# Journal Name

## ARTICLE

# Small Neutral Gd(III) Tags for Distance Measurements in Proteins by Double Electron–Electron Resonance Experiments

Mithun C. Mahawaththa,[†a] Michael D. Lee,[†b] Angeliki Giannoulis,[†c] Luke A. Adams,[†b] Akiva Feintuch,[c] James D. Swarbrick,[b] Bim Graham,[b] Christoph Nitsche,[a] Daniella Goldfarb*[c] and Gottfried Otting*[a]

Spin labels containing a Gd(III) ion have become important for measuring nanometer distances in proteins by double electron–electron resonance (DEER) experiments at high EPR frequencies. The distance resolution and sensitivity of these measurements strongly depend on the Gd(III) tag used. Here we report the performance of two Gd(III) tags, **propargyl-DO3A** and **C11** in DEER experiments carried out at W-band (95 GHz). Both tags are small, uncharged and devoid of bulky hydrophobic pendants. The **propargyl-DO3A** tag is designed for conjugation to the azide-group of an unnatural amino acid. The **C11** tag is a new tag designed for attachment to a single cysteine residue. The tags delivered narrower distance distributions in the *E. coli* aspartate/glutamate binding protein and the Zika virus NS2B-NS3 protease than previously established Gd(III) tags. The improved performance is consistent with the absence of specific hydrophobic or charge–charge interactions with the protein. In the case of the Zika virus NS2B-NS3 protease, unexpectedly broad Gd(III)–Gd(III) distance distributions observed with the previously published charged **C9** tag, but not the **C11** tag, illustrate the potential of tags to perturb a labile protein structure and the importance of different tags. The results obtained with the **C11** tag demonstrate the closed conformation in the commonly used linked construct of the Zika virus NS2B-NS3 protease, both in the presence and absence of an inhibitor.

## Introduction

Double electron-electron resonance (DEER, also called PELDOR) spectroscopy has become an established method for tracking protein conformations as it provides nanometre-scale distance distributions between two spin labels attached to the proteins at well-defined sites.[1] Most DEER applications have been carried out using the site-directed spin labelling approach with nitroxide spin labels.[2] Earlier on, the DEER measurements were carried out at X-band frequencies (~9.5 GHz), where sensitivity is limited, but recent developments of high-power Q-band spectrometers (~34 Gz) led to a considerable increase of sensitivity of DEER measurements with nitroxide spin labels.[3,4] Further increase in frequency to W-band (~95 GHz), while increasing sensitivity also introduces complications in data analysis arising from orientation selection.[5,6] In contrast, W-band has emerged as an efficient frequency for measuring nanometre distances in proteins using Gd(III) as spin labels.[7,8] DEER experiments

on Gd(III) labels are free of orientation selection and multi-spin effects.[8,9] Moreover, their chemical stability makes them ideal for in-cell distance measurements.[10-14] A Gd(III) ion can be used in conjunction with another Gd(III) ion to measure the Gd(III)–Gd(III) distance. Additionally, distances can also be measured with respect to a nitroxide,[15-17] a Mn(II) ion or to both in a triple spin system.[18] To use DEER distance measurements as a tool for the structure analysis of proteins, it is necessary to label the protein site-selectively with a tag that carries a Gd(III) ion. This poses stringent requirements on the Gd(III) tag, in particular, if the widths of the DEER distance distributions are to be interpreted in terms of protein structure variability.

So far no Gd(III) tag has been identified as universally perfect for all occasions. In fact, it is not obvious that such a tag exists. The 'ideal' Gd(III) tag has to fulfil a number of stringent criteria. (i) The tag must be chemically stable. Cyclen complexes of lanthanide ions show exceptional stability and have high affinity to the metal ions. They are thus the focus of the present work. (ii) The conjugation chemistry should be easy to implement and efficient so that the majority of the protein is readily labelled. For greatest utility, the tag should label the protein in the presence of solvent exposed cysteine residues. This can be achieved by genetic encoding of unnatural amino acids, which allow site-specific attachment of Gd(III) tags.[19] (iii) The Gd(III) ion within the tag must be held in a defined location relative to the protein. Most tags developed so far are attached to cysteine residues and thus result in a disulfide or thioether tether, which is usually flexible and, by allowing movements of the Gd(III) ion relative to the

protein, introduces broadening of the distance distributions in DEER experiments. Therefore, for improved distance resolution, the tags should feature either short or rigid linkers between the metal site and the protein backbone. While the peak of a DEER-derived distance distribution can often be predicted by simple modelling even for tags attached via long flexible tethers,[20-23] this depends on the assumption that the tag does not perturb the protein structure and does not engage in preferred, specific interactions with nearby amino acid residues. For tags with a net charge or hydrophobic pendants this may not be the case and the conformational ensemble is likely complicated by charge–charge or hydrophobic interactions to the protein. (iv) Finally, the EPR spectral properties of the Gd(III) spin label can play a major role in determining the sensitivity of DEER distance measurements. In principle, the best sensitivity is afforded by a spin label with small zero-field splitting (ZFS) to yield a narrow EPR spectrum. For such spin labels, however, the weak dipolar coupling approximation breaks down for shorter distances (below 4 nm) and the standard DEER data analysis artificially increases the width of the distance distribution.[24,25] Therefore, tags with a larger ZFS are preferred to probe short Gd(III)–Gd(III) distances.

Various cyclen-based Gd(III) tags have been synthesized and can be categorized according to their conjugation chemistry (Figure 1). The first two types of tags react with the thiol group of a cysteine residue to form either a disulfide bond (type **a**) or a thioether bond (type **b**). The third category, type **c**, reacts with an azide group from a genetically incorporated unnatural amino acid. The irreversible nature of type **b** and **c** conjugation makes them suitable for in-cell measurements.

The first cyclen-based Gd(III) tag used for DEER experiments was **C1** (Figure 1). It features a positive overall net charge and hydrophobic pendants.[20,26] Despite a relatively long and flexible tether, the maxima of the DEER distance distributions could be predicted with remarkable accuracy by rotamer libraries.[20] The tether of the related **C9** is shorter by two bonds, producing narrower DEER distance distributions.[27] Both **C1** and **C9**, however, feature a net positive charge and hydrophobic pendants, which may introduce difficult-to-control interactions with the protein surface.[28] In comparison, **C7** and **C8** are more compact in size and feature the same short linker as **C9**. Equally, they are enantiomerically pure with either *S* (**C7**) or *R* (**C8**) stereo-centres in the 2-hydroxypropyl pendant arms (Figure 1).[29] They deliver narrow distance distributions, have a large ZFS, and, as a special feature, display an EPR line-shape that is sensitive to the conjugation site.

Particularly narrow distance distributions can be achieved by attachment via a pair of proximal cysteines using double-arm tags such as **T1**. Even these tags, however, allow a range of Gd(III) ion positions.[23] Furthermore, they require more detailed knowledge of the 3D structure of the protein compared to their single-arm counterparts, as they depend on two, rather than one, adequately positioned cysteine residues. In an alternative approach, protein loops can be engineered to bind a Gd(III) ion through insertion of a lanthanide binding sequence of about 16 amino acid residues,[12,30] but this approach is structurally limited to loop regions, which may also be inherently mobile.

Cyclen tags containing three carboxylate pendants (derivatives of DO3A) such as **MTS-ADO3A** (Figure 1) form overall neutral Gd(III) complexes. In DEER distance measurements, **MTS-ADO3A** gave

broad distance distributions,[31,32] presumably due to a relatively long and flexible tether. The commercially available **maleimide DO3A** tag features a narrow EPR spectrum and has been used in in-cell DEER studies.[10,13] It has a rather long and potentially flexible tether and can also form different diastereomers upon conjugation to a cysteine thiol group, all of which contributes to broad distance distributions. In contrast, the **DO3MA-3BrPy** tag (Figure 1) is a new cyclen-based DO3A tag, which gives a shorter and more rigid tether on conjugation. It has successfully been used in in-cell DEER, but the ligation reaction is slower than for other Gd(III) tags.[14] **C3** (Figure 1) can be attached to a *p*-azidophenylalanine (AzF) residue by a copper-catalysed cyclo-addition (click) reaction.[19,33] AzF residues can readily be incorporated into proteins in response to an amber stop codon.[34,35] Although **C3** generates a rather long linker between Gd(III) ion and protein backbone, the tether is relatively rigid and, in most cases, the DEER distance distributions can accurately be modelled by random rotamer conformations.[19]
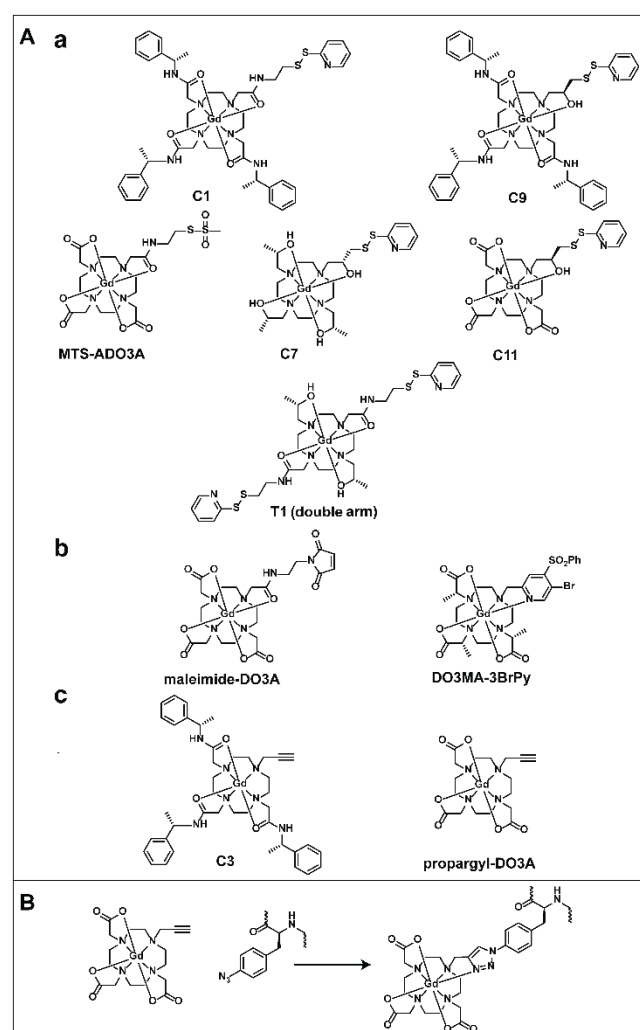


**Figure 1.** Gd(III) tags discussed in the present work (A). All tags are loaded with a Gd(III) ion but are referred to in the text by their name only, without explicitly stating the presence of the Gd(III) ion. Tags are grouped as type **a**, **b** and **c**, where types **a** and **b** are for attachment to cysteine residues, resulting in disulfide and thioether bonds, respectively, and type **c** undergoes copper-catalyzed cycloaddition reactions with azides as shown in (B). **C8** (not shown) is the enantiomer of **C7**.

The quest for the optimal Gd(III) tag for distance measurements led us to investigate two single-arm DO3A Gd(III) tags, the type **c propargyl-DO3A** tag and the type **b C11** tag, which feature small, overall uncharged cyclen–Gd(III) complexes devoid of large hydrophobic pendants. Gd(III) complexes of the **propargyl-DO3A** tag have been used before for magnetic resonance imaging (MRI) contrast agents[36,37] and DEER measurements of Gd(III)–Gd(III) distances in DNA.[38] The **C11** tag is a new DO3A tag designed for attachment to a single cysteine residue via a disulfide bond and a short tether. We tested the performance of these two tags on two proteins, the *E. coli* aspartate/glutamate binding protein (DEBP)[39] and the Zika virus NS2B-NS3 protease, which is an established drug target. Using DEBP, we compared the performance of the **propargyl-DO3A** tag with that of **C3** on the same mutants.[19] For the Zika virus NS2B-NS3 protease, we measured distances both in the presence and absence of inhibitor, using three different mutants for the **propargyl-DO3A** tag, four different mutants for **C11**, and the same four mutants for **C9**. In general, both **propargyl-DO3A** and **C11** performed better than the **C3** and **C9** tags, respectively. They produced relatively narrow distance distributions in DEER experiments and the maxima of the distance distributions were in good agreement with the crystal structures of both proteins. The **C11** tag produced narrow Gd(III)–Gd(III) distance distributions also for distances below 3 nm owing to its sufficiently large ZFS. Importantly, the distributions observed in the Zika virus protease were consistently much narrower than those obtained with the **C9** tag, which was previously shown to be an excellent tag for measuring narrow Gd(III)–Gd(III) distance distributions.[27]

## Results

### Performance of propargyl-DO3A

Two mutants of DEBP, Q80AzF/N146AzF and N48AzF/R169AzF, were labelled by reaction with **propargyl-DO3A** (Figure 2). The mutant N48AzF/R169AzF was chosen because previous experiments with the **C3** tag had produced an unexpectedly broad bimodal distance distribution, which was attributed to specific binding of the Gd(III) complex to the protein.[19] The Q80AzF/N146AzF mutant was chosen because it is a key construct to assess the conformational state of DEBP, where the amino acid binding site is accessible in the open and inaccessible in the closed conformation.[39] Labelled with the **C3** tag, this mutant also yielded a broader distance distribution than other pairs of tagging sites.[19] Repeating these measurements with the **propargyl-DO3A** tag, which features the same linker as **C3**, allows assessment of tag-dependent effects.
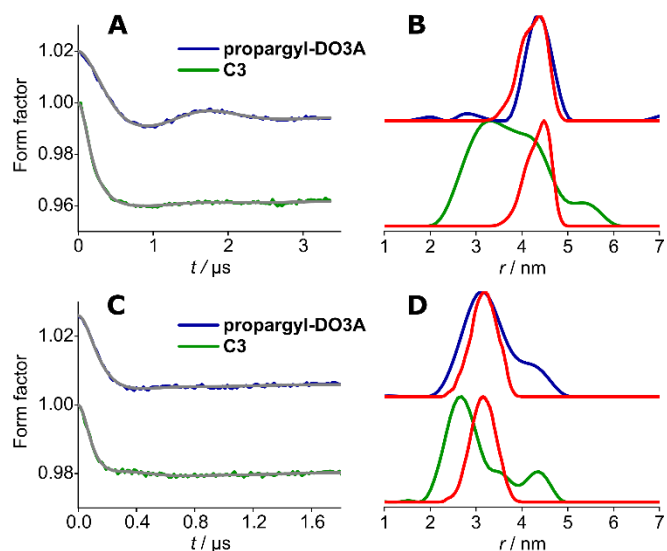
**Figure 2.** Crystal structure of the *E. coli* aspartate/glutamate binding protein (DEBP; PDB ID: 2VHA)[39] with models of the **propargyl-DO3A** tag ligated with AzF at four different sites, with the Gd(III) ions shown as black balls and the tag shown in blue for the mutant Q80AzF/N146AzF and in magenta for the mutant N48AzF/R169AzF.

Mass spectra indicated near-quantitative tagging yields of the DEBP mutants with the **propargyl-DO3A** and **C3** tags. Typical echo-detected (ED) EPR spectra of these mutants tagged with either **propargyl-DO3A** or **C3** are compared in Figure 3A. The spectrum of **propargyl-DO3A** is somewhat broader than that of **C3** and values of 1250 and 850 MHz were obtained by simulation for the centre of the axial ZFS parameter $D$ for these two tags, respectively (see Figure S4 in the Supporting Information). The yield parameter $\tau_{10\%}$ (the time at which the echo intensity reaches 10% of its initial value) determined from the echo decays of these samples is about 6 $\mu$s (Figure S5), which is within the range typical for Gd(III) tags in $D_2O$/glycerol-$d_8$ solutions (8:2 v/v).

complex in an unexpected location near the protein, indicating a specific binding interaction.[19] The measured distance distribution is also reproduced reasonably well by the distance distribution derived from a simple modelling approach, where all possible bond rotations of the tag, that are sterically compatible with the protein structure, are compiled in a rotamer library to predict the most abundant Gd–Gd distance and the width of the distance distribution (Figure 4B, red traces, and Figure S13).[40] To reproduce the distance distributions obtained with the **C3** tag would require explicitly taking into account hydrophobic and electrostatic interactions in molecular dynamics force fields, which is much more difficult to achieve.
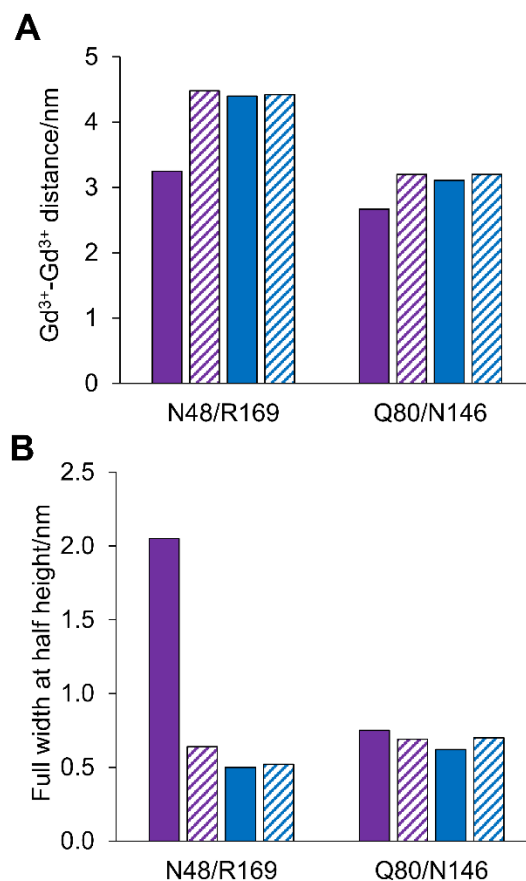
The DEER measurement of the mutant Q80AzF/N146AzF tagged with the **propargyl-DO3A** tag yielded a distance distribution similar in width to that obtained with the **C3** tag (0.6 nm; Figure 4C and D). Again, the experimental distance distribution was closely reproduced by modelling (Figure 4D, red trace), with a better match between experimental and predicted distances than for the **C3** tag.

In summary, the distances and distribution widths measured for both mutants with **C3** and **propargyl-DO3A** show that the **propargyl-DO3A** tag closely reproduces the maxima of the calculated distance distributions, while the predicted widths of the distance distribution are somewhat narrower than the experimental ones, which can be explained by some flexibility of the protein. In contrast, the **C3** tag delivered a biased distance distribution for the N48AzF/R169AzF mutant due to unexpected interactions between protein and tag and also for Q80AzF/N146AzF the agreement with the calculation is compromised. The comparison of the experimental and predicted distance distributions is summarized in Figure 5. Clearly, the availability of different tags provides a very important means to detect tag-specific artefacts.

Next we assessed the performance of the **propargyl-DO3A** tag with the Zika virus NS2B-NS3 protease. Previous NMR data had indicated that the C-terminal part of the NS2B co-factor, NS2Bc, may intermittently dissociate from NS3 in the commonly used construct, where NS2B and NS3 are connected by a $(Gly)_4$-Ser-$(Gly)_4$ polypeptide.[41] To assess the location of NS2Bc relative to NS3, we replaced T27 in NS3 by an AzF residue as well as either V67*, A77* or S85*, where the star identifies residues in the cofactor NS2B. Apart from the AzF mutations, the protein contained the covalent linker between NS2B and NS3 and the mutations C80S and C143S as described previously.[41-43] All three double-mutants, V67*/T27, A77*/T27 and S85*/T27, were designed to probe the presence of the closed conformation, which was determined by X-ray crystallography in the presence of the inhibitor cn-716.[41,44] In this conformation, NS2Bc contributes to forming the substrate binding site of the protease (Figure 6A). As cn-716 is known to stabilize the closed conformation,[41] we performed DEER measurements in the presence as well as in the absence of this inhibitor.



**Figure 4.** Comparison of DEER distance distributions obtained with the **propargyl-DO3A** and **C3** tags attached to DEBP. (A) DEER form factor of DEBP N48AzF/R169AzF labelled with the **propargyl-DO3A** tag and **C3** as noted in the figure. The time traces also show the fits obtained with the corresponding distance distributions shown in (B). The red line in (B, top) shows the distance distribution modelled based on the crystal structure (PDB ID: 2VHA[39]). (C) and (D) Same as (A) and (B), but for the mutant Q80AzF/N146AzF. The primary DEER data along with the background corrections and the uncertainty they introduce in the distance distribution are shown in Figures S10 and S11.



**Figure 5.** Bar graphs summarizing the experimental and calculated distances and widths of distance distributions for the DEBP mutants N48AzF/R169AzF and Q80AzF/N146AzF. Filled and hatched bars represent experimental and predicted data, respectively. Purple and blue bars pertain to measurements with **C3** and **propargyl-DO3A** tag, respectively. (A) Gd(III)–Gd(III) distances measured from the maxima of the distance distributions. (B) Full width at half maximum of the distance distributions.
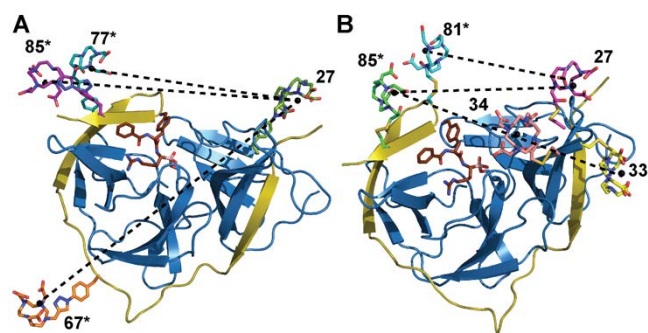
**Figure 6**. Crystal structure of the Zika virus NS2B–NS3 protease (PDB ID: 5LC0)[41] with Gd(III) tags modelled at different sites. NS2B and NS3 are shown as yellow and blue ribbons, respectively. Tagged residues and the inhibitor cn-716 (indicated in orange) are shown in a line representation. Distances measured by DEER experiments are shown as dotted lines. (A) Structure with AzF residues tagged with propargyl-DO3A in positions 67*, 77*, 85* and 27. (B) Structure with cysteine residues tagged with C11 in positions 81*, 85*, 27, 33 and 34.

The ED-EPR spectra and the echo decay curves of the Zika virus NS2B-NS3 protease mutants labelled with **propargyl-DO3A** are presented in Figure S7 and the echo-decays are presented in Figure S9A. The echo-decays were consistently shorter in the presence of inhibitor, in particular for the mutant A77*AzF/T27AzF. Nonetheless, the DEER results on the samples with inhibitor (Figure 7) yielded distance distributions that were in good agreement with the predicted distances, but the distributions were more than twice as broad as expected from rotamer library calculations (Figure 7C). Differences in the modulation depth arose from varying labelling efficiency, in part due to reduction of azido to amino groups during protein expression. Mass spectra indicated up to 20% reduction of each AzF residue in the Zika virus protease mutants and double-tagged samples were obtained with yields varying between 40 and 60% of the protein. As the **propargyl-DO3A** tag gave narrow distance distributions with DEBP, the experimentally observed widths of distance distributions suggest some conformational variability of the Zika virus protease. To verify the attribution of distance distribution widths to structural flexibility of the protein backbone, we also performed DEER measurements in the absence of the inhibitor. The flaviviral NS2B-NS3 proteases from dengue and West Nile viruses are known to include variable populations of open conformations in the absence of inhibitor, where the C-terminal part of NS2B is dissociated from NS3.[45-48] The distance distributions measured in the absence of inhibitor, however, were practically indistinguishable from those measured with inhibitor (Figure 7C). This indicates that the protease predominantly populates the closed conformation also in the absence of inhibitor. This is an interesting result, as previous NMR measurements could not distinguish between a model, where NS2B is largely dissociated from NS3, and a model, where NS2B undergoes local conformational changes without dissociating from NS3.[43]
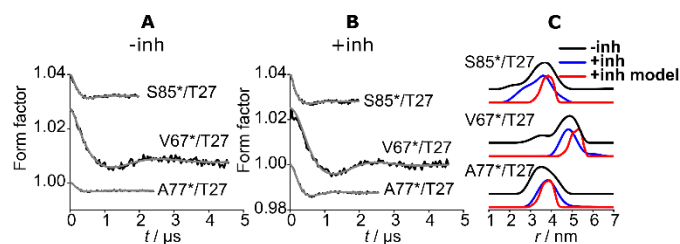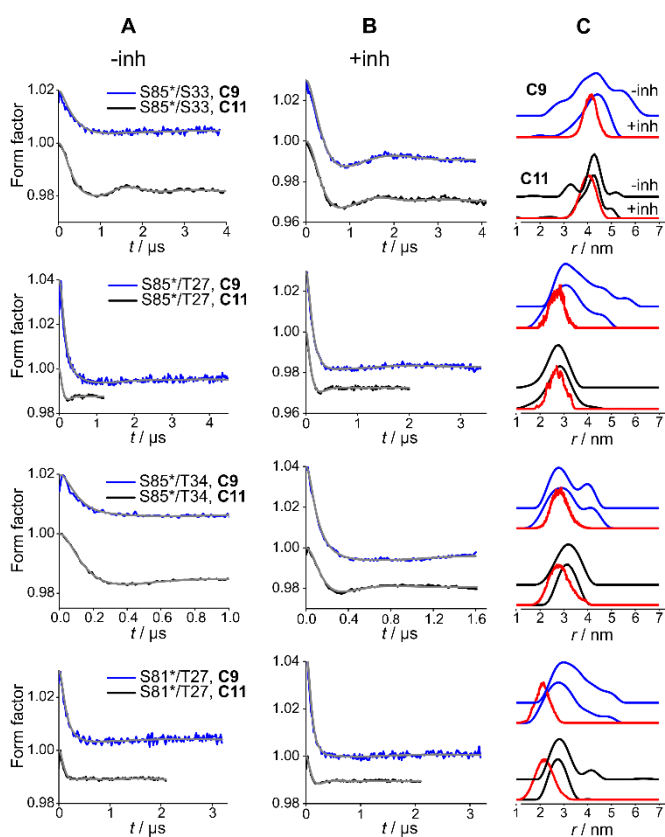


**Figure 7.** DEER distance distributions measured for the Zika virus NS2B-NS3 protease mutants labelled with **propargyl-DO3A**. (A) DEER form factors of the three mutants S85*AzF/T27AzF, V67*AzF/T27AzF and A77*AzF/T27AzF in the absence of inhibitor. The grey lines indicate the fits obtained with the distance distributions shown in (C). (B) Same as (A), but in the presence of the inhibitor cn-716.[41,44] (C) DEER distance distributions. The red traces are predicted distance distributions using the crystal structure 5LC0.[34] The primary DEER data along with the background corrections and the uncertainty they introduce in the distance distribution are shown in Figures S7 and S8. See Figure S10 for details on distance distribution predictions.

### Performance of C11

To compare the performance of the click tag **propargyl-DO3A** with those of cysteine-reactive tags, we also performed DEER measurements with the tags **C9** and **C11**. As the wild-type Zika virus NS2B-NS3 protease contains two highly solvent-exposed cysteine residues (Cys80 and Cys143), we first mutated both cysteine residues to serine.[41,43] Starting from this mutant, we prepared the cysteine double-mutants S81C*/T27C, S85C*/T27C, S85C*/S33C and S85C*/T34C (Figure 6B). Ligation with the **C9** and **C11** tags and preparation of samples with and without inhibitor cn-716 resulted in sixteen samples, for which DEER measurements were performed.

Representative ED-EPR spectra are displayed in Figure 3B, showing that the **C11** tag has a significantly broader EPR spectrum than **C9**, and slightly broader than the **propargyl-DO3A** tag. All other ED-EPR spectra are presented in the Supporting Information (Figure S7). Simulations of representative ED-EPR spectra are shown in Figure S8, yielding $D$ values of 800 and 1400 MHz for **C9** and **C11**, respectively.

Figure 8 shows the DEER results. The maxima of the distance distributions of both tags are similar for all mutants, as expected for their identical tether. The agreement with the distances predicted from the crystal structure was good except for the mutant S81*C/T27C, for which the experimental distances were about 0.5 nm longer than predicted. Most importantly, the **C11** tag delivered consistently narrower distance distributions than **C9** and it did not reveal any significant change in the distance distribution upon binding of the inhibitor. In contrast, the distance distribution obtained with **C9** was broader for all samples without the inhibitor, except for the S85C*/T34C mutant, where they were comparable. The widths of the experimental distance distributions obtained with **C11** were also much closer to the distributions simulated by rotamer library calculations (Figures 8C and 9).

**Figure 8.** DEER distance distributions measured for cysteine mutants of the Zika virus NS2B-NS3 protease labelled with **C9** (blue) and **C11** (black) tags. (A) DEER form factors without inhibitor. The grey line corresponds to the fit obtained with the distance distributions shown in (C). (B) Same as (A), but with inhibitor cn-716. (C) Distance distributions. The red traces correspond to the predictions based on rotamer libraries of the tags modelled on the closed conformation of the crystal structure 5LC0.[41] The primary DEER data along with the background corrections and the uncertainty they introduce in the distance distribution are shown in Figures S10 and S11.

The consistently broader widths of the distance distributions obtained for all mutants with **C9** versus **C11** could be associated with the narrower EPR linewidth of **C9** and its correspondingly smaller ZFS, which determines the validity of the weak coupling approximation commonly used in the derivation of distance distributions from DEER data. For short distances (below 4 nm) and small ZFS, we have shown the weak coupling approximation not to hold for Gd(III)–Gd(III) DEER experiments, as use of the standard data analysis software DeerAnalysis introduces artificial broadening and spurious peaks in the distance distribution, when the pseudo-secular terms of the dipolar interaction is neglected in the DEER data analysis.[24,25] This broadening can be reduced by conducting the DEER experiment with a large frequency difference, $\Delta\nu$, between the pump and observe pulses employing a dual-mode cavity, as has been experimentally confirmed with T4 lysozyme labelled with **C9**, where the Gd(III)–Gd(III) distance was around 4 nm.[49] Therefore, to verify the origin of the broadening in the **C9** samples, we carried out measurements on the S85C*/T27C mutant tagged with **C9** in the absence of inhibitor, using a dual-mode cavity with $\Delta\nu$ = 667 MHz (Figure S12). The distance distribution obtained was the same as that recorded with $\Delta\nu$ = 100 MHz (Figure 8), indicating that the extra broadening is

intrinsic with the tag and not due to neglecting pseudo-secular terms. This suggests that the chemical properties of **C9** (hydrophobicity and/or its positive net charge) significantly interfere with the association between NS2B and NS3. In agreement with this conclusion, the broadening of the distance distributions was particularly pronounced in the absence of inhibitor. In contrast, the uncharged and less hydrophobic **C11** tag delivered practically the same distance distributions in the presence or absence of inhibitor and, in general, these distance distributions were also closer to the distributions simulated by rotamer library calculations (Figures 8C and 9). Importantly, the narrow distance distribution obtained with **C11** even for distances below 3 nm shows that its ZFS is sufficiently large that the weak coupling approximation holds.

It is also interesting to compare the **C11** and **propargyl-DO3A** tags for the S85*/T27 mutants, where the **C11** tag produced a narrower distance distribution, which is in reasonable agreement with the predicted distance distribution width. In contrast, the **propargyl-DO3A** tag produced a wider distance distribution than predicted, like for the other two mutants of the Zika virus protease (Figure 9). This may reflect a greater sensitivity of the conformation of NS2B to tag-induced perturbations compared with DEBP, where the **propargyl-DO3A** tag produced narrow distance distributions as predicted (Figure 5B). Even though the disulfide-bonded **C11** tag tends to produce narrower distance distributions than the type **c propargyl-DO3A** tag, we conclude that both tags are well suited for investigating structural variability in the protein backbone, as the resulting tethers connecting the Gd(III) ion with the protein backbone are relatively rigid.

## Discussion

### Performance of the Gd(III) tags in DEER experiments

The two Gd(III) tags investigated in this work, **propargyl-DO3A** and **C11**, perform very well in Gd(III)–Gd(III) DEER distance measurements, yielding rather narrow distance distributions even for short distances. The absence of a net charge combined with a lower hydrophobicity of the tags appear to be an advantage, suggesting that it is important to limit the opportunities for
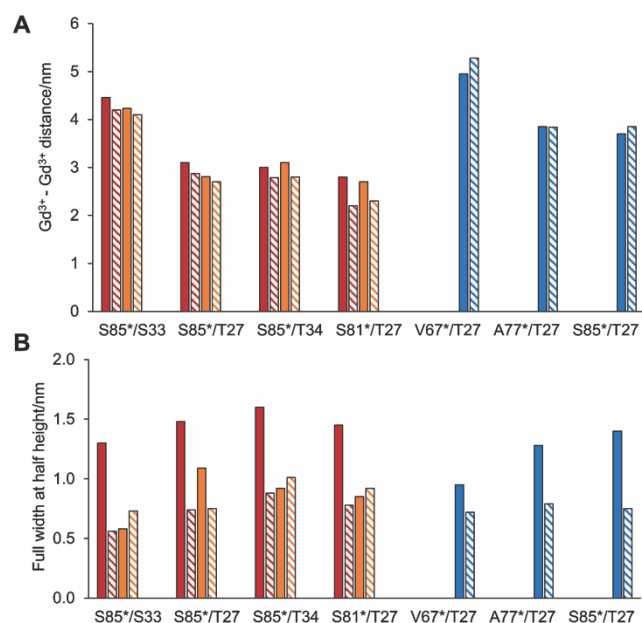
**Figure 9.** Bar graphs summarizing the experimental and predicted values of DEER distance distributions for the Zika virus NS2B-NS3 protease labelled with **C9**, **C11**, and **propargyl-DO3A** tags. (A) Graph showing Gd(III)–Gd(III) distances for the tags. Red bars represent the data for **C9**, orange bars represent the data for **C11** and blue bars represent the data for the **propargyl-DO3A** tag. The mutants are to cysteine for the data with **C9** and **C11** and to AzF for the data with **propargyl-DO3A**. Experimental and predicted data are represented by filled and hatched bars, respectively. (B) Graph showing the full width at half height of the distance distribution. Same colour scheme as in (A).

significant association energies of any non-covalent interactions between the cyclen moieties of the tags and the protein surface. Without such hard-to-predict interactions the maxima of distance distributions can readily be predicted by simple rotamer libraries. Perhaps most gratifying, however, was the observation that the **C11** tag yielded narrow and accurate distance distributions in the Zika virus NS2B-NS3 protease, when the **C9** tag yielded unexpected and significant broadening of DEER distance distributions. The **C9** tag appears to destabilize the interaction between NS2B and NS3, whereas the **C11** tag does not interfere with the native structure of the protease. Furthermore, the ZFS of **C11** is sufficiently large to suppress effects of the pseudo-secular terms of the dipolar Hamiltonian for Gd(III)–Gd(III) distances of about 3 nm. Recently, we have shown that also the enantiomeric **C7/C8** tags can produce narrow distance distributions.[29] These charged tags have the same short tether as **C11** and **C9**, are similarly small, and have a ZFS much larger than **C11**, which, as mentioned, simplifies the data analysis for short distances but reduces the sensitivity of DEER measurements.

Until now, the **C3** tag has been the only cyclen-Gd(III) tag designed for ligation to an AzF residue by copper-catalysed click chemistry. This ligation strategy is advantageous, as it is compatible with cysteine residues in the protein (such as in DEBP), allowing orthogonal labelling in combination with cysteine labelling and producing a tether that is stable under the reducing conditions found inside cells (even though the actual click reaction cannot be performed inside cells). Importantly, the resulting linkage (Figure 1B) has only a limited number of rotatable bonds, making the connection between the protein backbone and the Gd(III) ion quite rigid. Therefore, the **C3** and **propargyl-DO3A** tags can reliably report on

the backbone conformation, when AzF residues are introduced at sites, where the protein environment accommodates only a single rotamer for the dihedral angle $\chi_1$ of the AzF side chain.[19] While we previously observed specific non-covalent interactions between the cyclen moiety of the **C3** tag and the protein for one of six sites in DEBP,[19] the current work shows that the **propargyl-DO3A** tag at the same site is not subject to such detrimental non-covalent interactions.

While protein samples with site-specifically incorporated AzF residues are easy to produce in *E. coli* expression systems, a general drawback of the **C3** and **propargyl-DO3A** tags is the intolerance of some proteins towards the presence of copper ions required in the click reaction, resulting in precipitation during the reaction. Although this was not observed for the proteins of the present study, we have observed copper-induced precipitation in about half of the proteins we investigated in our laboratory. The design of tags for alternative attachment modes thus remains a topic of great interest.

The short tether of the **C11** tag makes it an attractive candidate for DEER measurements. Previously, the **C9** tag has been shown to give narrow distance distributions and accurate distance measurements, primarily due to a shorter tether between the Gd(III) ion and cysteine sulfur atom than for other cyclen tags.[27] The present study indicates that the smaller **C11** tag, which features the same linker, performs better largely because it is uncharged and/or less hydrophobic. Intuitively, a bulky tag should produce narrower distance distributions by restricting the conformational space accessible to the tag, but the present work suggests that this effect is relatively unimportant. In fact, the small size of the **C11** tag appears to be beneficial for narrow distance distributions, which may arise from weaker interactions between cyclen moiety and protein. Compared to the previously published DO3A tags **MTS-ADO3A** and **maleimide DO3A** (Figure 1),[31,32] the **C11** tag affords a shorter tether with the protein and a narrower distance distribution. Indeed, for distances greater than 4 nm, the width was among the narrowest measured for Gd(III)-Gd(III) distances (0.6 nm). Similarly narrow distance distributions have previously been obtained with the non-cyclen **PyMTA (**negatively charged**)** and **PyNPDA** (positively charged) tags on one ubiquitin mutant, but two other mutants produced greater widths, presumably due to protein flexibility in one of the labelling sites. The pyridyl tag **DO3MA-3BrPy**,[50] like **C11**, is small, uncharged and endowed with a sufficiently large ZFS to allow measurements of narrow distance distributions for short distances. It also has a more rigid tether, assuming that the pyridine nitrogen remains coordinated to the Gd(III) ion as designed. So far it has been tested only on one ubiquitin mutant (the same one that produced narrow distance distributions with **PyMTA** and **PyNPDA**), yielding a comparable distance distribution width (0.7 nm). These pyridyl tags all feature a rather large ZFS (> 1000 MHz).

### Open and closed conformations in the Zika virus NS2B-NS3 protease

The results obtained with the Zika virus protease indicate that the linked construct, where the C-terminus of NS2B is covalently connected via a (Gly)$_4$-Ser-(Gly)$_4$ polypeptide with the N-terminus of NS3, predominantly populates the closed conformation even in the absence of an inhibitor. This is an important result, as crystal structures of linked constructs do not show electron density for the

C-terminal β-hairpin of NS2B (PDB ID: 5TFN and 5T1V),[51] suggesting that it dissociates from NS3 in the absence of an inhibitor.

The question of open versus closed conformation has a long history in the field of flaviviral NS2B-NS3 proteases.[48] The proteases display full enzymatic activity only in the closed conformation, where NS2B participates in direct contacts with the substrate.[45] Linked constructs are easier to produce and more stable towards precipitation but have been found to destabilize the closed conformation.[52-54] Notably, the closed conformation is usually observed in the presence of inhibitors. Thus, in the case of the closely related NS2B-NS3 proteases from dengue and West Nile virus, crystal structures of the ligand-free proteins depict the C-terminal β-hairpin of NS2B completely separated from NS3,[45,46] although NMR experiments of a linked construct of the West Nile virus protease detected the closed conformation in solution.[55] In the presence of a tightly binding inhibitor, the linked construct of the dengue virus protease showed the closed conformation, both in solution[47,56] and in the single crystal.[57] In vivo, the covalent link between NS2B and NS3 is broken in the mature flaviviral proteases, and the absence of the linkage has been shown to promote the closed conformation in both the dengue virus[52,53] and the Zika virus protease.[58]

NMR data for the same linked construct of the Zika virus protease as investigated here showed the closed conformation in the presence of the inhibitor cn-716, which is known to stabilize the closed conformation, but were inconclusive in the absence of inhibitor.[43] Notably, for most of the mutants with the **C9** tag, the Gd(III)–Gd(III) distance distribution was narrower for samples prepared with the inhibitor, whereas the presence of inhibitor had no significant effect on the distance distribution widths measured with the smaller, neutral, and more hydrophilic **C11** tag. This indicates that the larger, charged, and more hydrophobic **C9** tag perturbs the conformation of the protease. For the Zika virus NS2B-NS3 protease, the Gd(III)–Gd(III) distance distributions measured with the **C11** tag establish the closed conformation as the predominant conformation not only in the presence but also in the absence of inhibitor.

## Conclusions

Conformational bias of tags and perturbation of protein structures are less likely for tags that are small, uncharged and hydrophilic, and these properties are essential for a reliable prediction of tag conformations based on a simple model of bond rotations restricted by steric interactions only. We have introduced two Gd(III) tags for DEER measurements, **propargyl-DO3A** and **C11**, which are small, uncharged and endowed with a short or rigid tether between the Gd(III) ion and protein backbone, resulting in narrow distance distributions with maxima in agreement with predictions. For the **C11** tag, the width of the distance distribution could also be predicted reasonably well, even for distances below 3 nm. This is an important result, which expands the scope of determining the structural variability of proteins by DEER analyses. In addition, our work calls for caution in the choice of tags by demonstrating that the widths of distance distributions cannot reliably be interpreted in terms of protein structure variability without assuring that the tags do not form unanticipated non-covalent interactions with the protein, which may destabilize the protein or bias the position of the

Gd(III) ion. In the absence of molecular dynamics force fields that allow the reliable prediction of the populations of different tag conformations and their effect on protein conformations, the availability of different tags clearly is of critical importance for conducting control experiments. Compared to previous tags, the **C11** tag performs outstandingly well, yielding accurate Gd(III)–Gd(III) distances between NS2B and NS3 of the Zika virus protease. These distances unequivocally show that, in the commonly used linked construct of the protease, the closed conformation prevails regardless of the presence of a high-affinity inhibitor.[41]

## Experimental Section

### Tag synthesis

**Propargyl-DO3A** was synthesised as described,[36] with minor modifications. The synthesis protocol of **C11** is described in the Supporting Information.

### Protein sample preparation

Expression constructs for the proteins DEBP and Zika virus NS2B-NS3 protease were based on the pETMCSIII vector.[59] Both proteins featured N-terminal His$_6$ tags followed by a TEV cleavage site. Samples with AzF were expressed in RF1-free B–95.ΔA E. coli cells.[60] To minimize the amount of usage of expensive unnatural amino acid, a 1 L cell-culture grown in LB medium was concentrated to 300 mL before induction with IPTG. Expression was conducted at 37 °C and limited to 3 hours after IPTG induction to limit the chemical reduction of AzF. Protein samples with double cysteine mutations were expressed in E. coli BL21(DE3) at 25 °C overnight after induction with IPTG. Cells were harvested by centrifugation at 5,000 g for 10 minutes and lysed by passing two times through a French Press (SLM Aminco, USA) at 830 bars. The lysate was centrifuged for 1 h at 34,000 g and the supernatant loaded onto a 5 mL Ni-NTA column (GE Healthcare, USA) pre-equilibrated with buffer A (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5% glycerol). The protein was eluted with buffer B (buffer A containing, in addition, 300 mM imidazole) and fractions were analyzed by 12% SDS-PAGE. Prior to ligation by the click reaction, the His$_6$-tags of proteins containing AzF were removed by digestion with TEV protease in 1:100 ratio overnight at 4 °C in buffer containing 50 mM Tris-HCl, pH 8.0, 300 mM NaCl and 1 mM β-mercaptoethanol. No attempt was made to remove the natural ligand (aspartate or glutamate) from DEBP.

Click reactions with the **propargyl-DO3A** and **C3** tags were performed as described.[19] The ligation of **C9** and **C11** tags was conducted by treating the protein with DTT for 1 h in buffer containing 50 mM Tris-HCl, pH 7.5, 300 mM NaCl and 1 mM DTT to make sure that all cysteine residues were reduced. Next, the sample was washed with wash buffer (50 mM Tris-HCl, pH 7.5, 300 mM NaCl) using a centrifugal filter unit (Amicon Ultra, molecular weight cut-off 10 kDa; Millipore, Billerica, USA) to remove DTT and the protein was incubated with 3-fold excess tag overnight at 25 °C. Samples were exchanged into EPR buffer (20 mM Tris-HCl, pH 7.5 (uncorrected pH meter reading), in D$_2$O) using a centrifugal filter unit and deuterated glycerol was added to a final concentration of 20%. All the protein samples were analysed using mass spectrometry to assess the tagging yields. An Elite Hybrid Ion Trap-Orbitrap mass spectrometer

(Thermo Scientific, USA) coupled with an UltiMate S4 3000 UHPLC (Thermo Scientific, USA) was used to analyse the mass and 7.5 pmol of sample were injected to the mass analyser via an Agilent ZORBAX SB-C3 Rapid Resolution HT Threaded Column (Agilent, USA). The distance distributions of the mutants were predicted using the program PyParaTools.[61]

### EPR Spectroscopy

All measurements were carried out at 10 K on a home-built pulse EPR spectrometer operating at W-band (94.9 GHz) frequencies.[62,63]

Echo-detected EPR spectra were recorded as a function of the echo intensity of a Hahn echo ($\pi/2 - \tau - \pi - \tau -$ echo) sequence sweeping the field. The spectra were recorded using $\pi/2$ and $\pi$ pulses of 15 ns and 30 ns, respectively, $\tau = 550$ ns and a repetition rate of 1 ms. A two-step phase cycle was applied.

Echo-decay spectra were recorded monitoring the echo intensity of a $\pi/2 - \tau - \pi - \tau -$ echo sequence at the maximum of the Gd(III) spectrum as a function of increasing $\tau$ values. The spectra were recorded using $\pi/2$ and $\pi$ pulses of 15 and 30 ns, respectively, and a repetition rate of 1 ms and 0.7 ms for the Zika virus protease and DEBP samples, respectively. A two-step phase cycle was applied.

DEER measurements were recorded using the standard four-pulse DEER ($\pi/2(v_{obs}) - \tau_1 - \pi(v_{obs}) - (\tau_1+t) - \pi(v_{pump}) - (\tau_2-t) - \pi(v_{obs}) - \tau_2 -$ echo) sequence using $\pi/2$ and $\pi$ pulses of 15 ns and 30 ns, respectively, $\tau_1 = 375$ ns for the Zika virus protease and $\tau_1 = 450$ ns for the DEBP samples, and a repetition rate of 800 µs. The delays t and the $\tau_2$ values were varied during the experiments. The pump pulse was set to the maximum of the Gd(III) spectrum, while the observe frequency was +100 MHz from the maximum. An eight-step phase cycle was applied. For the dual cavity measurements, the pump pulse length was 25 ns and those of the observe $\pi/2$ and $\pi$ pulses 15 and 30 ns, respectively. The repetition rate was 1 ms and $\tau_1 = 450$ ns. The pump frequency (94.5859 GHz) was set to the maximum of the EPR spectrum, while the frequency of the observe pulses was 667 MHz higher (95.2525 GHz).

### DEER data analysis

The primary DEER data were subjected to removal of background contributions that yielded a good fit and a reasonable Pake frequency pattern. The distance was obtained with Tikhonov regularization implemented into the DeerAnalysis2015 package.[64] The regularization parameter was chosen either by the L-curve criterion or by visual inspection in order to obtain reasonable fits.

Following Tikhonov regularization, the primary DEER data were subjected to statistical error analysis with respect to the background removal in order to obtain confidence intervals in the respective distance distributions. The distance distributions shown in the Supporting Information were validated in this way, using the validation tool of the DeerAnalysis2015 program, which allows considering different background corrections to yield the mean distance probability as well as upper and lower boundaries. The background validation was performed with a background start of 5% to 80% of the time window in 16 trials i.e. every 5% without the addition of random noise. All data sets were retained and the distance distributions normalized for display.

## Conflicts of interest

The authors declare no conflict of interest.

## References

1   G. Jeschke, *Annu. Rev. Phys. Chem.*, 2012, **63**, 419–446.
2   W. L. Hubbell, A. Gross, R. Langen and M. A. Lietzow, *Curr. Opin. Struct. Biol.*, **1998**, **8**, 649–656.
3   H. Ghimire, R. M. McCarrick, D. E. Budil and G. A. Lorigan, *Biochemistry*, **2009**, *48*, 5782–5784.
4   Y. Polyhach, E. Bordignon, R. Tschaggelar, S. Gandra, A. Godt and G. Jeschke, *Phys. Chem. Chem. Phys.*, 2012, **14**, 10762–10773.
5   G. W. Reginsson, R. I. Hunter, P. A. S. Cruickshank, D. R. Bolton, S. T. Sigurdsson, G. M. Smith and O. Schiemann, *J. Magn. Reson.*, 2012, **216**, 175–182.
6   Y. Polyhach, A. Godt, C. Bauer and G. Jeschke, *J. Magn. Reson.*, 2007, **185**, 118–129.
7   D. Goldfarb, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9685–9699.
8   A. Feintuch, G. Otting and D. Goldfarb, *Methods Enzymol.*, 2015, **563**, 415–457.
9   D.T. Edwards, T. Huber, S. Hussain, K.M. Stone, M. Kinnebrew, I. Kaminker, E. Matalon, M.S. Sherwin, D. Goldfarb and S. Han, *Structure,* 2014, **22**, 1677–1686.
10  A. Martorana, G. Bellapadrona, A. Feintuch, E. Di Gregorio, S. Aime and D. Goldfarb, *J. Am. Chem. Soc.*, 2014, **136**, 13458–13465.
11  M. Qi, A. Gross, G. Jeschke, A. Godt and M. Drescher, *J. Am. Chem. Soc.*, 2014, **136**, 15366–15378.
12  F.C. Mascali, H.Y.V. Ching, R.M. Rasia, S. Un and L.C. Tabares, *Angew. Chem. Int. Ed.*, 2016, **55**, 11041–11043.
13  F.-X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb and P. Selenko, *Nature*, 2016, **530**, 45–50.
14  Y. Yang, F. Yang, Y. J. Gong, J. L. Chen, D. Goldfarb and X.-C. Su, *Angew. Chem. Int. Ed.*, 2017, **56**, 2914–2918.
15  P. Lueders, G. Jeschke and M. Yulikov, *J. Phys. Chem. Lett.*, 2011, **2**, 604–609.
16  I. Kaminker, H. Yagi, T. Huber, A. Feintuch, G. Otting and D. Goldfarb, *Phys. Chem. Chem. Phys.*, 2012, **14**, 4355–4358.
17  L. Garbuio, E. Bordignon, E. K. Brooks, W. L. Hubbell, G. Jeschke and M. Yulikov, *J. Phys. Chem. B*, 2013, **117**, 3145–3153.
18  Z. Wu, A. Feintuch, A. Collauto, L. A. Adams, L. Aurelio, B. Graham, G. Otting and D. Goldfarb, *J. Phys. Chem. Lett.,* 2017, **8**, 5277–5282.
19  E. H. Abdelkader, A. Feintuch, X. Yao, L. A. Adams, L. Aurelio, B. Graham, D. Goldfarb and G. Otting, *Chem. Commun.*, 2015, **51**, 15898–15901.
20  H. Yagi, D. Banerjee, B. Graham, T. Huber, D. Goldfarb and G. Otting, *J. Am. Chem. Soc.*, 2011, **133**, 10418–10421.
21  G. Hagelueken, R. Ward, J. H. Naismith and O. Schiemann, *Appl. Magn. Reson.*, 2012, **42**, 377–391.
22  G. Jeschke, *Prog. NMR Spectrosc.*, 2013, **72**, 42–60.

23  P. Welegedara, Y. Yang, M. D. Lee, J. D. Swarbrick, T. Huber, B. Graham, D. Goldfarb and G. Otting, *Chem. Eur. J.,* 2017, **23**, 11694–11702.

24  A. Dalaloyan, M. Qi, S. Ruthstein, S. Vega, A. Godt, A. Feintuch and D. Goldfarb, *Phys. Chem. Chem. Phys*., 2015, **17**, 18464–18476.

25  N. Manukovsky, A. Feintuch, I. Kuprov and D. Goldfarb, *J. Chem. Phys*., 2017, **147**, 044201.

26  B. Graham, C. T. Loh, J. D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, G. Pintacuda, T. Huber and G. Otting, *Bioconjugate Chem.,* 2011, **22**, 2118–2125.

27  E. H. Abdelkader, M. D. Lee, A. Feintuch, M. R. Cohen, J. D. Swarbrick, G. Otting, B. Graham and D. Goldfarb, *J. Phys. Chem. Lett.,* 2015, **6**, 5016–5021.

28  E. Matalon, I. Kaminker, H. Zimmermann, M. Eisenstein, Y. Shai and D. Goldfarb, *J. Phys. Chem. B,* 2013, **117**, 2280–2293.

29  G. Prokopiou, M. D. Lee, A. Collauto, E. H. Abdelkader, T. Bahrenberg, A. Feintuch, M. Ramirez-Cohen, J. Clayton, J. D. Swarbrick, B. Graham, G. Otting and D. Goldfarb, *Inorg. Chem*., 2018, **57**, 5048–5059.

30  D. Barthelmes, M. Gränz, K. Barthelmes, K. N. Allen, B. Imperiali, T. Prisner and H. Schwalbe, *J. Biomol. NMR*, 2015, **63**, 275–282.

31  D. Thonon, V. Jacques and J. F. Desreux, *Contrast Media Mol. I.,* 2007, **2**, 24–34.

32  E. Matalon, T. Huber, G. Hagelueken, B. Graham, V. Frydman, A. Feintuch, G. Otting and D. Goldfarb, *Angew. Chem. Int. Ed.*, 2013, **52**, 11831–11834.

33  C. T. Loh, K. Ozawa, K. L. Tuck, N. Barlow, T. Huber, G. Otting and B. Graham, *Bioconjugate Chem.,* 2013, **24***,* 260–268.

34  J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang and P. G. Schultz, *J. Am. Chem. Soc.*, 2002, **124**, 9026–9027.

35  D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon and P. G. Schultz, *Biochemistry*, 2011, **50**, 1894–1900.

36  C. Vanasschen, N. Bouslimani, D. Thonon and J. F. Desreux, *Inorg. Chem*., 2011, **50**, 8946–8958.

37  L. Yan, L. Shen, H. Zhou, C. Wu, Y. Zhao, L. Wang, X. Fang, G. Zhang, J. Xu and W. Yang, *Tetrahedron*, 2016, **72**, 8578–8583.

38  Y. Song, T. J. Meade, A. V. Astashkin, E. L. Klein, J. H. Enemark and A. Raitsimring, *J. Magn. Reson*., 2011, **210**, 59–68.

39  Y. Hu, C.-P. Fan, G. Fu, D. Zhu, Q. Jin and D.-C. Wang, *J. Mol. Biol.*, 2008, **382**, 99–111.

40  A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting and D. Goldfarb, *J. Am. Chem. Soc.*, 2010, **132**, 9040–9048.

41  J. Lei, G. Hansen, C. Nitsche, C. D. Klein, L. Zhang and R. Hilgenfeld, *Science*, 2016, **353**, 503–505.

42  C. Nitsche, M. C. Mahawaththa, W. Becker, T. Huber and G. Otting, *Chem. Commun.*, 2017, **53**, 10894–10897.

43  M. C. Mahawaththa, B. J. G. Pearce, M. Szabo, B. Graham, C. D. Klein, C. Nitsche and G. Otting, *Antiviral Res.*, 2017, **142**, 141–147.

44  C. Nitsche, L. Zhang, L. F. Weigel, J. Schilz, D. Graf, R. Bartenschlager, R. Hilgenfeld and C. D. Klein, *J. Med. Chem*., 2017, **60**, 511–516.

45  P. Erbel, N. Schiering, A. D'Arcy, M. Renatus, M. Kroemer, S. P. Lim, Z. Yin, T. H. Keller, S. G. Vasudevan and U. Hommel *Nat. Struct. Mol. Biol.*, 2006, **13**, 372–373.

46  A. E. Aleshin, S. A. Shiryaev, A. Y. Strongin and R. C. Liddington, *Protein Sci*., 2007, **16**, 795–806.

47  L. de la Cruz, T. H. D. Nguyen, K. Ozawa, J. Shin, B. Graham, T. Huber and G. Otting, *J. Am. Chem. Soc.*, 2011, **133**, 19205–19215.

48  C. Nitsche, S. Holloway, T. Schirmeister and C. D. Klein, *Chem. Rev*., 2014, **114**, 11348–11381.

49  M. R. Cohen, V. Frydman, P. Milko, M.A. Iron, E.H. Abdelkader, M. D. Lee, J. D. Swarbrick, A. Raitsimring, G. Otting, B. Graham, A. Feintuch and D. Goldfarb, *Phys. Chem. Chem. Phys*., 2016, **18**, 12847–12859.

50  Y. Yang, Y.-J. Gong, A. Litvinov, H.-K. Liu, F. Yang, X.-C. Su and D. Goldfarb, *Phys. Chem. Chem. Phys.*, 2017, **19**, 26944–26956.

51  H. Lee, J. Ren, S. Nocadello, A. J. Rice, I. Ojeda, S. Light, G. Minasov, J. Vargas, D. Nagarathnam, W. F. Anderson and M. E. Johnson, *Antivir. Res*., 2017, **139**, 49–58.

52  Y. M. Kim, S. Gayen, C. Kang, J. Joy, Q. Huang, A. S. Chen, J. L. Wee, M. J. Ang, H. A. Lim, A. W. Hung, R. Li, C. G. Noble, L. T. Lee, A. Yip, Q.-Y. Wang, C. S. Chia, J. Hill, P. Y. Shi and T. H. Keller, *J. Biol. Chem*., 2013, **288**, 12891–12900.

53  L. de la Cruz, W.-N. Chen, B. Graham and G. Otting, *FEBS J.*, 2014, **281**, 1515–1533.

54  C. Kang, T. H. Keller and D. Luo, *Trends Microbiol*., 2017, **25**, 797–808.

55  X.-C. Su, K. Ozawa, R. Qi, S. G. Vasudevan, S. P. Lim and G. Otting, *PLoS Negl. Trop. Dis*., 2009, **3**, e561.

56  W.-N. Chen, C. Nitsche, K. B. Pilla, B. Graham, T. Huber, C. D. Klein and G. Otting, *J. Am. Chem. Soc.*, 2016, **138**, 4539–4546.

57  C. G. Noble, C. C. Seh, A. T. Chao and P. Y. Shi, *J. Virol*., 2012, **86**, 438–446.

58  Z. Zhang, Y. Li, Y. R. Loh, W. W. Phoo, A. W. Hung, C. B. Kang and D. Luo,  *Science*, 2016, **354**, 1597–1600.

59  C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love and N. E. Dixon, *Biochemistry*, 2000, **39**, 11989–11999.

60  T. Mukai, H. Hoshi, K. Ohtake, M. Takahashi, A. Yamaguchi, A. Hayashi, S. Yokoyama and K. Sakamoto, *Sci. Rep*., 2015, **5**, 9699.

61  M. Stanton-Cook, X.-C. Su, G. Otting and T. Huber, PyParaTools — Software for working with paramagnetic NMR data, http://comp-bio.anu.edu.au/mscook/PPT/, (accessed February 14, 2018).

62  D. Goldfarb, Y. Lipkin, A. Potapov, Y. Gorodetsky, B. Epel, A. M. Raitsimring, M. Radoul and I. Kaminker, *J. Magn. Reson*., 2008, **194**, 8–15.

63  F. M. Vigier, D. Shimon, V. Mugnaini, J. Veciana, A. Feintuch, M. Pons, S. Vega and D. Goldfarb, *Phys. Chem. Chem. Phys*., 2014, **16**, 19218–19228.

64  G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger and H. Jung, *Appl. Magn. Reson*., 2006, **30**, 473–498.

**Supporting information**

**Small neutral Gd$^{3+}$ tags for distance measurements in proteins by double electron–electron resonance experiments**

Mithun C. Mahawaththa, Michael D. Lee, Angeliki Giannoulis, Luke A. Adams, Akiva Feintuch, James D. Swarbrick, Bim Graham, Christoph Nitsche, Daniella Goldfarb, Gottfried Otting

**Synthesis of C11-Gd**

**Scheme 1.** Synthesis of **C11**-Gd. Reagents and conditions: (i) $K_2CO_3$, MeCN, reflux, 3 days; (ii) 2,2'-dipyridyldisulfide, triethylsilane, TFA, $CH_2Cl_2$, room temperature, 3 days, 20% (over two steps); (iii) $GdCl_3$, $H_2O$, pH 5, 80 °C, 2 h, 64%.

**Analytical instrumentation used in tag synthesis**

All 400 MHz NMR spectra were recorded on a Bruker Avance III 400 MHz NMR spectrometer. Chemical shifts are quoted in units of parts per million (ppm) and were referenced internally to the residual proteo-solvent resonance. Multiplicities and appearances of NMR resonances are abbreviated as: s, singlet; d, doublet; t, triplet; q, quartet; p, pentet; m, multiplet, app, apparent; br, broad. LC-MS data were acquired on an Agilent 1220/6120 LC/MS system, using ChemStation software for instrument control and data analysis. Preparative reverse-phase HPLC was performed on an Agilent 1260 Prep HPLC using an Alltima C8 column (250 mm x 22 mm, 5 micron).

**Material**

Tri-*tert*-butyl 2,2',2''-(1,4,7,10-tetraazacyclododecane-1,4,7-triyl)triacetate hydrobromide (($^t$Bu)$_3$DO3A·HBr) was purchased from Toronto Research Chemicals. (*R*)-1-chloro-3-(tritylthio)propan-2-ol was synthesised following literature procedures.[1]

**(*S*)-2,2',2''-(10-(2-hydroxy-3-(pyridin-2-yldisulfaneyl)propyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triyl)triacetic acid (C11)**

Potassium carbonate (1.5 g, 10.85 mmol) was added to a solution of ($^t$Bu)$_3$DO3A·HBr (400 mg, 0.67 mmol) and (*R*)-1-chloro-3-(tritylthio)propan-2-ol (225 mg, 0.75 mmol) in MeCN (15 mL). The mixture was heated to reflux for 3 days. After cooling to room temperature, insoluble salts were removed by filtration and the filtrate was concentrated under reduced pressure to yield a brown oil. The crude material was dissolved in a solution of 2,2'-dipyridyldisulfide (446 mg, 2.03 mmol) and triethylsilane (323 μL, 2.03 mmol) in CH$_2$Cl$_2$ (5 mL), before the dropwise addition of trifluoroacetic acid, TFA(5 mL). The solution was stirred for 3 days at room temperature, after which time LCMS analysis indicated complete *tert*-butyl group deprotection and conversion of the thiotrityl group to a pyridyldisulfide. Volatile solvents were removed by passing a gentle N$_2$ stream over the open reaction vessel. The resulting residue was taken up in CH$_2$Cl$_2$ (10 mL) and washed with a solution of TFA (0.1% v/v) in H$_2$O (10 mL), followed by H$_2$O (10 mL). The combined aqueous layers were purified by dry flash column chromatography, using Davisil® P60 C18 (35-70 µm) silica gel and a gradient from 0-20% MeCN in H$_2$O with 0.1% (v/v) TFA. Fractions containing pure product were lyophilized to yield **C11** as a beige solid. Yield: 73 mg (20%). **$^1$H NMR** (400 MHz, D$_2$O) δ 8.46 (m, 1H), 7.91 (m, 1H), 7.86 (d, *J* = 7.99 Hz, 1H), 7.37 (m, 1H), 4.14 (br s, 1H), 3.80–3.55 (m, 7 H), 3.41–2.90 (m, 23 H). **$^{13}$C NMR** (101 MHz, D$_2$O) δ 168.75, 158.00, 148.56, 139.35, 122.38, 66.11, 56.82, 55.81, 54.54, 49.49 (br), 43.36. **LC-MS**: *m/z* (ESI, 20V) 546.3 [M+H]$^+$.

**C11-Gd(III) tag**

**C11** (10 mg, 0.018 mmol) and GdCl$_3$ (5 mg, 0.20 mmol) were dissolved in H$_2$O (2 mL) and the pH of the solution was adjusted to ~ 5 by the addition of DIPEA. The solution was heated to 80 °C for 2 h, after which time LC-MS analysis indicated no uncomplexed **C11** remained. After cooling to room temperature, the complex was purified by reverse-phase HPLC (0.1% TFA and 5-95% MeCN over 20 min on a C8 preparative column). Fractions containing pure product were lyophilized to yield **C11-Gd(III)** as a beige solid. Yield: 8 mg (64%). **LC-MS:** *m/z* (ESI, 20V) 700.2 (complex isotope pattern) [M+H]$^+$.
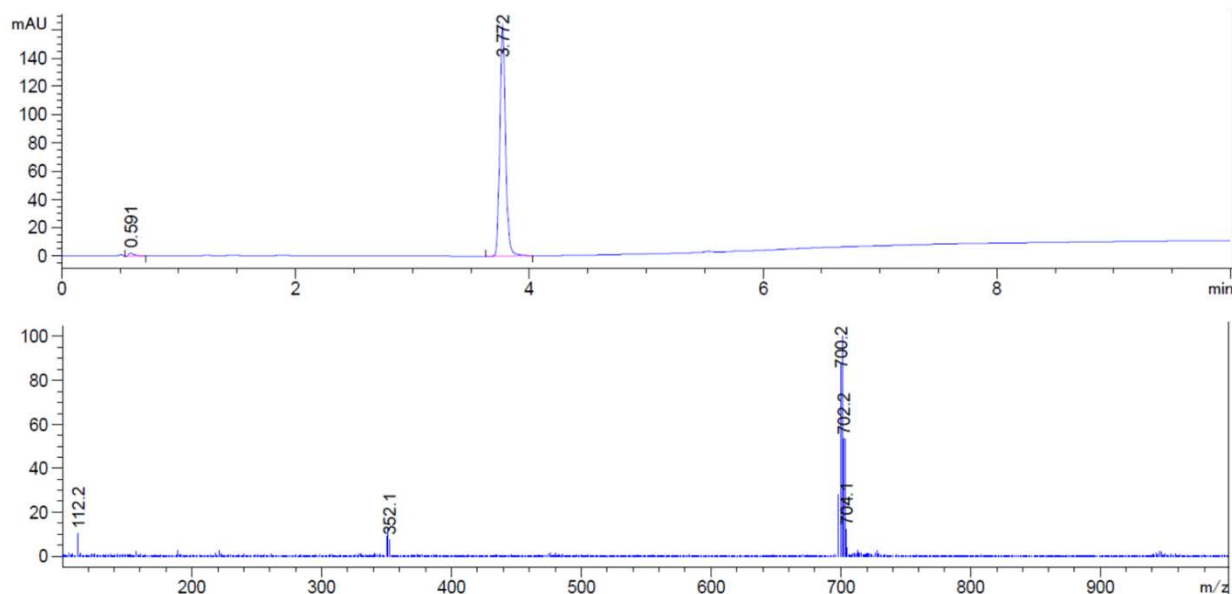
**NMR and LC-MS data**
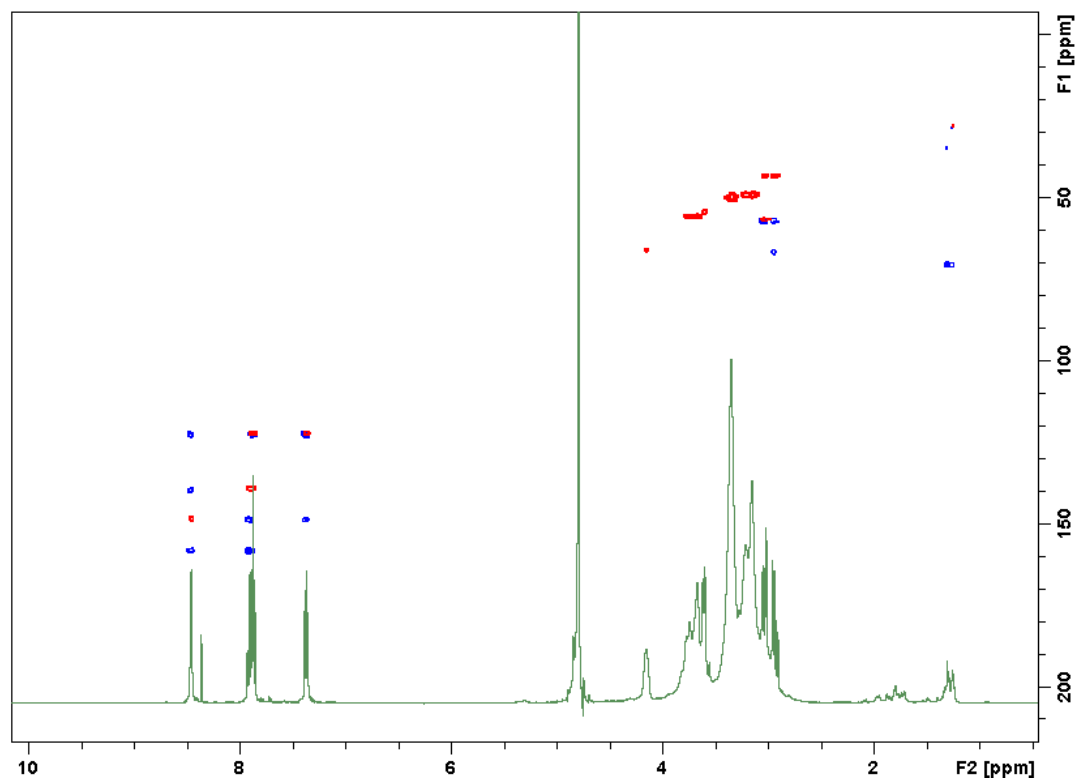


**Figure S1.** Superimposition of ${}^1$H-NMR (green), ${}^{13}$C-HSQC (red) and ${}^{13}$C-HMBC (blue) spectra of **C11** in D$_2$O at pH 3.
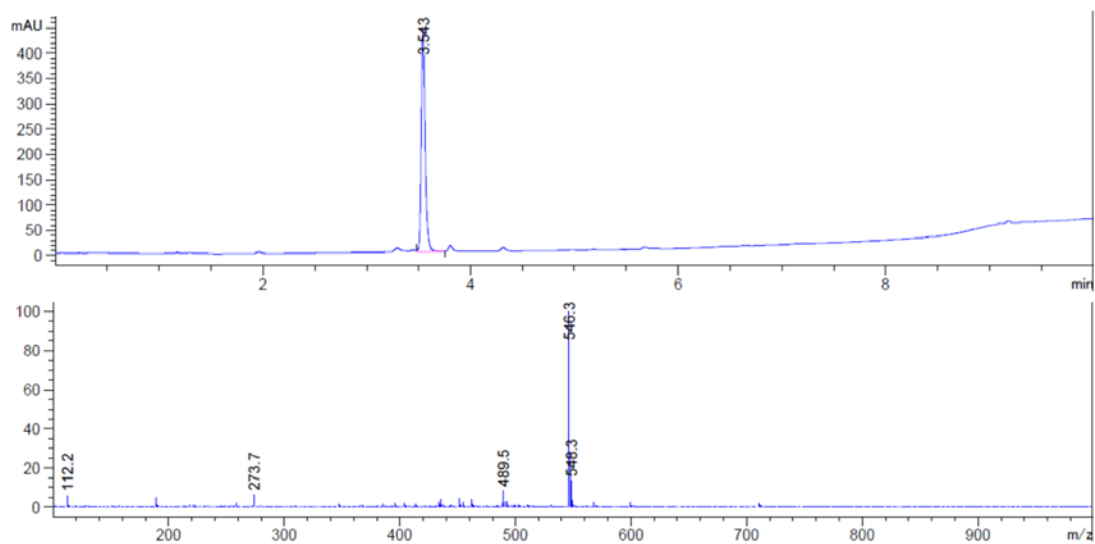


**Figure S2.** LC-MS UV trace at 254 nm (upper panel) and positive mass spectrum (lower panel) of **C11**.

**Figure S3.** LC-MS UV trace at 254 nm (upper panel) and positive mass spectrum (lower panel) of **C11** loaded with gadolinium.

**EPR measurements and analysis**

**Simulations of ED-EPR spectra of DEBP labeled with C3 or propargyl-DO3A**

The ED-EPR spectra were simulated using the 'pepper' function in the program EasySpin[2] and taking into account the distribution of the axial (*D*) and rhombic (*E*) parameters of the ZFS as suggested by Raitsimring *et al.*[3,4] The distribution over *D* is given by two Gaussian functions centered at –*D* and +*D* with equal widths $\sigma_D$ and equal weights. The probability distribution of *E*/*D* is given by

$$P(E/D) = (E/D) - 2(E/D)^2 \tag{1}$$

The frequency was 94.9 GHz and the temperature was 10 K. To obtain good fits we had to introduce a linewidth (*lw*). Table S1 shows the parameters used in the simulations of Figures S4 and S8. The simulations did not take into account the underestimation of the amplitude of the broad background arising from an adjustment of the nominal $\pi/2$ and $\pi$ pulses to the central transition, which has the highest transition probability. This may be one of the reasons for not obtaining better fits.
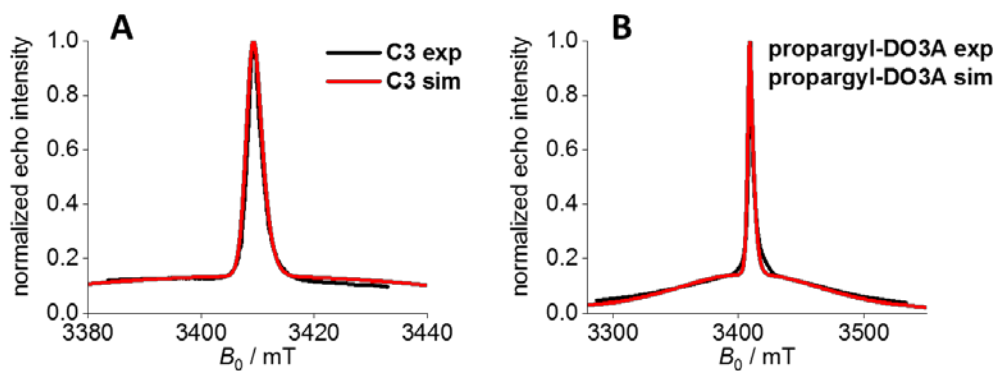
**Figure S4.** Simulations of the ED-EPR spectra of DEBP and comparison with experimental spectra. (A) N48/R169 mutant labelled with the **C3** tag. (B) N48/R169 mutant labelled with **propargyl-DO3A**.

**Table S1.** Summary of the parameters used to simulate the spectra presented in Figures S1 and S5.

| Tag | $D$/MHz | linewidth/mT |
|---|---|---|
| **C3** | 850 | 2 |
| **propargyl-DO3A** | 1250 | 1 |
| **C11**[a] | 1400 | 1 |
| **C9**[b] | 800 | 3 |

[a] Different weights of the two Gaussians were required for a reasonable fit. The relative weights for the -$D$ and $D$ parameters were 100:1.

[b] The $D$ value of **C9** is probably overestimated, as an overly large linewidth had to be added to reproduce the broad signal due to the transitions other than the central one.

**Echo decays of DEBP**
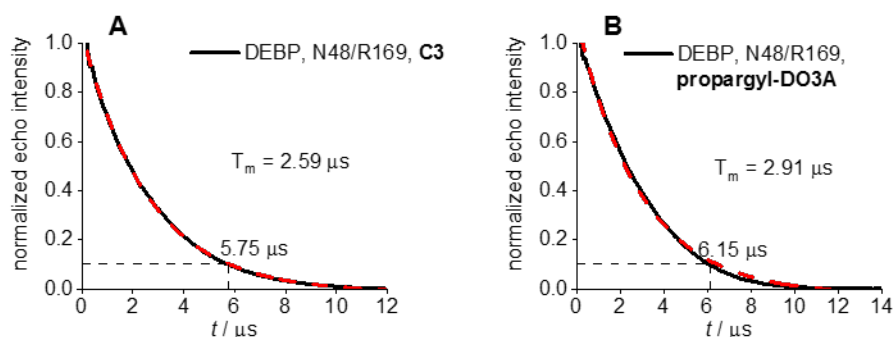


**Figure S5.** Echo decay data measured at W-band and 10 K. The $T_m$ values obtained by fitting a mono-exponential decay (shown as a red dashed line) are given as well as the values at which the echo intensity has decayed to 10% of its initial intensity. (A) DEBP mutant N48/R169 labelled with the **C3** tag. (B) DEBP mutant N48/R169 labelled with **propargyl-DO3A**.

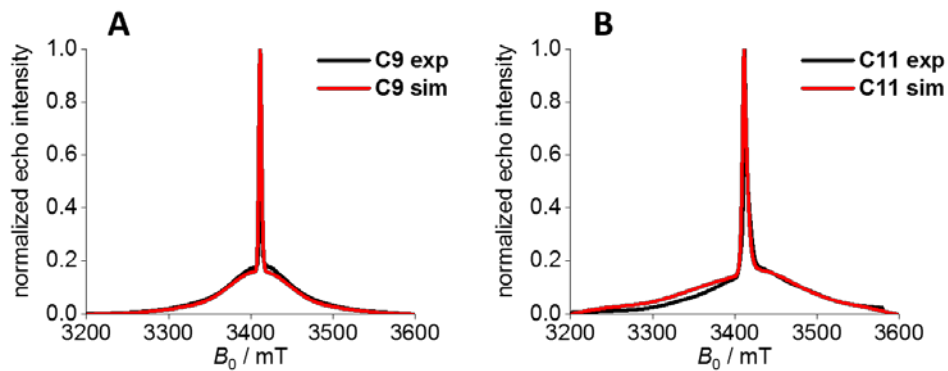**ED-EPR spectra of the Zika virus NS2B-NS3 protease mutants and simulations**
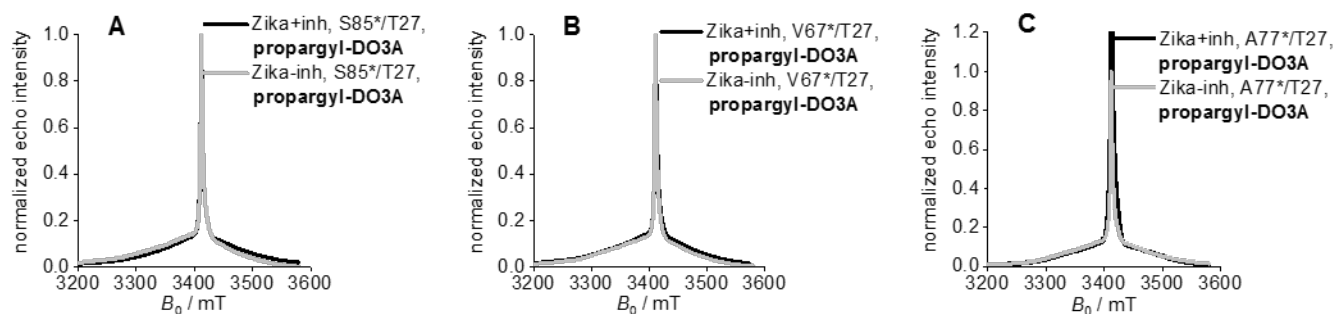


**Figure S6.** ED-EPR spectra of the Zika virus NS2B-NS3 protease mutants tagged with **propargyl-DO3A** measured at W-band at 10 K. Spectra recorded with and without the inhibitor cn-716 are shown in black and grey, respectively. (A) Mutant S85*AzF/T27AzF. (B) Mutant V67*AzF/T27AzF. (C) Mutant A77*AzF/T27AzF.
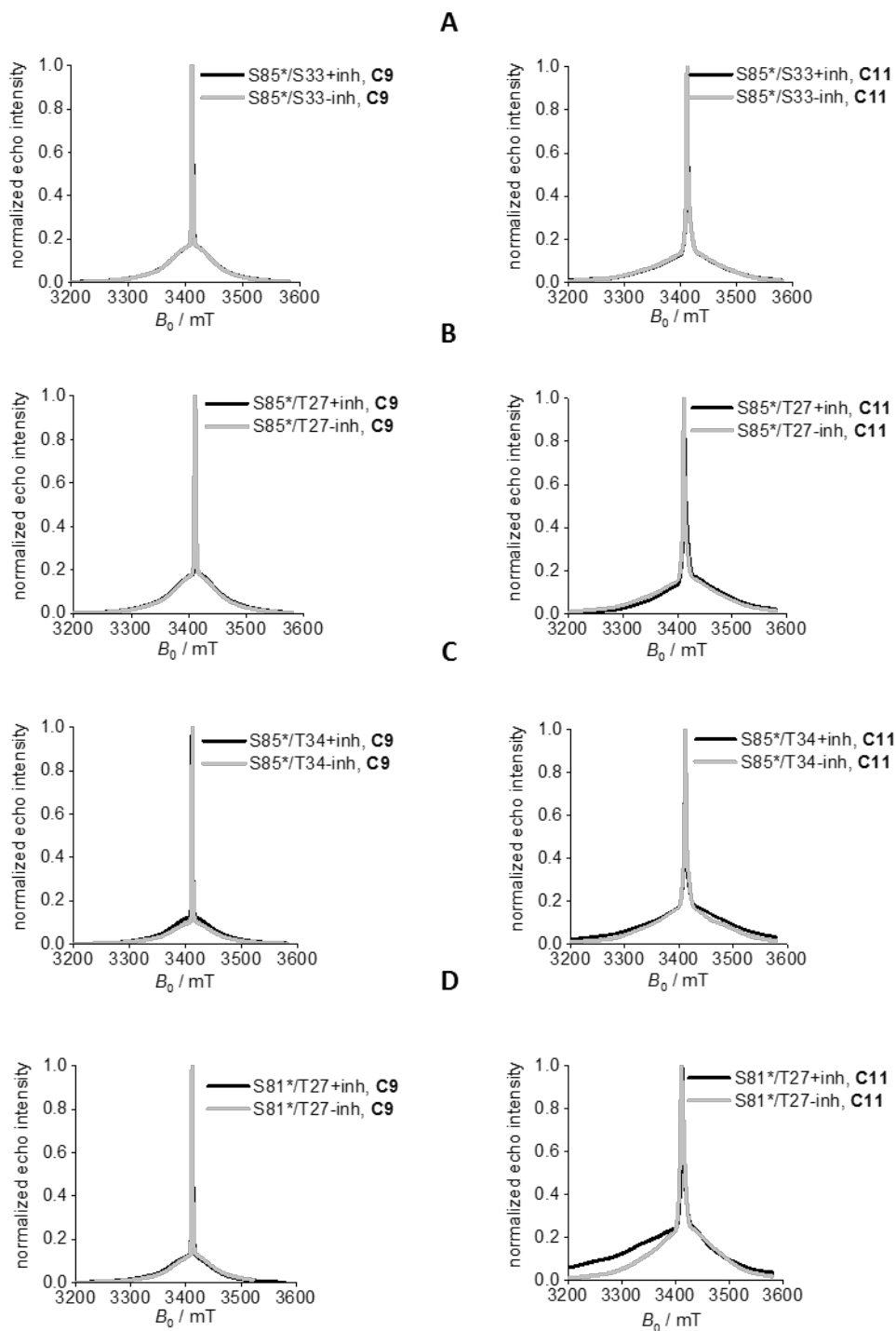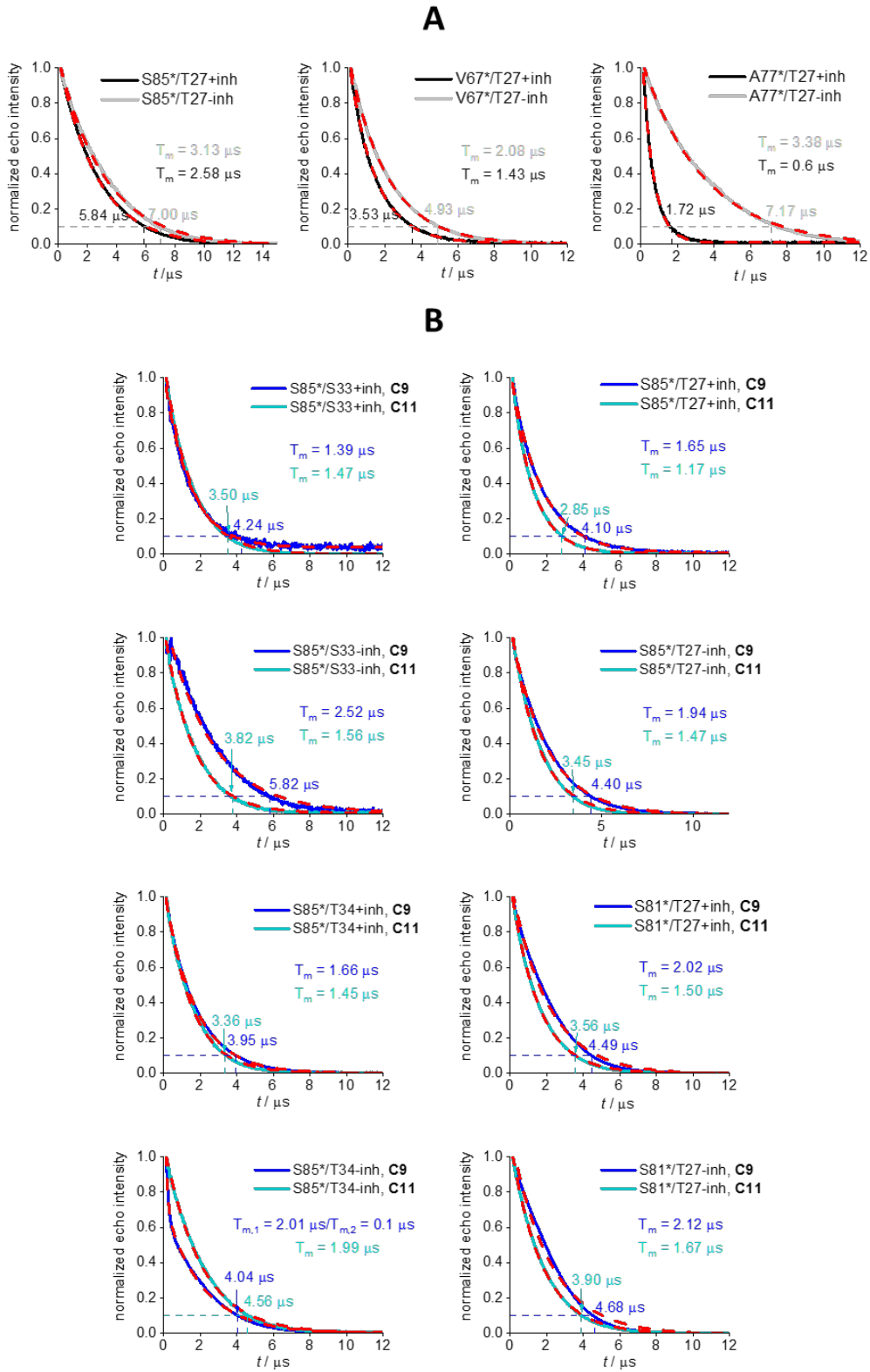
**Figure S7.** ED-EPR spectra of the Zika virus NS2B-NS3 protease mutants tagged with **C9** (left panels) or **C11** (right panels) measured at W-band at 10 K. Spectra recorded with and without the inhibitor cn-

716 are shown in black and grey, respectively. (A) Mutant S85*C/S33C. (B) Mutant S85*C/T27C. (C) Mutant S85*C/T34C. (D) Mutant S81*C/T27C.



**Figure S8.** Simulations of the ED-EPR spectra of Zika virus NS2B-NS3 protease mutant S85*C/T27C mutant and comparison with experimental spectra. (A) Labelled with the **C9** tag. (B) Labelled with the **C11** tag.

**Echo decay data for all Zika virus NS2B-NS3 protease samples**

**Figure S9.** Echo decay data measured at W-band at 10 K at the maximum of the ED-EPR spectra of the Zika virus protease mutants. The $T_m$ values are given, as well as the times at which the echo intensity had decayed to 10% of its initial intensity, according to the exponential decay obtained by fitting (mono-exponential for all mutants except for the sample "S85*/T34-inh, **C9**", where a biexponential decay function was used). The fits are shown by red dashed lines. The times of echo intensity decay to 10%, $\tau_{10\%}$, provide estimates of the duration of evolution times at which DEER experiments can be conducted. For the mutant "S81*/T27-inh, **C9**", we could not obtain a good fit and the $T_m$ value given here was obtained by fitting a monoexponential decay function. (A) Mutants (left to right): S85*AzF/T27AzF, V67*AzF/T27AzF, A77*AzF/T27AzF. Data recorded with and without the inhibitor cn-716 are shown in black and grey, respectively. (B) Double-cysteine mutants. The mutation sites and presence or absence of inhibitor are specified in each panel. Dark blue: data recorded with the **C9** tag. Light blue: data recorded with the **C11** tag.
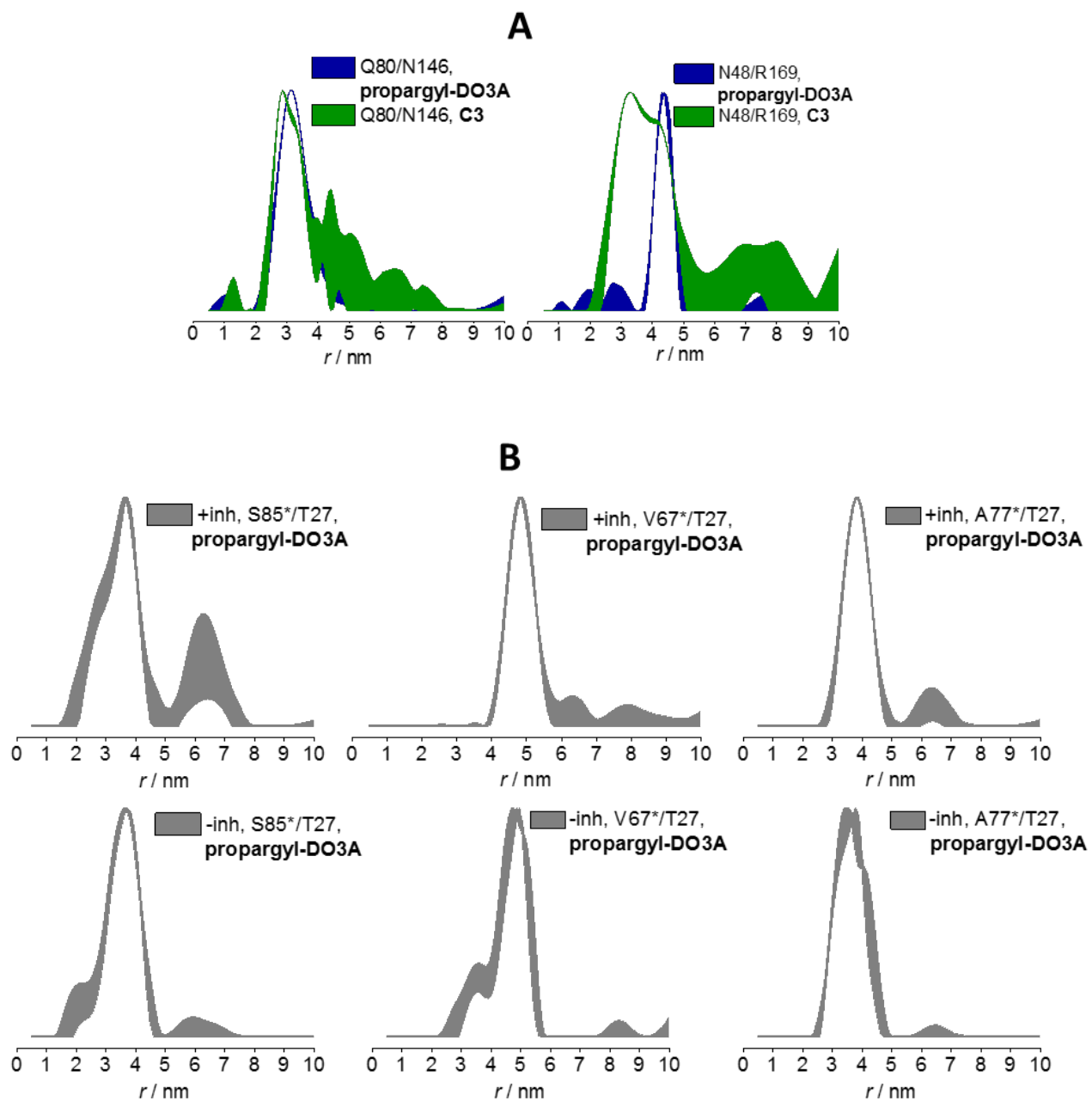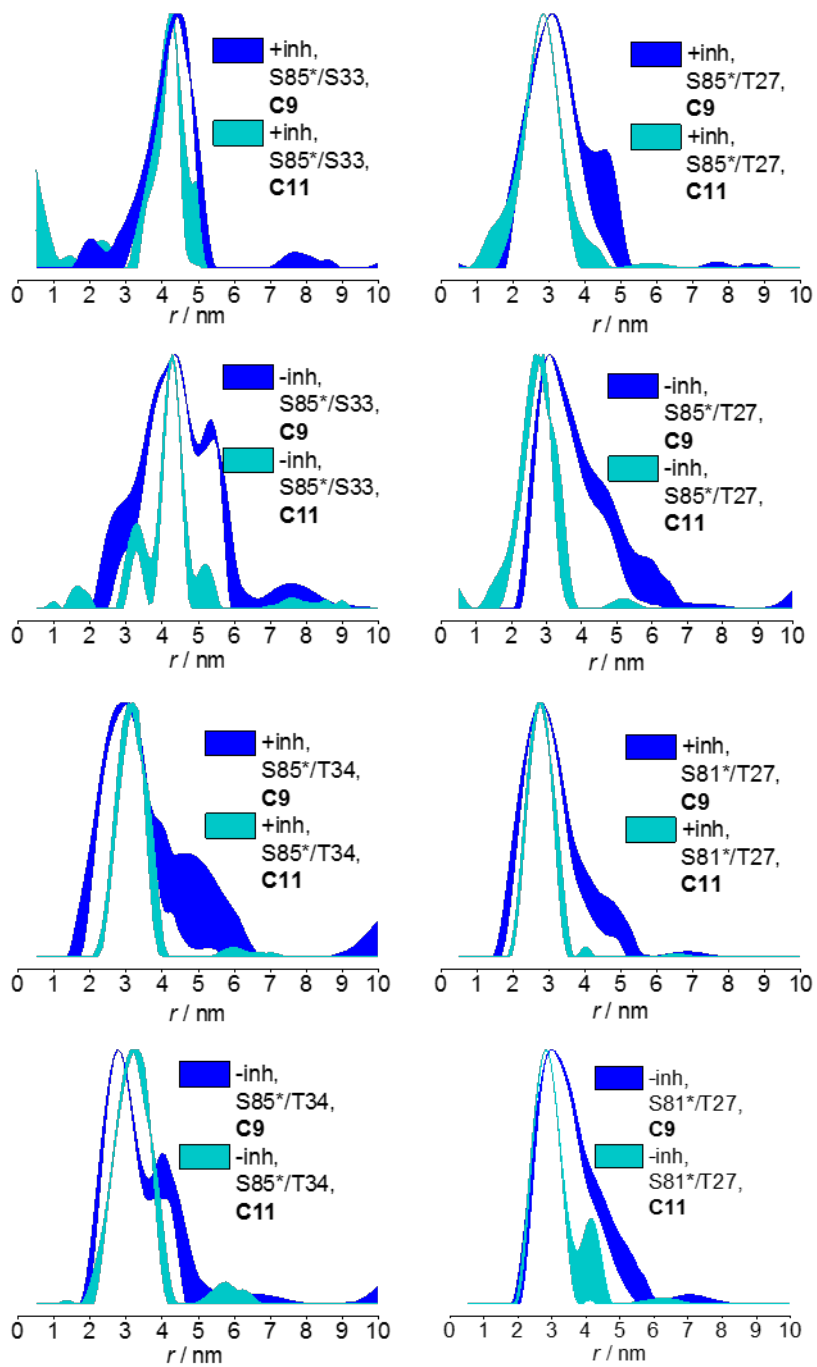
## Primary DEER data
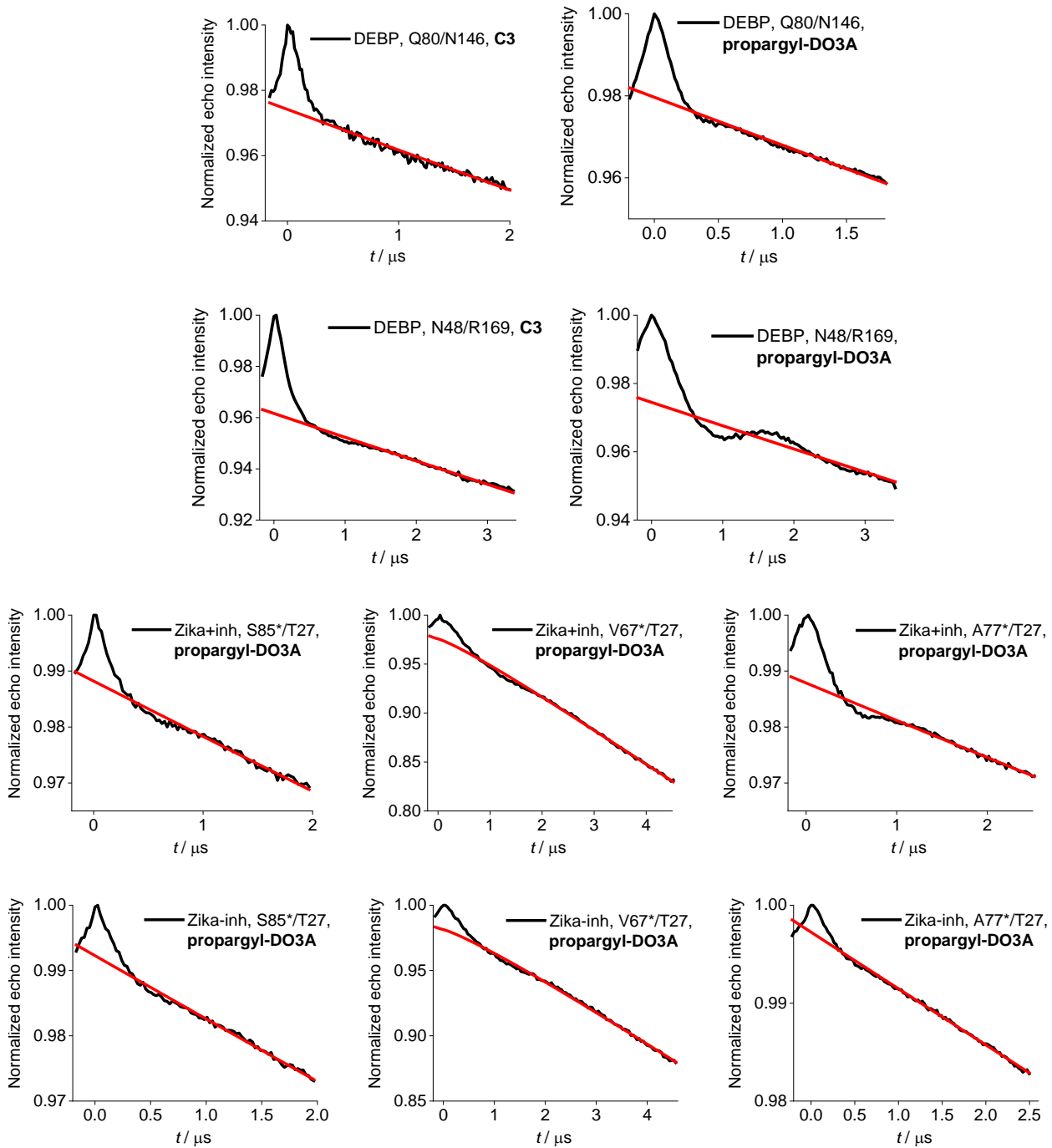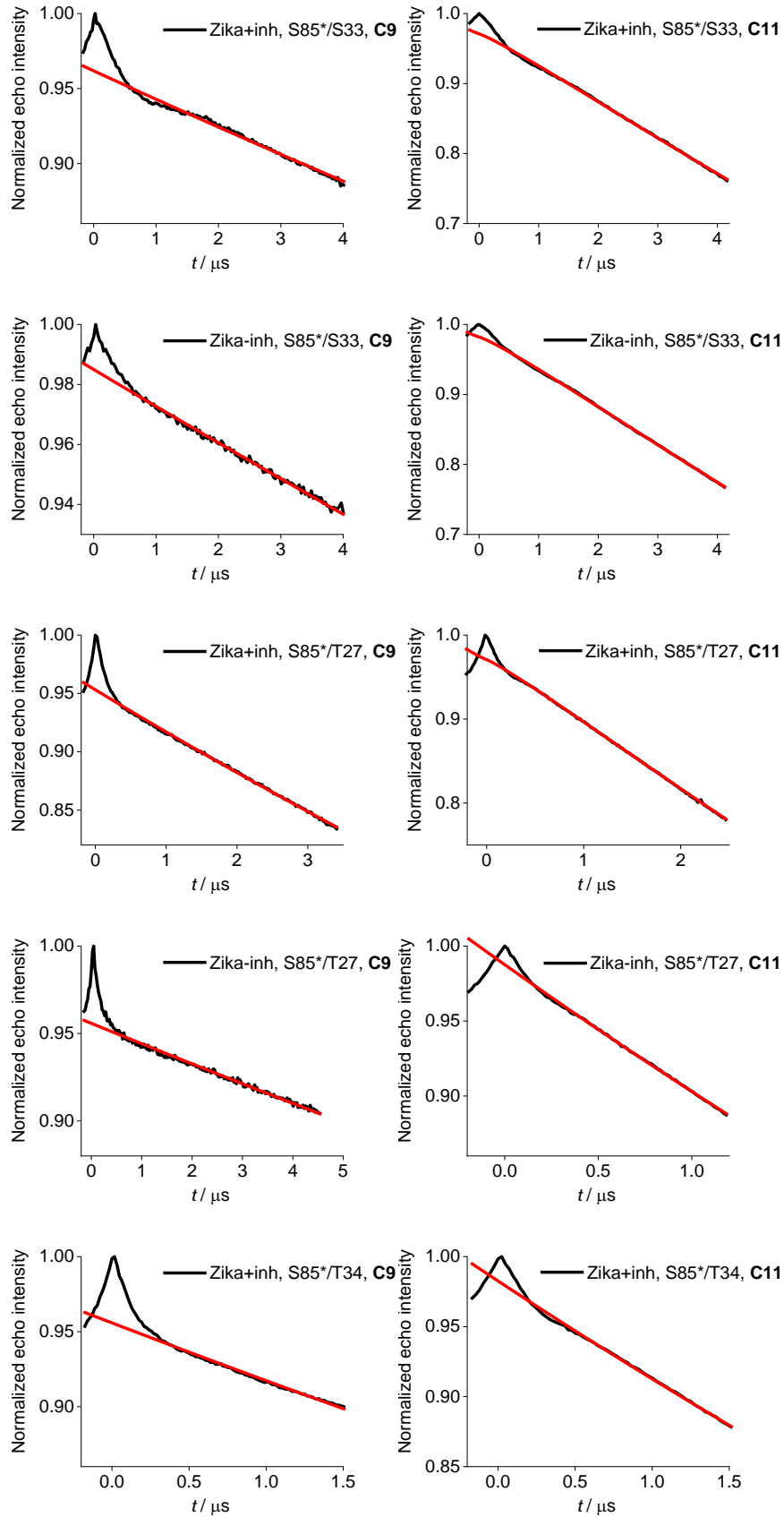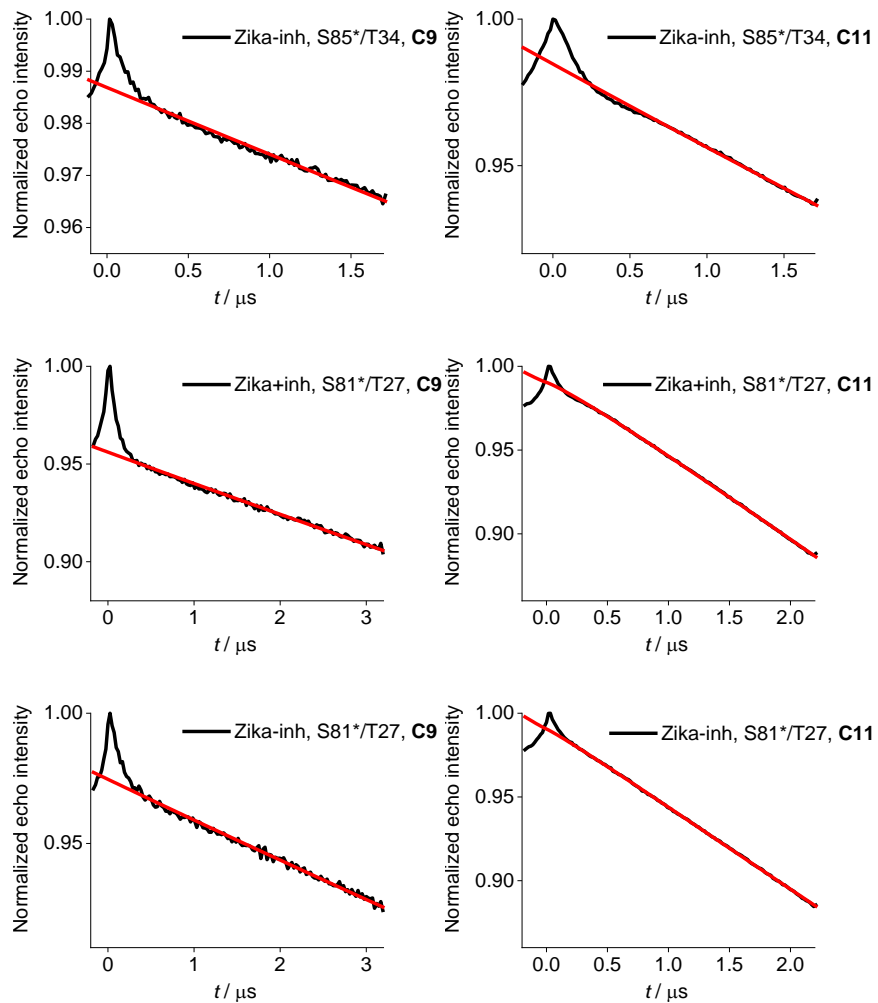
**Figure S10.** Primary DEER data (black) and fitted background decays (red). The protein, mutant, presence of inhibitor and tag are indicated in each panel.

**Validated distance distributions**

**Figure S11.** Distance distributions for all mutants including confidence intervals. The mutant, presence of inhibitor and tag are indicated in each panel. (A) DEBP mutants. (B) Zika virus NS2B-NS3 mutants.

**DEER with dual-mode cavity**

When performing Gd(III)−Gd(III) DEER distance measurements, one needs to consider the pseudo-secular terms of the dipolar Hamiltonian in order to reliably extract the distance distribution(s).[5] Neglecting the pseudo-secular terms has been found to be reasonable for distances above 3.4 nm,[6] but not necessarily for shorter Gd(III)−Gd(III) distances. In the latter case, processing of the time domain data using the 'common practice' DeerAnalysis software can cause artificial broadening of the distance distributions, as the software utilises a kernel function that assumes the weak coupling approximation to be valid. The effect is more pronounced for short Gd(III)−Gd(III) distances and small zero-field splittings (i.e. a narrow central EPR line corresponding to the $|−1/2 \rightarrow |+1/2\rangle$ transition of Gd(III)). It has been shown that this artificial broadening can be overcome experimentally by using large probe-pump offsets that increase the contribution from higher order transitions. In the present work, this was achieved by performing the DEER experiment using a dual cavity probe,[7] which allows the use of a large probe-pump offset (Figure S9). This enabled us to test whether the broad distance distribution, which we had measured for the S85*C/T27C mutant labelled with the **C9** tag and in the absence of inhibitor, originates from protein/tag flexibility or artificial broadening due to the neglect of pseudo-secular terms in the analysis of data recorded with the standard probe-pump offset of 100 MHz. Equally broad distance distributions were obtained in both setups, indicating intrinsic flexibility of the protein and/or tags.
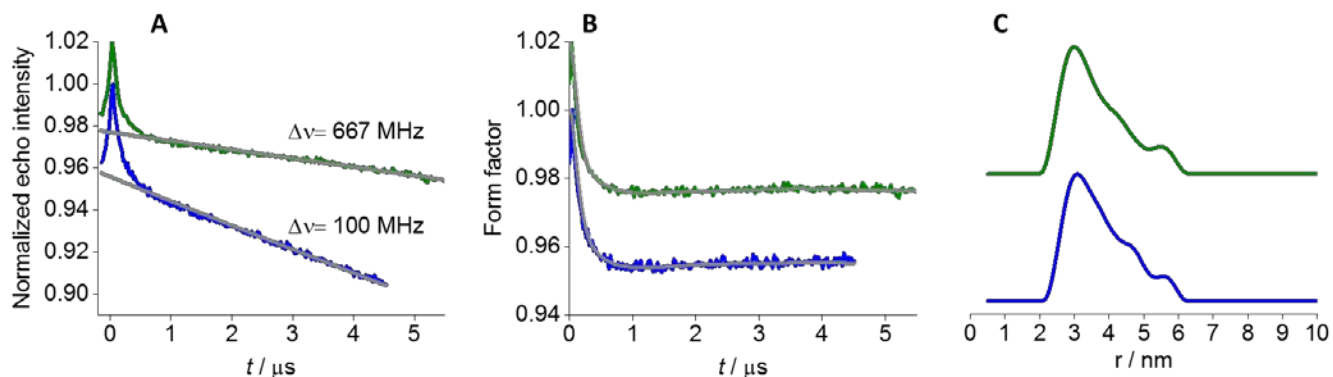


**Figure S12.** Comparison of the DEER traces of the Zika virus NS2B-NS3 protease mutant S85*C/T27C labeled with **C9** and without inhibitor recorded with two values of the $\Delta v$ parameter as noted in the figure. (A) Primary DEER data with the fitted background decay. (B) Corresponding form factor with the fitted data (in grey) obtained with the distance distributions shown in (C).

**Modelling of DEER distance distributions**

## References

1 M. D. Lee, C.-T. Loh, J. Shin, S. Chhabra, M. L. Dennis, G. Otting, J. D. Swarbrick and B. Graham, *Chem. Sci.*, 2015, **6**, 2614–2624.

2 S. Stoll and A. Schweiger, *J. Magn. Reson.*, 2006, **178**, 42–55.

3 A. M. Raitsimring, A. V. Astashkin, O. G. Poluektov and P. Caravan, *Appl. Magn. Reson.*, 2005, **28**, 281–295.

4 A. M. Raitsimring, A. V. Astashkin and P. Caravan, Highfrequency EPR and ENDOR characterization of MRI contrast agents, in Biological Magnetic Resonance, ed. L. Berliner and G. Hanson, Springer, New York, 2009, pp. 581–621.

5 N. Manukovsky, A. Feintuch, I. Kuprov and D. Goldfarb, *J. Chem. Phys.*, 2017, **147**, 044201.

6 A. Dalaloyan, M. Qi, S. Ruthstein, S. Vega, A. Godt, A. Feintuch and D. Goldfarb, *Phys. Chem. Chem. Phys.*, 2015, **17**, 18464–18476.

7 M. R. Cohen, V. Frydman, P. Milko, M.A. Iron, E.H. Abdelkader, M. D. Lee, J. D. Swarbrick, A. Raitsimring, G. Otting, B. Graham, A. Feintuch and D. Goldfarb, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12847–12859.

8 M. Stanton-Cook, X.-C. Su, G. Otting and T. Huber, http://comp-bio.anu.edu.au/mscook/PPT/, accessed February 14, 2018.

9 A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting and D. Goldfarb, *J. Am. Chem. Soc.*, 2010, **132**, 9040–9048.

10 M. Tropiano, C. J. Record, E. Morris, H. S. Rai, C. Allain and S. Faulkner, *Organometallics*, 2012, **31**, 5673–5676.

11 D. Parker, H. Puschmann, A. S. Batsanov and K. Senanayake, *Inorg. Chem.*, 2003, **42**, 8646–8651.