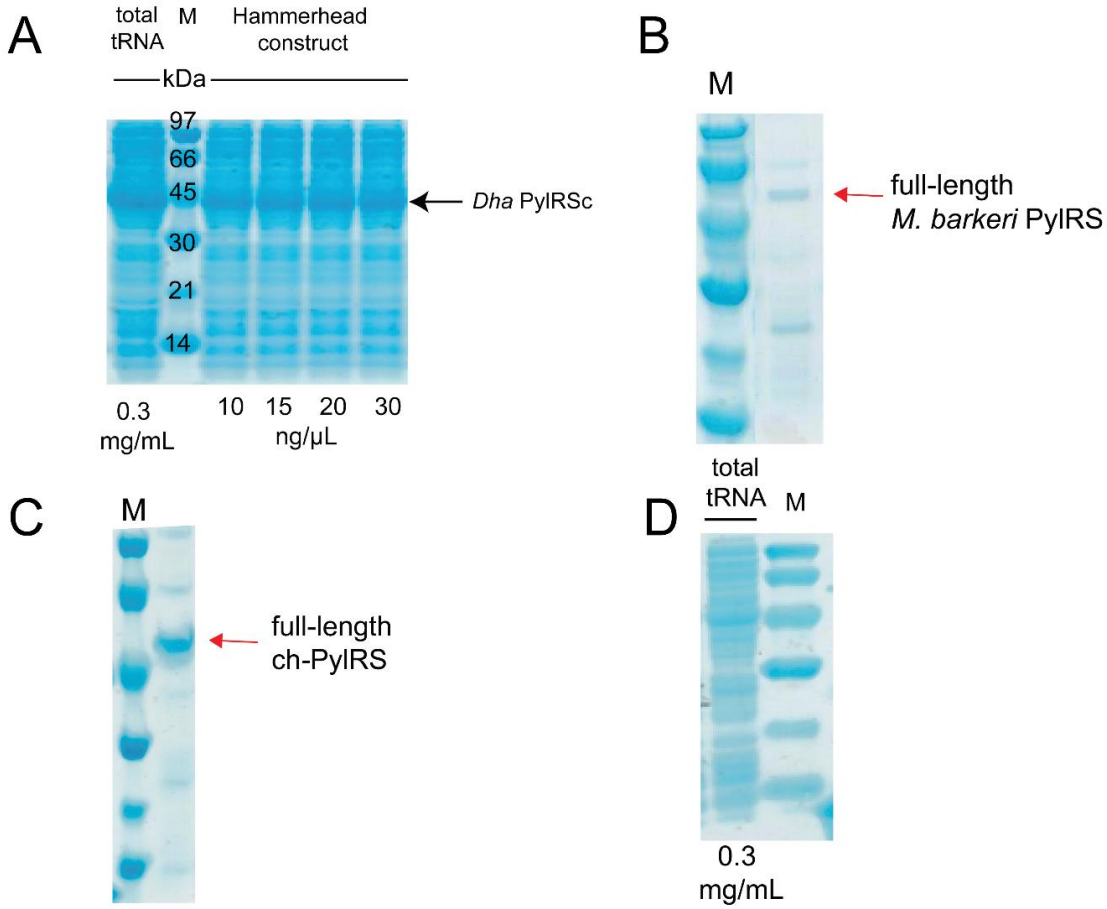


Figure 6. SDS-PAGE analysis of the wild-type M. barkeri PylRS and ch-PylRS purification and CFPS with Dha PylRSc and ch-PylRS. M: Marker lane. The same marker bands are present in all four gels. A) CFPS with Dha PylRSc, using either total tRNA extract (left-most lane) or the hammerhead construct of Dha $tRNA_{CUA}$ *(lanes 3–6). The band of Dha PylRSc added to the CFPS mixture is marked. B) Purified wild-type M. barkeri PylRS. C) Purified ch-PylRS. D) CFPS with ch-PylRS and total tRNA extract. The concentration of ch-PylRS (MW = 48.4 kDa) was too low (starting from 20 µM, but presumably precipitating during CFPS) to be visible on the gel.*

Although *E. coli* BL21(DE3) pLysS cells expressed full-length ch-PylRS enzyme in good yield, the protein started to precipitate during buffer exchange into buffer J. Thus, the maximum solubility of the stock solution obtained was 100 µM, limiting the final

concentration of ch-PylRS in the cell-free reaction mixture to 20 μM. The activity of the synthetase was evaluated by its ability to charge its cognate $tRNA_{CUA}$ with BocK. As shown by SDS-PAGE analysis (Figure 6D), full-length protein production was not detected.

Discussion

Chemla et al. (2015) identified the poor solubility of full-length PylRS as a major limiting factor in cell-free reaction systems performed with the PylRS/$tRNA_{CUA}$ pair and demonstrated a functioning CFPS system, where the PylRS was endogenously expressed during the preparation of the *E. coli* S30 extract. This bypasses the PylRS purification step and alleviates the problem associated with the poor solubility of the PylRS, possibly aided by binding of the NTD to cognate $tRNA_{CUA}$. This approach, however, limits the appeal of the system for the incorporation of structurally different UAAs because, unless the synthetase mutant is polyspecific, different synthetase mutants are required to incorporate different UAAs, and S30 extract preparation with different synthetase mutants is time consuming and expensive. It would be much more attractive, if purified *Dha* PylRSc could be used, which was reported to aminoacylate its cognate tRNA without a solubility-compromising tRNA-binding NTD (Nozowa et al., 2009; Herring et al., 2007; Jiang and Krzycki, 2012; Jiang, 2013; Suzuki et al., 2017; Yanagisawa et al., 2006). Indeed, the present work showed that the expression level of the synthetase and its solubility were well-suited for use in cell-free reaction systems but, unfortunately, despite a high concentration of *Dha* PylRSc in the cell-free reaction (Figure 6A), no full-length ZiPro V36U could be produced, neither with total *E. coli* tRNA extract containing *Dha* $tRNA_{CUA}$ nor with a hammerhead ribozyme construct with *Dha* $tRNA_{CUA}$. These results suggest that the purified *Dha* PylRSc enzyme was inactive, did not bind the UAA or failed to bind the amber suppressor tRNA due to insufficient affinity in the absence of the *Dha* PylRSn domain.

Notably, *Dha* PylRSc lacks the tRNA binding domain, retaining only the catalytic core that binds the UAA. A previous report mentioned that *Dha* PylRSc is active only *in vitro* but not *in vivo* (Herring et al., 2007), although the compositions of components available for recombinant protein synthesis are quite similar in both situations. As Jiang (2013) pointed out the benefit of an excess of *Dha* $tRNA_{CUA}$ for the *in vivo* activity of *Dha* PylRSc, the present work prepared the total *E. coli* tRNA extract with an excess of suppressor tRNA and, in addition, attempted CFPS with increasing quantities of *Dha*

$tRNA_{CUA}$ made in the cell-free reaction mixture from increasing concentrations of DNA template of its hammerhead construct. The failure to obtain full-length protein could be ascribed to the requirement that *Dha* $tRNA_{CUA}$ must have the correct CCA 3′ end to be active as tRNA. In principle, run-off transcription with the T7 RNA polymerase in the cell-free reaction mixture could result in an incorrect 3' end (Draper et al., 1998; Kholod et al., 1998; Milligan and Uhlenbeck, 1989; Triana-Alonso et al., 1995), but a similar tRNA production system based on *M. jannashii* $tRNA_{CUA}$ had successfully been used before in the lab (E. Abdelkader, personal communication). It is possible that the NTD of *Dha* PylRS is essential for amber codon suppression with UAAs in cell-free protein synthesis experiments.

As the recent crystal structure of the NTD of *M. mazei* PylRS revealed a zinc binding site (Suzuki et al., 2017), the present work shifted the focus on stabilizing the NTD of the full-length ch-PylRS enzyme, which is more soluble than wild-type full-length PylRSs. Conservation of the native NTD structure was attempted by provision of zinc during ch-PylRS synthesis in *E. coli* and maintaining 20 μM Zn^{2+} concentration in the buffers used during protein purification. For good protein yield, CFPS reactions must be conducted under reducing conditions, which is usually achieved with DTT, but DTT binds Zn^{2+} (Krężel et al., 2001) with micromolar binding affinity (Krężel and Maret, 2016). Therefore, reducing conditions were maintained using TCEP. Modest concentrations of zinc are tolerated in CFPS reactions (Carruthers et al., 2014) and the solubility or expression yields of the synthetase were not affected by the presence of Zn^{2+}. Although the protein was stable and soluble during FPLC purification, it started to aggregate upon removal of imidazole and during concentration. Protein aggregation could have been triggered by Zn^{2+} binding to the His_6 tag. A control experiment without Zn^{2+} in the buffer J, however, also resulted in precipitation of the ch-PylRS, suggesting that the precipitation is not simply due to binding of Zn^{2+} to the His_6 tag.

With the solubility of ch-PylRS limited to 100 μM, the concentration of ch-PylRS in the cell-free reaction mixture may have been too low to yield significant amounts of ZiPro V36U. To ensure maximum activity of the ch-PylRS, its purification and the subsequent CFPS reaction were carried out on the same day. The failure to produce full-length ZiPro V36U with BocK suggests either very rapid aging of the enzyme or the need for much higher concentrations of PylRS and $tRNA_{CUA}$.

Conclusion

At the time of the present work, *Dha* PylRSc was the only PylRS, for which activity had been reported in the absence of the NTD. Unfortunately, the activity could not be confirmed in CFPS reactions. Most recently, however, a new N-terminally truncated PylRS variant was reported to have an *in vivo* activity superior to those of the wild-type *M. mazei* and *M. barkeri* enzymes (Willsis and Chin, 2018). This may offer an outstanding solution for a functioning CFPS system with the orthogonal PylRS/$tRNA_{CUA}$ pair.

References

Albayrak, C. and Swartz, J. R. (2013) Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res.* **11**, 5949–5963.

Apponyi, M., Ozawa, K., Dixon, N. E. and Otting, G. (2008) Cell-free protein synthesis for analysis by NMR spectroscopy *Methods in Molecular Biology, Structural proteomics: high-throughput methods B. Kobe, M. Guss, T. Huber Eds., Humana Press, Totowa, USA* **426**, 257–268.

Carruthers, T. J., Carr, P. D., Loh, C. –T., Jackson, C. J. and Otting, G. (2014) Fe^{3+} located in the dinuclear metallo-β-lactamase IMP-1 by pseudocontact shifts. *Angew. Chem. Int. Ed.* **53**, 14268–14272.

Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V. and Alfonta, L. (2015) Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system. *Biotechnol. Bioeng.* **112**, 1663–1672.

Draper, D. E., White, S. A. and Kean, J. M. (1988) Preparation of specific ribosomal RNA fragments. *Methods Enzymol.* **164**, 221–237.

Herring, S., Ambrogellyb, A., Gundllapallib, S., O'Donoghueb, P., Polycarpob, C. R. and Söll, D. (2007) The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable *in vitro* but required for *in vivo* activity. *FEBS Lett.* **581**, 3197–3203.

Hong, S. H., Ntai, I., Haimovich, D., Kelleher, N. L., Isaacs, F. J. and Jewett, M. C. (2014) Cell-free protein synthesis from a release factor 1 deficient *Escherichia coli* activates efficient and multiple site-specific nonstandard amino acid incorporation. *ACS Synth. Biol.* **3**, 398–409.

Jiang, R. and Krzycki, J. A. (2012) PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. *J. Biol. Chem.* **287**, 32738–32746.

Jiang, R. (2013) Studies of the Pyrrolysyl-tRNA Synthetase. *Dissertation*. The Ohio State University.

Kholod, N., Vassilenko, K., Shlyapnikov, M., Ksenzenko, V. and Kisselev, L. (1998) Preparation of active tRNA gene transcripts devoid of 3′-extended products and dimers. *Nucleic Acids Res.* **26**, 2500–2501.

Krężel, A., Leśniak, L., Jeżowska-Bojczuk, M., Młynarz, P., Brasuñ, J., Kozłowski, H. and Bal, B. (2001) Coordination of heavy metals by dithiothreitol, a commonly used thiol group protectant. *J. Inorg. Biochem.* **84**, 77–88.

Krężel, A. and Maret, W. (2016) The biological inorganic chemistry of zinc ions. *Archives Biochem. Biophys.* **611**, 3–19.

Loscha, K. V., Herlt, A. J., Qi, R., Huber, T., Ozawa, K. and Otting, G. (2012) Multiple-site labeling of proteins with unnatural amino acids. *Angew. Chem. Int. Ed.* **51**, 2243–2246.

Milligan, J. F. and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51–62.

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.

Nozawa, K., O'Donoghue, P., Gundllapalli, S., Araiso, Y., Ishitani, R., Umehara, T., Söll, D. and Nureki, O. (2009) Pyrrolysyl-tRNA synthetase-tRNA^{Pyl} structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163–1167.

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.

Ozawa, K., Loscha, K. V., Kuppan, K. V., Loh, C. T., Dixon, N. E. and Otting, G. (2012) High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. *Biochem. Biophys. Res. Commun.* **418**, 652–656.

Seki, E., Yanagisawa, T. and Yokoyama, S. (2018) Cell-free protein synthesis for multiple site-specific incorporation of noncanonical amino acids using cell extracts from RF-1 deletion *E. coli* strains. *Methods Mol. Biol.* **1728**, 49–65.

Suzuki, T., Miller, C., Guo, L. –T., Ho, J. M. L., Bryson, D. I., Wang, Y. –S., Liu, D. R. and Söll, D. (2017) Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* **13**, 1261–1266.

Triana-Alonso, F. J., Dabrowski, M., Wadzack, J. and Nierhaus, K. H. (1995) Self-coded 3′-extension of run-off transcripts produces aberrant products during in vitro transcription with T7 RNA polymerase. *J. Biol. Chem.* **270**, 6298–6307.

Willis, J. C. W. and Chin, J. W. (2018) Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem.* doi: 10.1038/s41557-018-0052-5.

Wu, P. S. C., Ozawa, K., Lim, S. P., Vasudevan, S., Dixon, N. E. and Otting, G. (2007) Cell-free transcription/translation from PCR amplified DNA for high-throughput NMR studies. *Angew. Chem. Int. Ed.* **46**, 3356–3358.

Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K. and Yokoyama, S. (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N^{ε}-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **15**, 1187–1197.

Yanagisawa, T., Ishii, R., Fukunaga, R., Nureki, O. and Yokoyama, S. (2006) Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*. *Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun.* **62**, 1031–1033.

CHAPTER 7

CELL-FREE PROTEIN SYNTHESIS WITH THE POLYSPECIFIC G2 SYNTHETASE/ $tRNA_{CUA}$ PAIR

Contribution

I carried out all the experiments discussed in this chapter, including the construction of the G2 synthetase expression plasmid, expression and purification of the G2 synthetase, expression and extraction of *Mj* suppressor tRNA and execution of CFPS experiments. The CFPS experiments were carried out with a tetrazine amino acid provided by A/Prof. Bim Graham. The pEvol plasmid containing *Mj* suppressor tRNA was kindly provided by Prof. Peter Schultz and the analyses by mass spectrometry were performed by Dr Yao Wang.

Introduction

UAAs have been designed for many purposes, such as the provision of spectroscopic probes, mimics of post-translational modifications and reactive handles for biorthogonal reactions. Two major challenges in making proteins with these UAAs are their chemical synthesis and the evolution of aaRSs that are capable of charging their cognate tRNA with the UAA. Although the PylRS/$tRNA_{CUA}$ and *Mj* TyrRS/$tRNA_{CUA}$ pairs have been identified as outstanding orthogonal aaRS/tRNA pairs in *E. coli*, evolving their amino acid binding pocket to specifically and selectively bind the UAA in the presence of the natural 20 amino acids is difficult and most often unsuccessful. Fortunately, some of the evolved aaRS mutants have been reported to be capable of activating more than one UAA while maintaining the expected fidelity. These aaRSs have a high permissivity profile and are termed polyspecific aaRSs.

p-CNFRS, which is an active-site mutant of *Mj* TyrRS, is one of the widely used polyspecific aaRSs, which enables the incorporation of over a dozen different UAAs (Loh et al., 2018; Young et al., 2011). Moreover, it maintains high fidelity both in *in vivo* protein synthesis (Young et al., 2011) and CFPS (Ozawa et al., 2012). In subsequent work, Cooley et al. (2014) reported another polyspecific aaRS, named G2, which was derived from *Mj* TyrRS and reported to incorporate bulky UAAs. Originally, the G2 synthetase had been produced by Peeler et al. (2010) to incorporate 4-(2′-bromoisobutyramido)phenylalanine (BibaF), which is an initiator of polymer growth from proteins. Site-specific incorporation of BibaF into an amber mutant of super-folder green fluorescent protein (sfGFP) was confirmed by MS/MS analysis. The G2 synthetase is the active-site mutant Y32G/L65E/F108W/Q109M/D158S of the wild-type *Mj* TyrRS. It has been reported to also incorporate the FRET probe acridon-2-ylalanine (Acd), although there no mass spectrometric evidence was provided to prove the site-specific incorporation of this UAA by the G2 synthetase (Speight et al., 2013). Notably, the G2 synthetase reported by Cooley et al. (2014) contained the L162K mutation in addition to the G2 synthetase reported by Peeler et al. (2010) and Speight et al. (2013). This is important, as residue 162 is part of the amino acid binding pocket and Cooley et al. (2014) used this G2 synthetase variant to demonstrate the polyspecificity of the enzyme, evaluating the ability to incorporate twelve different UAAs into sfGFP in response to an amber stop codon. The performance of the G2 synthetase was established by comparing the yields of sfGFP obtained in the presence of UAAs versus those without UAAs.

Based on the reports of polyspecificity and selectivity for sterically demanding UAAs, we hypothesized that the G2 synthetase could work with UAAs that are not efficiently activated by *p*-CNFRS. Therefore, the G2 synthetase version reported by Cooley et al. (2014) was prepared and a CFPS system with the G2 synthetase and its cognate $tRNA_{CUA}$ established. As the enzyme had been reported to work as well with 4-(6-methyl-s-tetrazine-3-yl)aminophenylalanine (Tet-F; Figure 1A) as with BibaF (Cooley et al., 2014), we used Tet-F as the UAA to evaluate purified recombinant G2 synthetase. In particular, the capability of the G2 synthetase to incorporate Tet-F into an amber mutant of the model protein ERp29 S114U (where U indicates the UAA) was tested. As control experiments, CFPS reactions were also carried out with the G2 synthetase and its cognate $tRNA_{CUA}$ without adding a UAA in the cell-free reaction mixture.

Experimental procedure

Plasmid construction

The gene encoding the G2 synthetase (Cooley et al., 2014) was cloned into the T7 expression vector pETMCSI (Neylon et al., 2000) between the *Nde*I and *Eco*RI sites.

G2 synthetase expression and purification

The G2 synthetase was produced in *E. coli* BL21(DE3) cells. The cells were transformed with the plasmid pETMCSI-G2 and a starter culture was grown for 16 h at 37 ^{o}C in LB medium supplemented with 100 µg/mL ampicillin. Overnight cultures were inoculated into fresh LB medium (1:100 dilution), supplemented with ampicillin and grown at 37 ^{o}C. After the cultures reached an OD_{600} value of 0.5–0.8, IPTG was added to a final concentration of 1 mM. The cells were harvested following overnight expression. The cell pellet was resuspended in buffer A (20 mM Tris-HCl, pH 8) and lysed by French press at 12,000 psi and at 4 ^{o}C. The lysate was clarified by centrifugation (17,300 *g*, 40 min, 4 ^{o}C) and loaded onto an anion exchange diethyl-aminoethyl (DEAE) column 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, 1 M NaCl). Finally, the fractions containing the G2 synthetase were pooled and dialysed into buffer C (50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM DTT, 5% glycerol). The concentrated protein was stored at -80 ^{o}C.

Expression and purification of total tRNA containing *Mj* suppressor $tRNA_{CUA}$

The total tRNA containing *Mj* $tRNA_{CUA}$ was purified from *E. coli* BL21(DE3) cells containing plasmid pEvol-*Mj* $tRNA_{CUA}$ (Young et al., 2010). A starter culture was grown for 16 h at 37 ºC in LB medium supplemented with 33 μg/mL chloramphenicol. Overnight cultures were inoculated into fresh LB medium (1:100 dilution), supplemented with 33 μg/mL chloramphenicol and 0.18% glucose and grown at 37 ºC. Culture growth and purification of the total tRNA proceeded as described in the Chapter 6 for total tRNA containing *Dha* $tRNA_{CUA}$. About 20 mg of tRNA was purified from 6 L of LB medium. The concentration of total tRNA was measured by its absorbance at 260 nm.

Cell-free protein synthesis with the G2 synthetase

The PCR-amplified gene for ERp29 S114U was used as DNA template (Wu et al., 2007) to evaluate the activity of the G2 synthetase/$tRNA_{CUA}$ pair in the CFPS protocol. The protein was expressed by continuous exchange CFPS with the total tRNA and synthetase. The CFPS reactions were carried out at 30 ºC for 16 h as described previously (Apponyi et al., 2008; Loscha et al., 2012; Ozawa et al., 2012) with minor modifications. The purified G2 synthetase was added to the inner reaction mixture at a final concentration of 100 μM and tRNA was provided in the inner reaction mixture as 0.3 mg/mL total *E. coli* tRNA containing *Mj* $tRNA_{CUA}$. 1 mM Tet-F was used in the outer reaction mixture. The volumes of the inner and the outer reaction mixtures were 200 μL and 2 mL, respectively. Proteins were purified with 1 mL Co-NTA gravity columns equilibrated with buffer D (50 mM phosphate buffer, pH 7.5, 300 mM NaCl) and washed with buffer E (same as buffer C but with 20 mM imidazole). Proteins were then eluted with buffer F (same as buffer D but with 300 mM imidazole).

Results

Figure 1B shows the SDS-PAGE analysis of ERp29 S114U, after purification of the cell-free reaction mixture on a Co-NTA column. As anticipated, the G2 synthetase/*Mj* $tRNA_{CUA}$ pair produced full-length ERp29 in the presence of 1 mM Tet-F in the cell-free reaction mixture. Unexpectedly, however, full-length protein was also produced in the absence of Tet-F, indicating that amber codon suppression was also achieved with natural

amino acids. ESI-MS analysis suggests the incorporation of tyrosine both in the presence and absence of the Tet-F in the cell-free reaction mixture (Figure 1C and D).

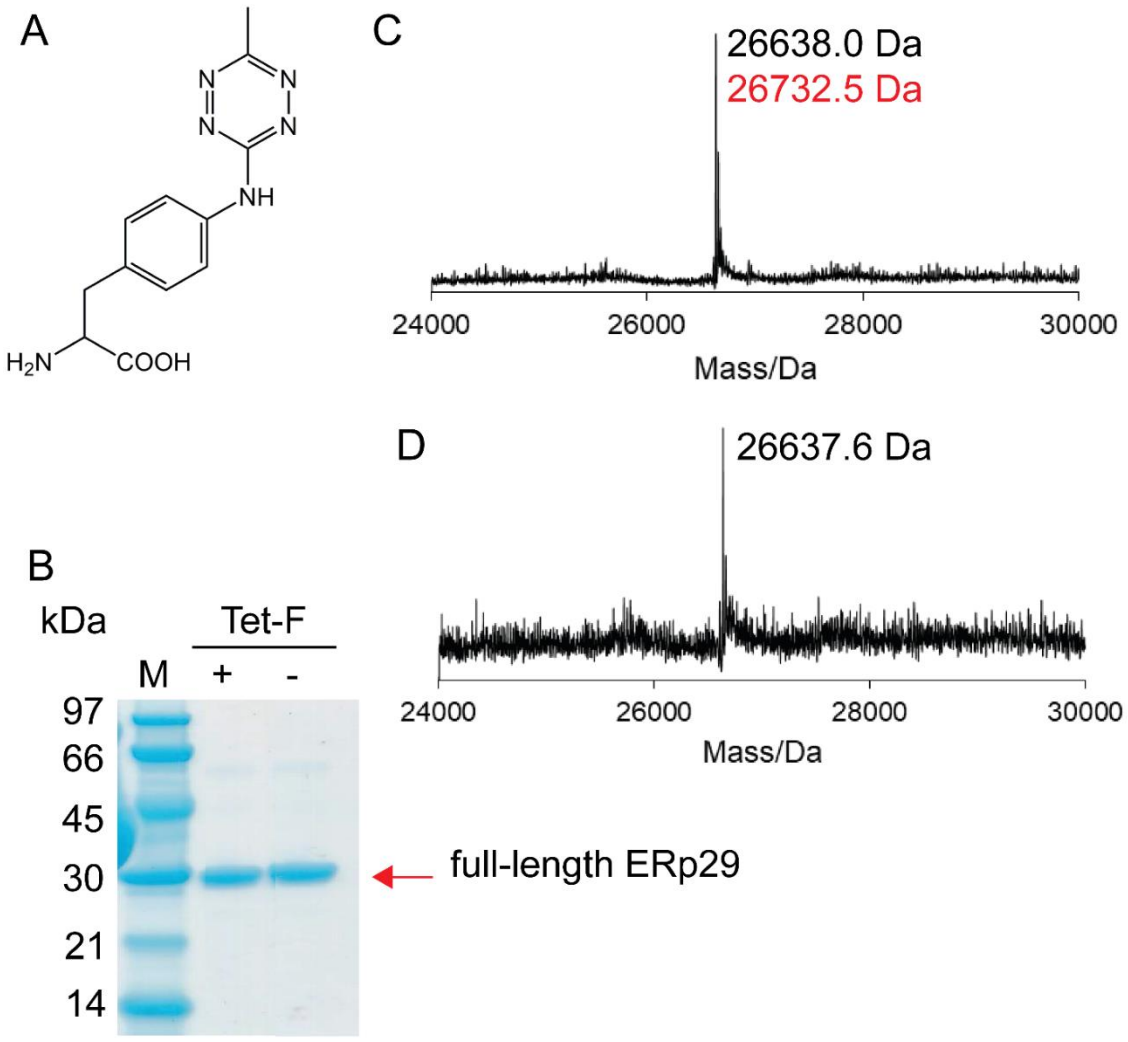


Figure 1. Cell-free synthesis of ERp29 S114U with the G2 synthetase. A) Chemical structure of Tet-F. B) SDS-PAGE analysis of ERp29 S114U expressed in the presence and absence of Tet-F, following purification on a Co-NTA column. C) ESI-MS analysis of ERp29 S114U expressed with 1 mM Tet-F during CFPS. The expected mass is shown in red. D) ESI-MS analysis of ERp29 S114U expressed without adding Tet-F during CFPS.

Discussion

Failure to achieve the expected fidelity in suppressing the amber stop codon with a UAA arises, if the aaRS is not perfectly orthogonal, i.e. supports the insertion of natural amino acids, and an alternative mechanism arises from an imperfect match of the amber codon by near-cognate *E. coli* tRNA (Aerni et al., 2014; Gan and Fan, 2017; Mukai et al., 2010; Nilsson and Rydén-Aulin, 2003; O'Donoghue et al., 2012). Usually, much lower yields of full-length protein in the absence of the UAA serve as indicators of natural amino acid incorporation by near-cognate tRNA, but reduced yields were not observed in the case of the G2 enzyme. MS analysis of the intact protein and MS/MS analysis of trypsin-digested protein provide a sensitive way of confirming UAA incorporation into the target protein and site-specificity of the UAA incorporation. In the present case, the mass of ERp29 S114U could be compared with the masses obtained in previous experiments with different UAAs, where the site-specificity of UAA incorporation had been established.

The discrepancy between the present results and the literature results is disappointing. Speight et al. (2013) successfully incorporated fluorescent Acd with the help of BrbRS3, which has a similar amino acid sequence as the G2 synthetase introduced by Peeler et al. (2010). Although MS evidence was not reported by Speight et al. (2013) to confirm the site-specific incorporation of the UAA, Sungwienwong et al. (2017) recorded MS data of protein with Acd, which showed the simultaneous presence of an unwanted species attributed to a byproduct of Acd synthesis, and the MS does not indicate the incorporation of a natural amino acid. Confusingly, the amino acid sequence of the synthetase utilized by Sungwienwong et al. (2017), which is named AcdRS1, contains the L162K mutation in addition to the mutations present in BrbRS3 (Speight et al., 2013), although the article states that the synthetase reported by Speight et al. (2013) was utilized. The same unacknowledged discrepancy applies to the amino acid sequences reported by Peeler et al. (2010) and Cooley et al. (2014). Except for the article by Sungwienwong et al. (2017), none of the other studies with the G2 synthetase mutant that we utilized (Cooley et al., 2014; Rauch et al., 2016; Sungwienwong et al., 2018; Xuan et al., 2016) present MS evidence for UAA incorporation, although the protein yield was reported to increase in the presence of a UAA. Nonetheless, the reports show a clear difference in the yield of full-length protein in the presence and absence of UAA, suggesting that the G2 synthetase does not promote the incorporation of natural amino acids.

To explain our results, one may note that the published results were obtained *in vivo*, whereas the present data were obtained by CFPS with the purified G2 synthetase. CFPS is carried out in the presence of 1 mM concentration of each of the natural amino acids, including tyrosine. The *in vivo* concentration of tyrosine is much lower, as also indicated by the low micromolar Michaelis constant (K_m = 3.3 μM; Hamano-Takaku et al., 2000) of the *E. coli* tyrosyl-tRNA synthetase. Hence, these contradictory observations may be attributed to the high concentration of tyrosine in CFPS, which may favour tyrosine binding in the amino acid binding pocket of the G2 synthetase over the UAA binding. In contrast, however, the *p*-CNFRS mutant of *Mj* TyrRS maintains its fidelity both *in vivo* and in CFPS.

Regarding amber suppression by near-cognate tRNA, one would expect to observe glutamine incorporation more than tyrosine incorporation, as glutamyl-tRNA is the strongest competitor (Nilsson and Rydén-Aulin, 2003).

Conclusion

In our hands, use of purified G2 synthetase in an established CFPS system did not deliver the expected fidelity in unnatural amino acid incorporation.

References

Aerni, H. R., Shifman, M. A., Rogulina, S., O'Donoghue, P. and Rinehart, J. (2014) Revealing the amino acid composition of proteins within an expanded genetic code. *Nucleic Acids Res.* **43**, e8–e8.

Apponyi, M., Ozawa, K., Dixon, N. E. and Otting, G. (2008) Cell-free protein synthesis for analysis by NMR spectroscopy *Methods in Molecular Biology, Structural proteomics: high-throughput methods B. Kobe, M. Guss, T. Huber Eds., Humana Press, Totowa, USA* **426**, 257–268.

Cooley, R. B., Karplus, P. A. and Mehl, R. A. (2014) Gleaning unexpected fruits from hard-won synthetases: Probing principles of permissivity in non-canonical amino acid–tRNA synthetases. *ChemBioChem.* **15**, 1810–1819.

Gan, Q. and Fan, C. (2017) Increasing the fidelity of noncanonical amino acid incorporation in cell-free protein synthesis. *Biochim. Biophys. Acta* **1861**, 3047–3052.

Hamano-Takaku, F., Iwama, T., Saito-Yano, S., Takaku, K., Monden, Y., Kitabatake, M., Söll, D. and Nishimura, S. (2000) A mutant *Escherichia coli* tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine. *J. Biol. Chem.* **275**, 40324–40328.

Loh, C. T., Adams, L. A., Graham, B. and Otting, G. (2018) Genetically encoded amino acids with *tert*-butyl and trimethylsilyl groups for site-selective studies of proteins by NMR spectroscopy. *J. Biomol. NMR* **71**, 287–293.

Loscha, K. V., Herlt, A. J., Qi, R., Huber, T., Ozawa, K. and Otting, G. (2012) Multiple-site labeling of proteins with unnatural amino acids. *Angew. Chem. Int. Ed.* **51**, 2243–2246.

Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S. and Sakamoto, K. (2010) Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* **38**, 8188–8195.

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.

Nilsson, M. and Rydén-Aulin, M. (2003) Glutamine is incorporated at the nonsense codons UAG and UAA in a suppressor-free *Escherichia coli* strain. *Biochim. Biophys. Acta* **1627**, 1–6.

O'Donoghue, P., Prat, L., Heinemann, I. U., Ling, J., Odoi, K., Liu, W. R. and Söll, D. (2012) Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA^{Pyl} for genetic code expansion. *FEBS Lett.* **586**, 3931–3937.

Ozawa, K., Loscha, K. V., Kuppan, K. V., Loh, C. T., Dixon, N. E. and Otting, G. (2012) High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites. *Biochem. Biophys. Res. Commun.* **418**, 652–656.

Peeler, J. C., Woodman, B. F., Averick, S., Miyake-Stoner, S. J., Stokes, A. L., Hess, K. R., Matyjaszewski, K. and Mehl, R. A. (2010) Genetically encoded initiator for polymer growth from proteins. *J. Am. Chem. Soc.* **132**, 13575–13577.

Rauch, B. J., Porter, J. J., Mehl, R. A. and Perona, J. J. (2016) Improved incorporation of noncanonical amino acids by an engineered tRNA^{Tyr} suppressor. *Biochemistry* **55**, 618–628.

Speight, L. C., Muthusamy, A. K., Goldberg, J. M., Warner, J. B., Wissner, R. F., Willi, T. S., Woodman, B. F., Mehl, R. A. and Petersson, E. J. (2013) Efficient synthesis and *in vivo* incorporation of acridon-2-ylalanine, a fluorescent amino acid for lifetime and Förster resonance energy transfer/luminescence resonance energy transfer studies. *J. Am. Chem. Soc.* **135**, 18806−18814.

Sungwienwong, I., Ferrie, J. J., Jun, J. V., Liu, C., Barrett, T. M., Hostetler, Z. M., Ieda, N., Hendricks, A., Muthusamy, A. K., Kohli, R. M. and Chenoweth, D. M. (2018) Improving the fluorescent probe acridonylalanine through a combination of theory and experiment. *J. Phys. Org. Chem.* e3813.

Sungwienwong, I., Hostetler, Z. M., Blizzard, R. J., Porter, J. J., Driggers, C. M., Mbengi, L. Z., Villegas, J. A., Speight, L. C., Saven, J. G., Perona, J. J. and Kohli, R. M. (2017) Improving target amino acid selectivity in a permissive aminoacyl tRNA synthetase through counter-selection. *Org. Biomol. Chem.* **15**, 603–3610.

Wu, P. S. C., Ozawa, K., Lim, S. P., Vasudevan, S., Dixon, N. E. and Otting, G. (2007) Cell-free transcription/translation from PCR amplified DNA for high-throughput NMR studies. *Angew. Chem. Int. Ed.* **46**, 3356–3358.

Xuan, W., Li, J., Luo, X. and Schultz, P. G. (2016) Genetic incorporation of a reactive Isothiocyanate group into proteins. *Angew. Chem. Int. Ed.* **128**, 10219–10222.

Young, D. D., Young, T. S., Jahnz, M., Ahmad, I., Spraggon, G. and Schultz, P. G. (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894–1900.

Young, T. S., Ahmad, I., Yin, J. A. and Schultz, P. G. (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **395**, 361–374.