Enzymatic Spin-labeling of Protein N- and C-Termini for Electron Paramagnetic Resonance Spectroscopy

Robert Dunleavy[1], **Sidddarth Chandrasekaran**[1], **Brian R. Crane**[1,*]

[1]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA.

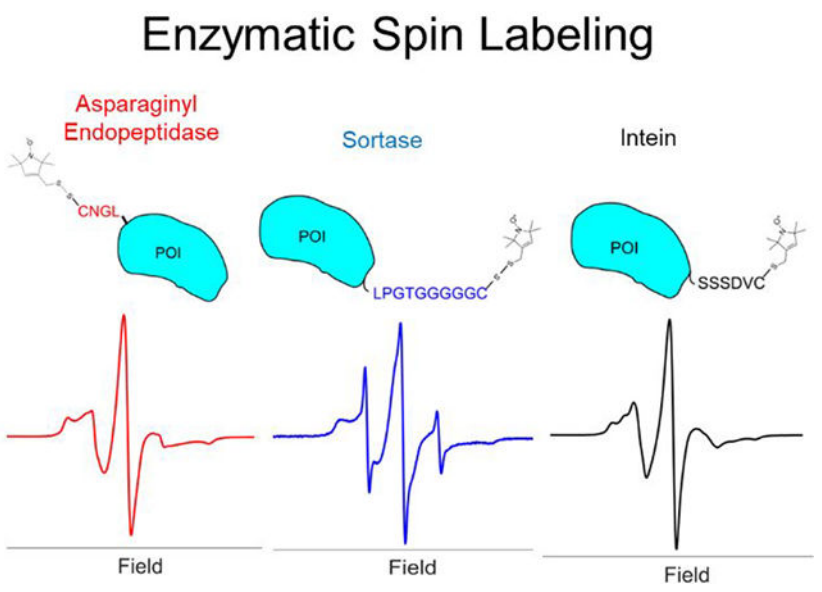
Abstract

Electron Paramagnetic Resonance (EPR) spectroscopy is a powerful tool for investigating the structure and dynamics of proteins. The introduction of paramagnetic moieties at specific positions in a protein enables precise measurement of local structure and dynamics. This technique, termed site-directed spin-labeling, has traditionally been performed using cysteine-reactive radical-containing probes. However, large proteins are more likely to contain multiple cysteine residues and cysteine labeling at specific sites may be infeasible or impede function. To address this concern, we applied three peptide-ligating enzymes (Sortase, Asparaginyl endopeptidase, and Inteins) for nitroxide labelling of N- and C-termini of select monomeric and dimeric proteins. Continuous wave and pulsed EPR (DEER) experiments reveal specific attachment of nitroxide probes to ether N-termini (OaAEP1) or C-termini (Sortase, Intein) across three test proteins (CheY, CheA, iLOV), thereby enabling a straightforward, highly specific and general method for protein labelling. Importantly, the linker length (3, 5, and 9 residues for OaAEP1, Intein, and Sortase reactions, respectively) between the probe and the target protein has a large impact on the utility of distance measurements by pulsed EPR, with longer linkers leading to broader distributions. As these methods are only dependent on accessible N- and C- termini, we anticipate application to a wide range of protein targets for biomolecular EPR spectroscopy.

Graphical Abstract

*For Correspondence: bc69@cornell.edu.

Competing interests: The authors declare that no competing interests exist.



Introduction

Electron Paramagnetic Resonance (EPR) spectroscopy is a valuable technique for investigating the structure and dynamics of biomolecules. EPR has been used to investigate a variety of biophysical systems ranging from characterization of protein complexes[1], conformational changes[2], and protein secondary structure[3]. EPR spectroscopy characterizes structure and dynamics of proteins difficult to study through other techniques such as NMR, X-ray crystallography, and cryo-electron microscopy; these advantages are particularly realized for studies of protein complexes and large intrinsically disordered proteins[4]. Furthermore, EPR spectroscopy can provide information complimentary to that obtained by these higher resolution techniques[5]. All EPR measurements rely upon the presence of an unpaired electron, typically an organic radical or a transition metal ion. The requirement of an unpaired election makes EPR spectroscopy particularly valuable as little to no background is observed from other non-paramagnetic species. However, many biomolecules lack paramagnetic cofactors and are therefore EPR silent. To overcome this challenge, several methods have been developed to introduce organic radicals or metal ions in site-specific locations in proteins – termed site-directed spin-labeling[6].

One of the most common spin-labeling strategies involves reacting MTSL (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl methanethiosulfonate spin label) with exposed cysteine residues to produce a nitroxide radical linked via a disulfide bond to the protein of interest (termed an R1 label[7]). Nitroxide radicals are commonly used spin probes in biomolecular EPR owing to their relative stability and because their spectra is sensitive to the motional regime commonly seen in biomolecules[3]. Although extensively used throughout the EPR community, MTSL labelling can be disadvantageous when dealing with proteins containing multiple cysteine sites, as mutation of these residues is laborious and can lead to loss of activity or function.

To circumvent this issue, several non-cysteine based spin-labeling strategies have been developed including labeling of tyrosine residues[8,9], the use of noncanonical amino acids (ncAAs)[10], and metal binding motifs[11]. The use of nCAAs allows site specific probe incorporation, either by ncAAs with a nitroxide side chain for *in vivo* labeling[12] or by ncAAs with side chains reactive to functionalized nitroxides for *in vitro* labelling[13]. As an alternative to nitroxide radicals, gadolinium and copper based probes have been used to react with select ncAAs[14,15], offering the potential for orthogonal labelling with nitroxides. A copper(II) binding ncAA has been investigated for *in vitro* DEER measurements among single proteins and protein complexes[15]. Gadolinium(III) in particular exhibits high stability for *in vivo* applications[16] and has been used for *in-cell* DEER measurements[17]. In contrast with ncAA approaches, copper(II) binding double histidine motifs along α-helices and β-sheets provide exceptionally narrow DEER distributions as compared to nitroxide R1 labelling[18]. Though the above techniques offer superior labeling, the use of expensive precursors and synthetic resources can be limiting for ncAA based approaches, especially for difficult to express proteins and non-bacterial expression systems. Copper and gadolinium based labeled could prove challenging in cases of non-specific binding either leading to unwanted DEER signal or poor stability and solubility in the presence of free metal ions. Additionally, the lack of structural information can hamper the applicability of double histidine motifs for novel and especially disordered targets. Moreover, any type of surface modification, but especially those altering ion coordination, may affect soluble protein expression levels.

To address these issues, we sought to develop non-cysteine based methods for N and C-terminal labeling of proteins, independent of secondary structure or protein molecular weight. Recently, several enzymes have been investigated for protein labeling applications, we sought to apply these methods for biomolecular EPR. We focused our attention on three well-studied enzymes that all catalyze peptide ligation reactions: Sortase A, Asparaginyl endopeptidase, and Inteins.

Sortase A is a cell surface trans-peptidase from *S. aureus* that conjugates proteins with a C-terminal LPXTG signal sequence to polyglycine (GGGGG) modified lipid II, thereby anchoring proteins to the cell surface[19]. Sortase A has been extensively used for N and C-terminal labeling[20,21], as well as labeling of internal loops[20]. Common probe attachments have included fluorophores[22], lipids[23], and PEG conjugates[24]. Many efforts have been undertaken to engineer Sortase A for labeling approaches, including those that produce faster ligation rates[25], alternate recognition sequences[26], and decreased calcium dependence[27]. We have previously investigated the use of Sortase A to track conformational changes of the C terminus of *Drosophila* cryptochrome using DEER measurements[28]. Here we apply Sortase A in comparison with Asparaginyl endopeptidases and Intein labeling.

Asparaginyl endopeptidases (AEPs) are a class of enzymes found in plants responsible for peptide cyclization and protein processing[29]. AEPs cleave a C-terminal asparagine residue and ligate an incoming peptide, which is variant specific[29]. Popular AEPs used for protein labeling applications include *C. ternatea* butelase-1 and *O. affinis* AEP1, for both N and C-terminal labelling,[30,31] although other AEPs have also been investigated[29]. Whereas butelase-1 displays high catalytic activity, its inability for recombinant production

requires multi-step purification from natural sources. However, *O. affinis* AEP1 (OaAEP1, 66% Identity to butelase-1) is readily expressed in *E. coli* with a single point mutation (C247A) that recovers 34% of the endopeptidase activity[32]. OaAPE1 preferentially reacts with a (N/GL) recognition sequence to allow ligation with a short three residue linker.

Inteins are a class of single turnover enzymes that catalyze the excision and ligation of two flanking protein sequences[33]. Inteins have been widely used in protein labeling[33], purification[34] and semi-synthesis[35]. Protein labelling applications have exclusively used trans-inteins[33], which rely upon the interaction of two domains, termed Intein-N and Intein-C, flanked by natural sequences termed exteins, Extein-N and Extein-C. These N and C-inteins associate with high affinity (Kd ~ nM)[36] to form a complex, where nucleophilic attack, transesterification, and acyl arrangement by the N and C-inteins results in the ligation of N and C exteins and excision of the parent N and C inteins[33]. Trans-Inteins have relatively conserved domains and are small enough to enable attachment to proteins of interest, N-inteins range from 100 to 130 residues while C-inteins range from 30 to 50 residues[37]. The sequence of the flanking N and C exteins (5–10 residues) greatly affects the kinetics of ligation, with native sequences exhibiting the fastest rates[38].

Herein we characterize the yields, labeling kinetics and resulting EPR behavior of three test proteins labeled with nitroxide-modified peptides by Sortase, OaAEP1, and an intein. The results provide a straightforward and effective approach for site-selective spin-labeling of proteins on their N- and C-termini.

Results and Discussion

For N-terminal labelling (Fig 1a), we utilized OaAEP1 for the attachment of a spin-labeled CNGL peptide (R1-CNGL). OaAEP1 requires an accessible N-terminal GL for conjugation; this site can be produced by preincubation of the protein of interest (POI) with TEV protease (ENLYFQ/GL), as TEV is tolerant to both glycine and serine at the P1’ position[39]. Incubation of R1-CNGL with GL-POI leads to the attachment of R1-CNGL to the POI’s N-terminus, with a 3 residue (NGL) linker. We expressed OaAEP1 (C247A) in *E. coli*; for ease of purification a TEV site was introduced to cleave the C terminal C Cap for enzyme activation, rather than C cap cleavage at acidic conditions as used previously[32].

For C-terminal labelling (Fig 1b), we employed either Sortase or a gp41-1 intein. As Sortase requires a poly-glycine donor peptide, we used R1-labelled GGGGC. Sortase catalyzes the attachment of GGGGC-R1 to a C-terminal LPGTG, leaving a conjugated product of POI-LPGTGGGGGC-R1, i.e. a 9-residue linker between the C-terminus and R1. Sortase typically requires calcium as a cofactor; however, we utilized a variant that has been engineered for activity without calcium (Sort7+)[27] to compensate for lower stability or solubility of some targets in the presence of calcium.

Additionally, we also applied the trans gp41-1 intein for C-terminal labelling due to its extremely fast kinetics, which are the fastest for any trans-intein known to date[40]. To shorten the final ligation product, but still achieve reasonable ligation rates, we omitted the native N-extein (TRSGY) while retaining the C-extein (SSSDV). Fusion of the gp41n intein (13

kDa) to a C-terminus of the POI and incubation of labelled gp41c-SSSDVC-R1, resulted in gp41n/c excision and a final ligation product of POI-SSSDVC-R1 with a 5-residue linker. The gp41c C-extein contains a catalytic Ser residue, as opposed to a Cys residue, and thus does not compete with labeling of the terminal Cys residue with MTSL.

We chose to utilize R1 labelled peptides for attachment as MTSL-based labelling methods are commonly used in biomolecular EPR and are a feasible option for laboratories without resources for solid phase peptide synthesis. For CNGL and GGGGC probes, the peptides were commercially synthesized and incubated with excess MTSL before purification by preparative Thin Layer Chromatography. Few side products were observed in the final products as seen in electrospray ionization (ESI) mass spectra (Fig 2a). As gp41c-SSSDVCG is a difficult target (62 residues) for solid phase peptide synthesis, this protein was expressed with a C-terminal SUMO tag in *E. coli*. Gp41c-SSSDVC was isolated after affinity purification of the SUMO fusion followed by the addition of TEV and thrombin proteases (Fig 2a, Fig S1), followed by labeling with MTSL. Room temperature X-band continuous wave (CW spectra) (Fig 2b) of the purified probes yielded spectra in the fast motion regime (τ_c < 1 ns). Correlation times were calculated with Easyspin[41] (Fig S2).

In order to estimate the time required for complete labelling (Fig 2c), product formation was monitored by either the loss of the gp41n intein on an SDS-PAGE gel or by attachment of a fluorescent fluorescein (FAM)-modified peptide for the OaAEP1 and Sortase reactions. For OaAEP1 conjugation of FAM-GNGL, nearly complete labelling was achieved after 4 hours with (1:100 molar eq) OaAEP1, which is similar to prior reports that achieve more than 90% labelling after 2 hours[42]. Sortase-mediated attachment of GGGGC-FAM was similar, with 5 hours for complete labeling with equimolar Sortase. Previous reports of Sortase-mediated conjugation have reported complete labelling on the order of hours[21]. We chose to have no additional linker residues present between the C-terminus of our POI and the gp41n intein, so as to minimize total linker length. However, this modification came at the expense of decreased labelling rates. Complete labelling was seen after 6 hours with 10-fold excess gp41c, considerably longer than that observed for the native N extein sequence which splices on the order of minutes[43]. Trans intein reactions are sensitive to the presence of extein sequences; deletion of the gp41-1c extein has been shown to produce an approximately 10-fold reduction in the labeling rate[43]. For all of the EPR samples, overnight incubation was used to maximize labelling efficiency.

Three spin-labeled proteins were investigated by both continuous wave and pulsed EPR spectroscopy: the chemotaxis signaling protein CheY (13 kDa) from *T. maritima*, the light activated fluorescent protein iLOV from *A. thaliana* (12 kDa), and the kinase module of the histidine kinase CheA from *T. maritima* (Residues 291–538, 56 kDa dimer). *T. maritima* CheY and CheA were chosen because they represent thermostable proteins of monomeric and dimeric oligomeric states, respectively. The fluorescent protein iLOV[44] was selected because of its ability to form a stable flavin mononucleotide (FMN) radical upon exposure to blue light, thereby providing a convenient second spin for pulsed distance measurements. Additionally, all three proteins have high-resolution (<2 Å) crystal structures available, which aided interpretation of the distance measurements between spin labels.

Overnight incubation of substoichiometric amounts of OaAEP1 (1:100 molar eq) and TEV (1:50) with excess R1-CNGL to CheY, iLOV and CheA, generated proteins that produced well defined X-band CW spectra (Fig 3a) with slow motional features, consistent with N-terminal attachment of R1-CNGL. Likewise, overnight incubation of Sortase (1:10) and excess GGGGC-R1 with CheY, iLOV and CheA led to similar spectra (Fig 3b), consistent with C-terminal attachment of GGGGC-R1. Lastly, overnight incubation of CheY, iLOV, and CheA gp41n fusions with excess (10:1) gp41c-SSSDVC-R1 yielded spectra (Fig 3c) consistent with attachment of SSSDVC-R1 to the available C-termini.

Overall, the X-band CW spectra of labeled proteins display features consistent with slow, fast, and intermediate motional regimes. The fast motional features in spectra such as CheA-LPGTGGGGGC-R1 are unlikely to be caused by free spin-labeled peptide because the ratio of the low to high field peak to peak amplitude (I^{+1}/I^{-1}) is much greater (2.02) in this spectrum than in the free GGGGC-R1 (1.22) spectrum. Additionally, the high field peak linewidth of this fast component (1.88 G) is broader than in the free GGGGC-R1 spectra (1.41 G). Therefore, we attribute the fast motional component of the R1-CNGL-iLOV and CheY/CheA-LPGTGGGGGC-R1 spectra to some fraction of fast tumbling spin labels attached to the protein. Due to the combination of backbone dynamics and mobility of the attached peptide linker, a distribution of spin label mobility is possible, especially if interactions are present between the spin label or peptide linker and the protein surface. This mobility distribution likely depends on the flexibility of the termini, the linker length/identity, and the local residues surrounding the labeling site.

After confirming labeling with CW EPR, Double Electron Electron Resonance (DEER) measurements were performed after introduction of second spins to evaluate spin-spin separations within the modified proteins. Both CheY and iLOV were labelled either N-terminally with OaAEP1 or C-terminally with gp41-1 intein. As CheY contains a native cysteine (Cys81), this residue was labelled with MTSL to serve as one spin, with either N- or C-terminal nitroxide moieties serving as the second spin. Illumination of iLOV with blue light (450 nm), produced an FMN neutral semiquinone (NSQ) radical, which was flash cooled in liquid N2 before measurement. Owing to the CheA homodimeric state, only C-terminal labelling was required to produce inter subunit spin separations.

N- and C-terminally labelled proteins were subject to four pulse DEER measurements performed at 60 K on a commercial Q-band spectrometer (Bruker E580). Background subtracted time domain data (Fig S3) produced similar distance distributions when analyzed with a Singular Value Decomposition based method[45]; or by Tikhonov regularization with DeerAnalysis[46] (Fig S4). Compared to labeling with OaAEP1 and intein, Sortase-labelled samples exhibited broad distance distributions and low modulation depths (Fig S5) most likely due to the long linker length (9 residues) of Sortase probe attachment. Expected distances (Table 1) were estimated from the deposited crystal structures with the inclusion of linker distance calculated using the Worm Like Chain Model[47]. The Worm Like Chain model, commonly used to describe flexible polymers, models short peptide sequences with a range of flexibility[48]. A persistence length of 3 Å, used in this study, is appropriate for flexible linkers, whereas higher persistence lengths (> 9 Å) more accurately model rigid linkers, such as those containing a high proline composition[49].

For CheY, nitroxide-nitroxide separations between Cys81 and NT or CT yielded relatively narrow distance distributions with expected distances within a few angstroms of calculated distances (Fig 4a, Table 1). For iLOV, flavin to N-terminal nitroxide DEER (Fig 4b) revealed a significantly narrower distance distribution (24 ± 0.5 Å) as compared to flavin to C-terminal DEER (22 ± 4 Å), thereby suggesting that the iLOV N-terminus is more constrained than the C-terminus. Measurements performed with longer N- and C-termini (7 and 5 additional residues, respectively) led to broad distance distributions (Fig S5), thereby suggesting that total linker length is an important variable to consider when aiming for well-defined DEER distances. For CheA, C-terminal nitroxide-to-nitroxide distances (Fig 4c) were slightly less than anticipated (35 ± 6 Å, expected 45 Å), perhaps because the nitroxide probes are interacting somewhat with the protein surface. We attribute the appearance of a second broad peak at 55 Å to higher-order associations of the protein, though interpretation of distances more than 4–5 nm at 3 μs evolution time is unreliable. Some such species may be disulfide linked because a small proportion dimerized subunits are seen on a nonreducing SDS-PAGE Gel (Fig S6). Overall, OaAEP1 and intein based N and C-terminal labelling yielded DEER distributions in reasonable agreement with calculated distances for selected test proteins.

Several effects may be responsible for the range of distance distributions seen across the three test proteins. In the case of extremely narrow distributions, as seen with R1-CNGL-iLOV, it is possible that interaction between the peptide linker and the protein surface confines the R1 label. Such a configuration would be difficult to predict beforehand, although calculation with the Worm Like Chain model can provide estimates of the distance. In the case of extremely broad distributions, as seen with LPGTGGGGGC-R1 labeled samples, it is likely that either this linker does not interact with the protein surface or that this linker length produces numerous conformations. It would be expected that non-interacting peptide linkers would lead to fast motional features seen in CW spectra. Most of the DEER spectra contain a broad component, perhaps indicative of non-interacting or fully extended peptide linkers. Overall, the use of shorter linkers (SSSDVC-R1 and R1-CNGL) is in most cases preferable, being more likely to produce useful DEER distributions.

Whereas we chose to utilize OaAEP1 for N-terminal labeling, OaAEP1 can also perform C-terminal labeling as previously shown for various targets[30]. Our attempts at using OaAEP1 for C-terminal labelling were met with poor (<10%) labeling efficiencies, likely due to the use of short (1–2 residues) linkers before the C-terminal asparagine and the protein of interest. Previous labelled targets have used longer linkers before the C-terminal asparagine, ranging from 12 (GFP)[30], 6 (ubiquitin)[30] and 4 residues (Nanobody)[42]. For DEER applications, shorter linker lengths are advantageous to limit the distance distribution width. We suggest using a linker as short as possible while still retaining reasonable labelling efficiency. Likewise, Inteins can also be used for N-terminal labelling. However, MTSL based labelling of N-intein conjugated probes would likely be problematic as the majority of N-terminal inteins have at least one catalytically active cysteine. Fortunately, a cysteine-less intein[50] has been investigated for both N and C-terminal labelling, exhibiting fast kinetics both *in vitro* and *in vivo*.

These N- and C-termini directed labeling methods provide some advantages over previously reported spin-labeling strategies, namely ease of reagent preparation and applicability to a wide range of proteins. Termini labeling may be less likely to perturb protein function and stability as compared to other site directed methods because internal labeling has a greater potential to perturb residues critical for activity and solubility. In theory, these methods allow labeling of any protein with accessible N- and C-termini. The enzymatic methods here are easily performed, as both Sortase A and OaAEP1 are readily expressed and purified from *E. coli* and donor peptides are readily labelled with MTSL. Their high degree of specificity also enables orthogonal labeling with different probes, i.e. N-terminal labelling with OaAEP1 and C-terminal labelling through either Sortase or intein. Such an approach could be advantageous in investigating N or C-terminal domain movements, for example in response to added ligand or the addition of binding partners. Additionally, N and C-terminal labeling can also be performed with gadolinium(III) or copper(II) chelating peptides, allowing distances to be measured between transitional metals and nitroxides. In combination with either ncAAs or traditional cysteine mutants, this would enable three-way labelling, a particular advantage in DEER experiments. While we opted to utilize R1 labelled peptides, more rigid spin labels such as TOAC[51] and TOPP[52] could also be conjugated to donor peptides, offering narrower DEER distributions.

Some disadvantages of the labeling strategies presented here include limited site flexibility and the influence of total linker length. The ability to only label N- and C-termini is restrictive compared to the residue specific labelling of cysteine and ncAA based approaches. Additionally, linker lengths have a large influence on the usefulness of acquired DEER spectra, we have experienced more than 6 residues often leads to broad distance distributions. The linkers presented here are significantly longer than R1 and ncAA based approaches, which may limit applicability in cases where accurate DEER distances are required. Depending on the application, spin labels more rigid than R1 can be utilized, such as the bivalent nitroxide RX[53] and copper binding double histidine motifs[11]. The termini labeling methods are similar to the labeling of surface exposed loops, which can vary in terms of local flexibility. In prior investigation using Sortase A for C terminal labelling of a light sensing cryptochrome[28], we found narrow distance distributions (2–8 Å HWHM) and crystallographic evidence of a partially ordered linker. Therefore the 9 residue linker left after Sortase labeling may adopt ordered conformations in a protein-specific manner. We have experienced decreased modulation depths and broader distance distributions with increasing linker length. Although such samples yield less reliable distance distributions, information can be obtained on the presence of shorter (< 60 Å) separations, thereby enabling further study after linker optimization. Cases with flexible or disordered N- and C-termini can be improved through the use of select termini truncations, guided by crystal structures or structure predictions. Another potential issue may be that certain proteins may not tolerate the C-terminal addition of the 13 kDa gp41N intein fragment, resulting in misfolding and poor stability.

The enzymatic methods presented here provide unique advantages and disadvantages depending on the system of study. For example, these methods fair well when working with systems with multiple cysteines such as large proteins or complexes, where ncAA approaches are not feasible. A particular advantage for protein complexes may be realized

because cysteine null variants may affect complex formation. However, the lack of residue specificity can be limiting in cases where site accuracy is required, such as the targeting of specific domain motions. In these cases, SDSL with MTSL or ncAAs may be preferable. In certain cases, where termini contain the region of interest either an enzymatic method or MTSL labeling could be equally effective. We anticipate the enzymatic methods presented here, either used alone or in combination with other SDSL methods, will enable a greater number of targets to be studied using continuous wave and pulsed EPR spectroscopy.

Conclusions

In conclusion, we utilized Sortase, OaAEP1, and inteins to attach spin-labelled peptides to the N and C termini of select proteins. Overnight incubation of proteins of interest with spin-labeled donor peptides and associated enzymes led to well defined slow motional CW features, consistent with probe attachment. Distance distributions of N and C-terminally labelled proteins agreed with expected distances across a diverse set of test proteins and spin probes. We anticipate these labelling methods to be a useful addition to biomolecular EPR spectroscopy, as they enable a general method for N and C-terminal spin-labeling of in principle any protein that can be produced in sufficient quantities.

Experimental Methods

Purification of Labelled Peptides

Peptides (GGGGC and CNGL) were purchased at >95% purity from Biomatik (Cambridge, Ontario, Canada) and MTSL (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate) was purchased from SantaCruz Biotechnology. Peptides were dissolved in a 50:50 mixture of acetonitrile (AcN) and water containing 50 mM HEPES (pH 8). To this mixture, a slight molar excess of MTSL was added and left to react overnight at room temperature. The reaction was checked by thin-layer chromatography on silica with the mobile phase containing 3:1:1 n-butanol:acetic acid: water. Products were separated by preparative TLC using the same solvent system; the silica band containing the desired product was cut out from the plate and dissolved in a small amount of 50:50 ACN/Water. The product was first spun down to remove any silica/contaminants and then checked for purity by LC–MS. This supernatant was transferred into a round bottom flask and rotovapped. The dry (final) product was dissolved in a small amount of solvent (50 mM Tris, 150 mM NaCl, 5% glycerol for GGGGC, 50% CAN for CNGL) and stored at −80 °C. The same procedure was followed for (GGGGC-FAM), except fluorescein maleimide (Cayman Chemicals) was used in place of MTSL. For OaAEP1 labelling, FAM-GNGL was purchased from (Biomatik, 96% Purity) and used directly.

Purification of gp41c-SSSDVC-R1

The following sequence, containing a N terminal His tag and a C-terminal SUMO tag was cloned into pET28a.

MGSSHHHHHHGENLYFQGLKKILKIEELDERELIDIEVSGNHLFYANDILTHNSSSDV
CGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL

MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGYPYDVPD YA*

gp41c-SSDVCG-SUMO in pET28a was transformed into BL21 (DE3) (New England Biolabs) and purified using standard protein expression and purification protocols. Briefly, a single colony was grown overnight in LB at 37 °C, diluted into LB (10 mL in 1 L), and grown until mid-exponential phase (OD 0.5–0.8). Expression was induced with 0.2 mM IPTG and cells were allowed to grow overnight at room temperature. Harvested cells were lysed in lysis buffer (20 mM Tris (pH 8), 500 mM NaCl, 5% (v/v) glycerol) and subjected to immobilized metal affinity chromatography (IMAC) with a gradient of 10, 25, and 500 mM Imidazole. The 500 mM elution containing gp41c-SUMO was diluted to 1 mg/mL with (20 mM Tris (pH 8), 150 mM NaCl, 1 mM CaCl2 and 5% glycerol) and incubated overnight at 4C with thrombin. The next day, both beta-mercaptoethanol (1 mM) and TEV (1:50 mg) was added (1 mM) and cleavage of the N terminal His tag was allowed to proceed overnight.

Cleaved gp41c/SUMO was concentrated (1–2 mM) and incubated with excess BME (20 mM) before desalting and incubation with MTSL (in 2–3x concentration excess) overnight at room temperature. Unreacted MTSL was removed by desalting and gp41c was concentrated to 5 mM, aliquoted and stored at −20 °C. The final storage buffer contains 20 mM Tris (pH 8), 150 mM NaCl, 5% glycerol.

As SUMO contains no cysteines, coelutes on SEC with gp41c and has a similar pI (4.6) to gp41c (4.9), we opted not to pursue further chromatography for higher purify gp41c.

Purification of Sortase 7+

Staphylococcus aureus Sortase A 7+ in pET30b (Addgene 105602) was transformed into BL21 (DE3) (New England Biolabs) and purified using standard protein expression and purification protocols. Briefly, a single colony was grown overnight in LB at 37 °C, diluted (10 mL in 1 L) into LB, and grown until mid-exponential phase (OD 0.5–0.8). Expression was induced with 0.2 mM IPTG and cells were allowed to grow overnight at room temperature. Harvested cells were lysed in lysis buffer (20 mM Tris (pH 8), 500 mM NaCl, 5% (v/v) glycerol) and subjected to IMAC with a gradient of 10, 25, and 500 mM Imidazole. The 500 mM elution was then further purified using a preparative Superdex 200 size exclusion column (26/600, GE Healthcare) in 25 mM Tris (pH 8), 150 mM NaCl, and 10% glycerol (vol/vol). Fractions containing sortase were pooled, concentrated to 40 mg/mL and stored at −20 C after the addition of glycerol to 50%.

Purification of OaAEP1 (C243A)

OaAEP1 (Addgene 89482) was modified in several ways before expression:

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSARDGDYLHLPSEVSRFFRPQE TNDDHGEDSVGTRWAVLIAGSKGYANYRHQAGVCHAYQILKRGGLKDENIVVFMY DDIAYNESNPRPGVIINSPHGSDVYAGVPKDYTGEEVNAKNFLAAILGNKSAITGGSG KVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLYADELNDALKKKHASGTYKSLVFY LEACESGSMFEGILPEDLNIYALTSTNTTESSWAYYCPAQENPPPPEYNVCLGDLFSVA WLEDSDVQNSWYETLNQQYHHVDKRISHASHATQYGNLKLGEEGLFVYMGSNPE

NLYFQGANDNYTSLDGNALTPSSIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQ IFKAMSHRVHIDSSIKLIGKLLFGIEKCTEILNAVRPAGQPLVDDWACLRSLVGTFETH CGSLSEYGMRHTRTIA NICNAGISEEQMAEAASQACASIPLE

1. The C243A mutation was introduced via Q5 Site Directed Mutagenesis
2. The N terminal SUMO tag was deleted as a significant portion of free SUMO was present in initial purifications, suggesting cleavage between SUMO and the OaAEP1 N terminus.
3. A TEV site was introduced before the C-terminal cap because removal of the cap is necessary for activity and (1) cleavage of the cap at low pH (4) led to significant precipitation creating low yields of OaAEP1 and (2) expression without the C cap was unsuccessful.

OaAEP1 (C243A) was transformed into BL21 (DE3) (New England Biolabs) and purified using standard protein expression and purification protocols. Briefly, a single colony was grown overnight in LB at 37 °C, diluted into LB (10 mL in 1L), and grown until mid-exponential phase (OD 0.5–0.8). Expression was induced with 0.2 mM IPTG and cells were allowed to grow overnight at room temperature. Harvested cells were lysed in lysis buffer (20 mM Tris (pH 8), 500 mM NaCl, 5% (v/v) glycerol) and subjected to IMAC with a gradient of 10, 25, and 500 mM Imidazole. The 500 mM elution was desalted into 20 mM Tris (pH 8), 150 mM NaCl, 5% glycerol, and 1 mM DTT. An engineered form of TEV protease (Addgene 92414) was added at 1:100 mg. After confirming cleavage of the C cap with SDS-PAGE, the reaction mixture was desalted to remove DTT, concentrated, aliquoted and stored at −80 C.

Purification of Labelling Targets

Standard expression protocols in BL21 (DE3) were performed followed by IMAC. No issues arose for purification of CheY, iLOV, and CheA. For each protein, the IMAC elution was desalted into 25 mM Tris (pH 8), 150 mM NaCl, and 5% glycerol (v/v). Proteins were either stored at −20 °C for long-term storage or at 4 °C for immediate (1–2 day) use.

Labelling Reactions for CW and Pulsed EPR

Each protein target was diluted (50 – 100 uM) into 25 mM Tris (pH 8), 150 mM NaCl, and 5% glycerol (v/v) with a reaction volume of 5 mL. Indicated amounts of either OaAEP1 (1:100 molar eq), Sortase (1:1 molar eq), and gp41c (10:1 molar eq) were added. For OaAEP1 and Sortase reaction, 2–4 molar excess of peptide was added. All reactions were incubated overnight at RT before being concentrated and injected on an S200 (10/300, Cytiva) SEC column. Fractions were concentrated to >100 uM for analysis by CW or pulsed EPR. CW samples were concentrated in 25 mM Tris (pH 8), 150 mM NaCl, and 5% glycerol (v/v), whereas DEER samples were buffer exchanged into 20 mM Tris, 150 mM NaCl, 25% d8-glycerol prepared in D_2O.

CW and Pulsed EPR Spectroscopy

CW EPR spectroscopy experiments were carried out at room temperature at X-band (~9.4 GHz) with a modulation amplitude of 2 G on a Bruker E500 spectrometer equipped

with a super Hi-Q resonator. All pulse-EPR measurements were carried out at Q-band (~35 GHz) on a Bruker E580 spectrometer equipped with a 10 W solid-state amplifier (150 W equivalent TWTA) and an arbitrary waveform generator. DEER measurements were performed at 60 K in an EN 5107D2 cavity with a cryogen-free insert/temperature controller. DEER was carried out using four pulses (π/2-τ1-π-τ1-πpump-τ2-π-τ2-echo) with 16-step phase cycling. The pump and probe pulses were separated by 56 MHz (~20 G) for nitroxide-nitroxide distances. For flavin-nitroxide distances, the pump (flavin) and probe (nitroxide) pulses were separated by 84 MHz (~30 G). A typical π/2 pulse length was between 16 and 20 ns.

EPR Data Analysis

For CW fitting, the garlic.m function in easyspin (5.2.35) was used in MATLAB (2020a). Values for A, g, lw, and tau were varied using the Nelder/Mead simplex and Levenberg-Marquardt search algorithms. For DEER data processing, both a SVD based method developed at Cornell (https://denoising.cornell.edu/) and Tikhonov regularization using DEERAnalysis (Version 2022, https://epr.ethz.ch/software.html) developed at ETH Zurich were used.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Curt Dunam of the National Resource for Advanced Electron-Spin Resonance Spectroscopy (ACERT, Cornell University) for assistance with spectrometer maintenance and repair. This research was supported by NIH grants R01GM066775 (B.R.C), S10OD021543 (ACERT), P41GM103521 (ACERT) and Molecular Biophysics Training Grant T32GM826730 (R.D).

References

(1). Jeschke G DEER Distance Measurements on Proteins. Annu. Rev. Phys. Chem. 2012, 63, 419–446. 10.1146/annurev-physchem-032511-143716. [PubMed: 22404592]

(2). McHaourab HS; Steed PR; Kazmier K Toward the Fourth Dimension of Membrane Protein Structure: Insight into Dynamics from Spin-Labeling EPR Spectroscopy. Structure 2011, 19 (11), 1549–1561. 10.1016/j.str.2011.10.009. [PubMed: 22078555]

(3). Torricella F; Pierro A; Mileo E; Belle V; Bonucci A Nitroxide Spin Labels and EPR Spectroscopy: A Powerful Association for Protein Dynamics Studies. Biochim. Biophys. Acta - Proteins Proteomics 2021, 1869 (7), 140653. 10.1016/j.bbapap.2021.140653. [PubMed: 33757896]

(4). Drescher M EPR in Protein Science: Intrinsically Disordered Proteins. Top. Curr. Chem. 2012, 321, 91–120. 10.1007/128. [PubMed: 21826602]

(5). Jeschke G The Contribution of Modern EPR to Structural Biology. Emerg. Top. Life Sci. 2018, 2 (1), 9–18. 10.1042/ETLS20170143. [PubMed: 33525779]

(6). Berliner L; Reuben J Spin Labeling: Theory and Applications; Plenum Press: New York, 1989.

(7). Hubbell WL; Gross A; Langen R; Lietzow MA Recent Advances in Site-Directed Spin Labeling of Proteins. Curr. Opin. Struct. Biol. 1998, 8, 649–656. [PubMed: 9818271]

(8). Lorenzi M; Puppo C; Lebrun R; Lignon S; Roubaud V; Martinho M; Mileo E; Tordo P; Marque SRA; Gontero B; Guigliarelli B; Belle V Tyrosine-Targeted Spin Labeling and EPR Spectroscopy: An Alternative Strategy for Studying Structural Transitions in Proteins. Angew. Chemie - Int. Ed. 2011, 50 (39), 9108–9111. 10.1002/anie.201102539.

(9). Gmeiner C; Klose D; Mileo E; Belle V; Marque SRA; Dorn G; Allain FHT; Guigliarelli B; Jeschke G; Yulikov M Orthogonal Tyrosine and Cysteine Site-Directed Spin Labeling for Dipolar Pulse EPR Spectroscopy on Proteins. J. Phys. Chem. Lett 2017, 8 (19), 4852–4857. 10.1021/acs.jpclett.7b02220. [PubMed: 28933855]

(10). Fleissner MR; Brustad EM; Kálai T; Altenbach C; Cascio D; Peters FB; Hideg K; Schultz PG; Hubbell WL Site-Directed Spin Labeling of a Genetically Encoded Unnatural Amino Acid. Proc. Natl. Acad. Sci. U. S. A. 2009, 106 (51), 21637–21642. 10.1073/pnas.0912009106. [PubMed: 19995976]

(11). Cunningham TF; Putterman MR; Desai A; Horne WS; Saxena S The Double-Histidine Cu2+-Binding Motif: A Highly Rigid, Site-Specific Spin Probe for Electron Spin Resonance Distance Measurements. Angew. Chemie - Int. Ed. 2015, 54 (21), 6330–6334. 10.1002/anie.201501968.

(12). Schmidt MJ; Borbas J; Drescher M; Summerer D A Genetically Encoded Spin Label for Electron Paramagnetic Resonance Distance Measurements. J. Am. Chem. Soc. 2014, 136 (4), 1238–1241. 10.1021/ja411535q. [PubMed: 24428347]

(13). Kucher S; Korneev S; Tyagi S; Apfelbaum R; Grohmann D; Lemke EA; Klare JP; Steinhoff HJ; Klose D Orthogonal Spin Labeling Using Click Chemistry for in Vitro and in Vivo Applications. J. Magn. Reson. 2017, 275, 38–45. 10.1016/j.jmr.2016.12.001. [PubMed: 27992783]

(14). Abdelkader EH; Feintuch A; Yao X; Adams LA; Aurelio L; Graham B; Goldfarb D; Otting G Protein Conformation by EPR Spectroscopy Using Gadolinium Tags Clicked to Genetically Encoded P-Azido-l-Phenylalanine. Chem. Commun. 2015, 51 (88), 15898–15901. 10.1039/c5cc07121f.

(15). Merz GE; Borbat PP; Muok AR; Srivastava M; Bunck DN; Freed JH; Crane BR Site-Specific Incorporation of a Cu2+ Spin Label into Proteins for Measuring Distances by Pulsed Dipolar Electron Spin Resonance Spectroscopy. J. Phys. Chem. B 2018, 122 (41), 9443–9451. 10.1021/acs.jpcb.8b05619. [PubMed: 30222354]

(16). Giannoulis A; Ben-Ishay Y; Goldfarb D Characteristics of Gd(III) Spin Labels for the Study of Protein Conformations, 1st ed.; Elsevier Inc., 2021; Vol. 651. 10.1016/bs.mie.2021.01.040.

(17). Martorana A; Bellapadrona G; Feintuch A; Gregorio E Di; Aime, S.; Goldfarb, D. Probing Protein Conformation in Cells by EPR Distance Measurements Using Gd3+ Spin Labeling. J. Am. Chem. Soc. 2014, 136 (38), 13458–13465. 10.1021/ja5079392. [PubMed: 25163412]

(18). Gamble Jarvi A; Bogetti X; Singewald K; Ghosh S; Saxena S Going the DHis-Tance: Site-Directed Cu2+Labeling of Proteins and Nucleic Acids. Acc. Chem. Res. 2021. 10.1021/acs.accounts.0c00761.

(19). Pishesha N; Ingram JR; Ploegh HL Sortase A: A Model for Transpeptidation and Its Biological Applications. Annu. Rev. Cell Dev. Biol. 2018, 34, 163–188. 10.1146/annurev-cellbio-100617-062527. [PubMed: 30110557]

(20). Guimaraes CP; Witte MD; Theile CS; Bozkurt G; Kundrat L; Blom AEM; Ploegh HL Site-Specific C-Terminal and Internal Loop Labeling of Proteins Using Sortase-Mediated Reactions. Nat. Protoc. 2013, 8 (9), 1787–1799. 10.1038/nprot.2013.101. [PubMed: 23989673]

(21). Theile CS; Witte MD; Blom AEM; Kundrat L; Ploegh HL; Guimaraes CP Site-Specific N-Terminal Labeling of Proteins Using Sortase-Mediated Reactions. Nat. Protoc. 2013, 8 (9), 1800–1807. 10.1038/nprot.2013.102. [PubMed: 23989674]

(22). Antos JM; Ingram J; Fang T; Pishesha N; Truttmann MC; Ploegh HL Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. Curr. Protoc. protein Sci. 2017, 89, 15.3.1–15.3.19. 10.1002/cpps.38.

(23). Antos JM; Miller GM; Grotenbreg GM; Ploegh HL Lipid Modification of Proteins through Sortase-Catalyzed Transpeptidation. J. Am. Chem. Soc. 2008, 130 (48), 16338–16343. 10.1021/ja806779e. [PubMed: 18989959]

(24). Chen L; Cohen J; Song X; Zhao A; Ye Z; Feulner CJ; Doonan P; Somers W; Lin L; Chen PR Improved Variants of SrtA for Site-Specific Conjugation on Antibodies and Proteins with High Efficiency. Sci. Rep. 2016, 6, 1–12. 10.1038/srep31899. [PubMed: 28442746]

(25). Li J; Zhang Y; Soubias O; Khago D; Chao FA; Li Y; Shaw K; Byrd RA Optimization of Sortase A Ligation for Flexible Engineering of Complex Protein Systems. J. Biol. Chem. 2020, 295 (9), 2664–2675. 10.1074/jbc.RA119.012039. [PubMed: 31974162]

(26). Dorr BM; Ham HO; An C; Chaikof EL; Liu DR Reprogramming the Specificity of Sortase Enzymes. Proc. Natl. Acad. Sci. U. S. A. 2014, 111 (37), 13343–13348. 10.1073/pnas.1411179111. [PubMed: 25187567]

(27). Jeong HJ; Abhiraman GC; Story CM; Ingram JR; Dougan SK Generation of Ca2+-Independent Sortase A Mutants with Enhanced Activity for Protein and Cell Surface Labeling. PLoS One 2017, 12 (12), 1–15. 10.1371/journal.pone.0189068.

(28). Chandrasekaran S; Schneps CM; Dunleavy R; Lin C; DeOliveira CC; Ganguly A; Crane BR Tuning Flavin Environment to Detect and Control Light-Induced Conformational Switching in Drosophila Cryptochrome. Commun. Biol. 2021, 4 (1). 10.1038/s42003-021-01766-2.

(29). Tang TMS; Luk LYP Asparaginyl Endopeptidases: Enzymology, Applications and Limitations. Org. Biomol. Chem. 2021, 19 (23), 5048–5062. 10.1039/d1ob00608h. [PubMed: 34037066]

(30). Tang TMS; Cardella D; Lander AJ; Li X; Escudero JS; Tsai YH; Luk LYP Use of an Asparaginyl Endopeptidase for Chemo-Enzymatic Peptide and Protein Labeling. Chem. Sci. 2020, 11 (23), 5881–5888. 10.1039/d0sc02023k. [PubMed: 32874509]

(31). Harris KS; Durek T; Kaas Q; Poth AG; Gilding EK; Conlan BF; Saska I; Daly NL; Van Der Weerden NL; Craik DJ; Anderson MA Efficient Backbone Cyclization of Linear Peptides by a Recombinant Asparaginyl Endopeptidase. Nat. Commun. 2015, 6. 10.1038/ncomms10199.

(32). Yang R; Wong YH; Nguyen GKT; Tam JP; Lescar J; Wu B Engineering a Catalytically Efficient Recombinant Protein Ligase. J. Am. Chem. Soc. 2017, 139 (15), 5351–5358. 10.1021/jacs.6b12637. [PubMed: 28199119]

(33). Shah NH; Muir TW Inteins: Nature's Gift to Protein Chemists. Chem. Sci. 2014, 5 (2), 446–461. 10.1039/c3sc52951g. [PubMed: 24634716]

(34). Wood DW; Camarero JA Intein Applications: From Protein Purification and Labeling to Metabolic Control Methods. J. Biol. Chem. 2014, 289 (21), 14512–14519. 10.1074/jbc.R114.552653. [PubMed: 24700459]

(35). Conibear AC; Watson EE; Payne RJ; Becker CFW Native Chemical Ligation in Protein Synthesis and Semi-Synthesis. Chem. Soc. Rev. 2018, 47 (24), 9046–9068. 10.1039/c8cs00573g. [PubMed: 30418441]

(36). Braner M; Kollmannsperger A; Wieneke R; Tampé R "Traceless" Tracing of Proteins – High-Affinity Trans-Splicing Directed by a Minimal Interaction Pair. Chem. Sci. 2016, 7 (4), 2646–2652. 10.1039/c5sc02936h. [PubMed: 28660037]

(37). Saleh L; Perler FB Protein Splicing in Cis and in Trans. Chem. Rec. 2006, 6 (4), 183–193. 10.1002/tcr.20082. [PubMed: 16900466]

(38). Shah NH; Eryilmaz E; Cowburn D; Muir TW Extein Residues Play an Intimate Role in the Rate-Limiting Step of Protein Trans -Splicing. J. Am. Chem. Soc. 2013, 135 (15), 5839–5847. 10.1021/ja401015p. [PubMed: 23506399]

(39). Kapust RB; Toözseór J; Copeland TD; Waugh DS The P1′ Specificity of Tobacco Etch Virus Protease. Biochem. Biophys. Res. Commun. 2002, 294 (5), 949–955. 10.1016/S0006-291X(02)00574-0. [PubMed: 12074568]

(40). Carvajal-vallejos P; Mootz HD; Schmidt SR Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources *. 2012, 287 (34), 28686–28696. 10.1074/jbc.M112.372680.

(41). Stoll S; Schweiger A EasySpin, a Comprehensive Software Package for Spectral Simulation and Analysis in EPR. J. Magn. Reson. 2006, 178 (1), 42–55. 10.1016/j.jmr.2005.08.013. [PubMed: 16188474]

(42). Rehm FBH; Harmand TJ; Yap K; Durek T; Craik DJ; Ploegh HL Site-Specific Sequential Protein Labeling Catalyzed by a Single Recombinant Ligase. J. Am. Chem. Soc. 2019, 141 (43), 17388–17393. 10.1021/jacs.9b09166. [PubMed: 31573802]

(43). Carvajal-Vallejos P; Palliss?? R; Mootz HD; Schmidt SR Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources. J. Biol. Chem. 2012, 287 (34), 28686–28696. 10.1074/jbc.M112.372680. [PubMed: 22753413]

(44). Kopka B; Magerl K; Savitsky A; Davari MD; Röllen K; Bocola M; Dick B; Schwaneberg U; Jaeger KE; Krauss U Electron Transfer Pathways in a Light, Oxygen, Voltage (LOV) Protein

Devoid of the Photoactive Cysteine. Sci. Rep. 2017, 7 (1), 1–16. 10.1038/s41598-017-13420-1. [PubMed: 28127051]

(45). Srivastava M; Freed JH Singular Value Decomposition Method to Determine Distance Distributions in Pulsed Dipolar Electron Spin Resonance. J. Phys. Chem. Lett. 2017, 8 (22), 5648–5655. 10.1021/acs.jpclett.7b02379. [PubMed: 29099190]

(46). Ibáñez LF; Jeschke G; Stoll S DeerLab: A Comprehensive Software Package for Analyzing Dipolar Electron Paramagnetic Resonance Spectroscopy Data. Magn. Reson. 2020, 1 (2), 209–224. 10.5194/mr-1-209-2020.

(47). Rubinstein M; Colby RH Polymer Physics; Oxford University Press, 2003.

(48). Zhou HX Polymer Models of Protein Stability, Folding, and Interactions. Biochemistry 2004, 43 (8), 2141–2154. 10.1021/bi036269n. [PubMed: 14979710]

(49). Kjaergaard M; Glavina J; Chemes LB Predicting the Effect of Disordered Linkers on Effective Concentrations and Avidity with the “Ceff Calculator” App, 1st ed.; Elsevier Inc., 2021; Vol. 647. 10.1016/bs.mie.2020.09.012.

(50). Bhagawati M; Terhorst TME; Füsser F; Hoffmann S; Pasch T; Pietrokovski S; Mootz HD A Mesophilic Cysteine-Less Split Intein for Protein Trans-Splicing Applications under Oxidizing Conditions. Proc. Natl. Acad. Sci. U. S. A. 2019, 116 (44), 22164–22172. 10.1073/pnas.1909825116. [PubMed: 31611397]

(51). Schreier S; Bozelli JC; Marín N; Vieira RFF; Nakaie CR The Spin Label Amino Acid TOAC and Its Uses in Studies of Peptides: Chemical, Physicochemical, Spectroscopic, and Conformational Aspects. Biophys. Rev. 2012, 4 (1), 45–66. 10.1007/s12551-011-0064-5. [PubMed: 22347893]

(52). Stoller S; Sicoli G; Baranova TY; Bennati M; Diederichsen U TOPP: A Novel Nitroxide-Labeled Amino Acid for EPR Distance Measurements. Angew. Chemie - Int. Ed. 2011, 50 (41), 9743–9746. 10.1002/anie.201103315.

(53). Fleissner MR; Bridges MD; Brooks EK; Cascio D; Kálai T; Hideg K; Hubbell WL Structure and Dynamics of a Conformationally Constrained Nitroxide Side Chain and Applications in EPR Spectroscopy. Proc. Natl. Acad. Sci. U. S. A. 2011, 108 (39), 16241–16246. 10.1073/pnas.1111420108. [PubMed: 21911399]

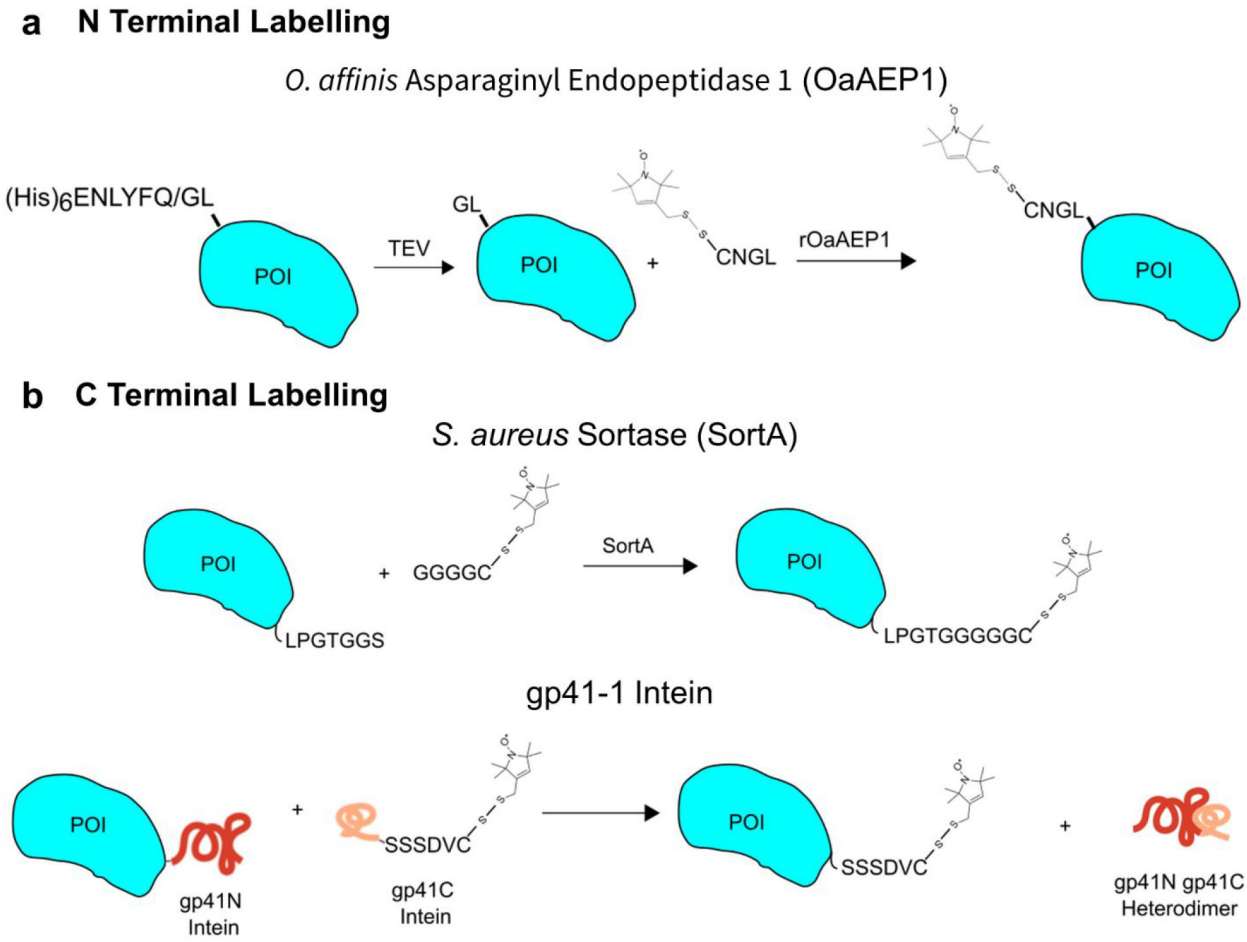


Figure 1: N and C-terminal Attachment of Spin-labelled Peptides

a) Labelling scheme for attachment of an R1-CNGL peptide to the N-terminus of a protein of interest (POI) using an engineered Asparaginyl Endopeptidase 1 (OaAEP1). b) C-terminal labelling strategy using either a GGGGC-R1 peptide with Sortase or attachment of a SSSDVC-R1 peptide with a split gp41-1 intein. Donor peptides were labeled with MTSL which produces an R1 label linked via a disulfide bond.

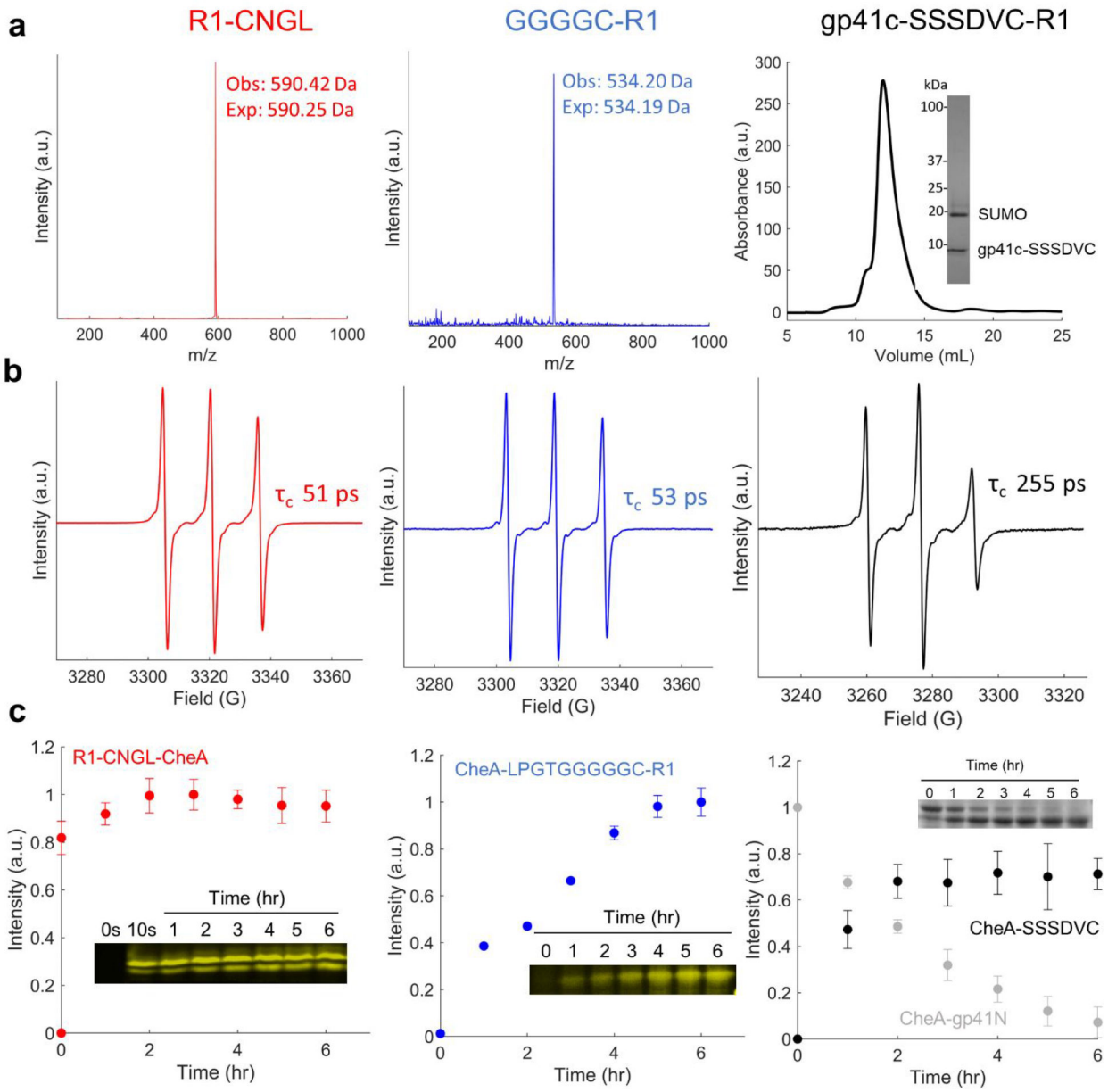


Figure 2. Characterization and Attachment of Labelled Peptides

a) Purity of R1 labelled probes: R1-CNGL for attachment with OaAEP1 (ESI-MS), GGGGC-R1 for attachment with Sortase (ESI-MS), and R1 labelled gp41c intein (SEC, SDS-PAGE). b) X-band Continuous Wave (CW) EPR spectra for spin-labelled probes. Simulations and fitting parameters for calculation of the correlation time (τ_c) are provided in Fig S2. c) Rates of probe attachment using fluorescein labelled CNGL and GGGGC. Kinetics of the split gp41 intein reaction were monitored by the loss of the gp41N-intein fragment (12 kDa).

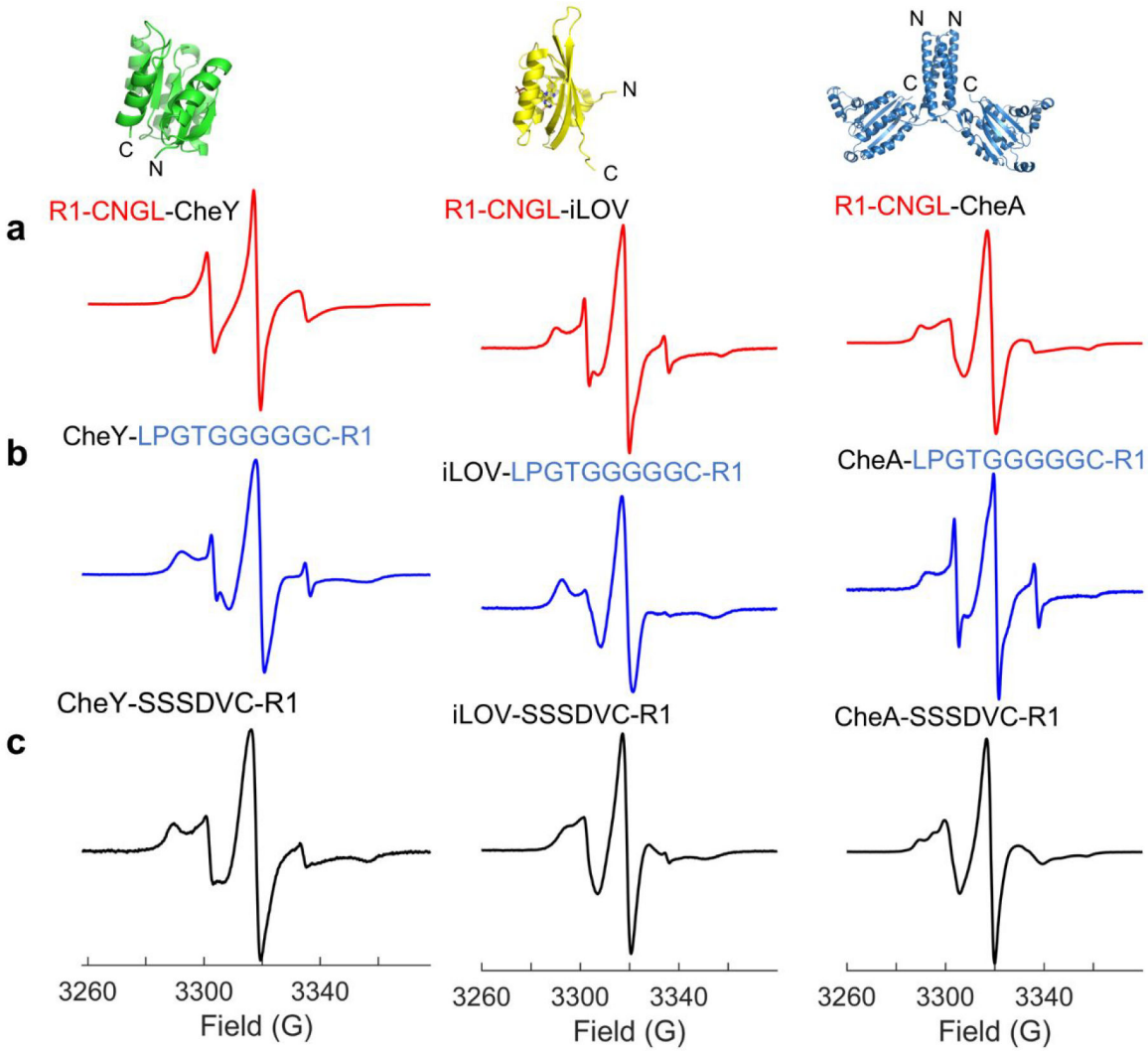


Figure 3. X band CW EPR Spectra of Labelled Proteins.

CW X-band EPR spectra of attached (a) R1-CNGL (b) GGGGC-R1 and (c) SSSDVC-R1 for CheY, iLOV, and CheA. N-terminal (N) and C-terminal (C) labeling positions are indicated for CheY (PDB:1TMY), iLOV (PDB:4EES), and CheA (PDB:4XIV).

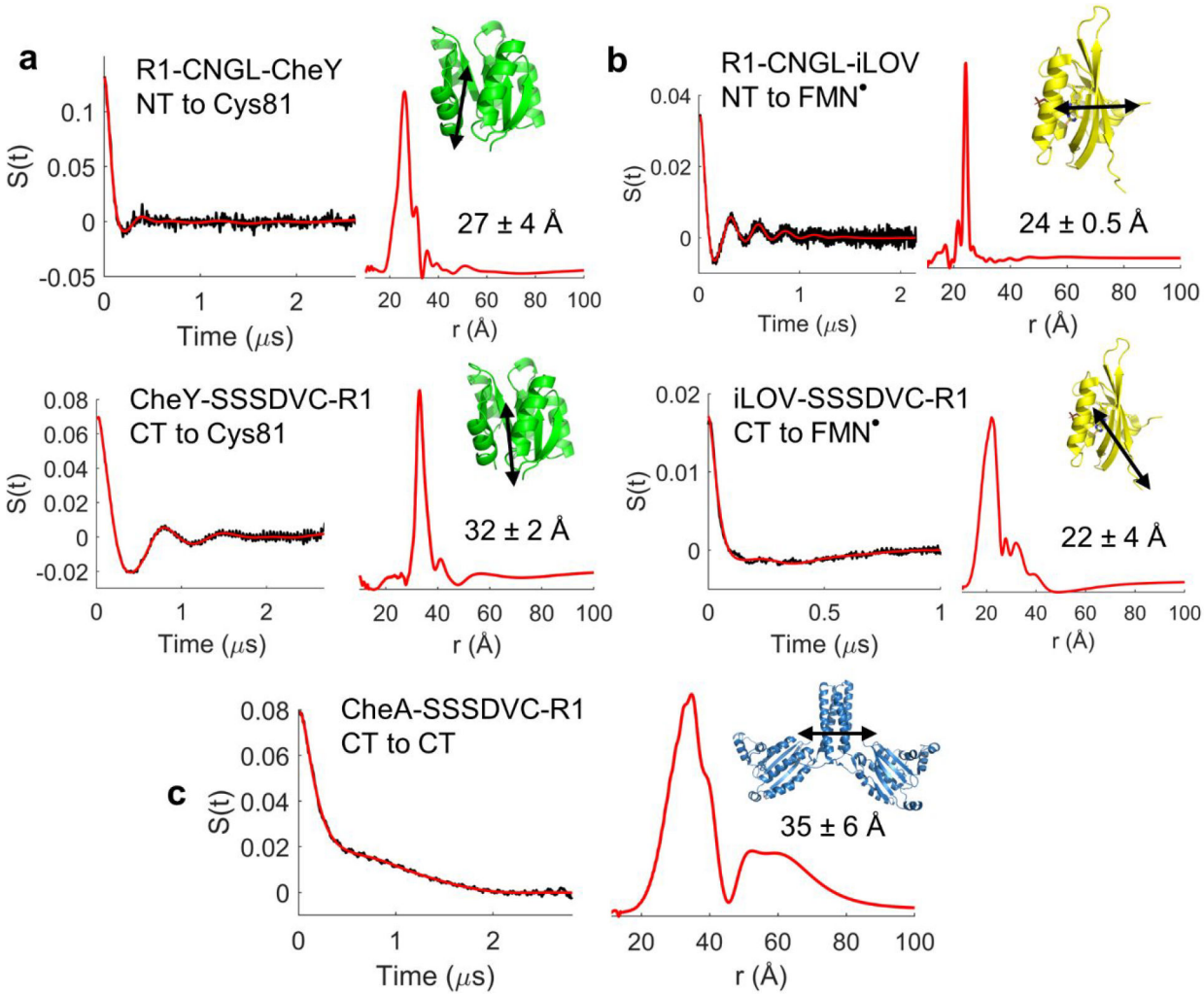


Figure 4. Q Band DEER Spectra and Distance Distributions of Select Proteins.

a) DEER Spectra of N (top) and C (bottom) terminally labeled CheY with R1-labelled Cys81 as the second spin. b) DEER Spectra of N (top) and C (bottom) terminally labelled light-activated iLOV with the FMN neutral semiquinone serving as the second spin. c) DEER spectra of C-terminally labeled dimeric CheA. For all spectra, DEER signal (black) and SVD reconstruction (red) are shown after background subtraction.

Table 1:
Observed and Expected DEER Distances.

Estimated linker length includes the number of residues in the attached peptide, as well as the disordered terminal residues not observed in the crystal structures.

		Crystal Structure Distance (Å)	Estimated Linker Length (Å) (N_{res})	Expected Distance (Å)	Observed Distance (Å)
CheY	NT to Cys81 (OaAEPl)	20.5	11.6 (6)	32.1	27 ± 4
	CT to Cys81 (gp41c)	18.0	13.2 (7)	31.2	32 ± 2
iLOV	NT to FMN (OaAEP1)	11.0	10.6 (4)	21.6	24 ± 0.5
	CT to FMN (gp41c)	13.6	12.4 (6)	26.0	22 ± 4
CheA	CT to CT (gp41c)	32.6	12.4 (6)	45.0	35 ± 6