Putting a spin on it: amyloid aggregation from oligomers to fibrils

Zurlo, E.

5 A two-armed probe for in-cell DEER measurements on protein

The double electron-electron resonance (DEER) technique combined with site-directed spin labeling (SDSL) is a popular method for bio-molecular structure studies. The application of DEER to protein systems within living cells puts rigorous restraints on the spin-label. The probe needs to be immobilized relative to the protein and both the linkage and the paramagnetic center need to be stable under the reducing conditions of the cell. Here, three two-armed Gd(III) complexes, Gd(III)-CLaNP13a/b/c, which differ in the length of the arms that links the Gd(III) chelating cage to the protein, were synthesized. Rather than the disulphide linkage employed for protein attachment in most other CLaNP molecules, thioether linkage is used here. We doubly labeled the N55C/V57C/K147C/T151C variant of T4Lysozyme and performed paramagnetic 95 GHz DEER on these constructs. DEER experiments for spin-labeled variants of T4 lysozyme were performed in buffer, in cell lysate and in *D. discoideum* cells. A narrow distance distribution was detected for all constructs. The DEER derived distances are on the order of 4.5 nm and are in agreement with the expected values based on the metal ion positions derived from paramagnetic NMR results. The results indicate that the probes are mostly rigid relative to the protein due to the dual attachment sites. The thioether linkages make the probes suitable for in-cell protein studies.

Chapter 5

5.1 Introduction

The studies about the structure of proteins are most commonly performed in vitro on isolated and purified protein samples. However, the proteins in a natural environment are subjected to much more complex conditions, for example due to the interaction with a vast range of other molecules and proteins or under other factors that are strongly dissimilar from an aqueous solution. Therefore, recent studies are focused into finding alternative ways to investigate the activity of the proteins in environments which resemble the natural ones, for example in cell lysate or directly within a cell. Electron Paramagnetic Resonance (EPR) is a technique which has seen an increase of interest due its ability to study these systems under such conditions[56,162,163].

Double electron-electron resonance (DEER) spectroscopy, also known as pulsed electron-electron double resonance (PELDOR), is a pulsed EPR method that measures, through the dipolar electron–electron coupling between two paramagnetic species, distances in the range of 1.5–8 nm, and up to 16 nm under special perdeuterated conditions[55,58,163–166]. This technique is characterized by high precision and reliability, and beyond mere mean distances, DEER yields distance distributions, which provide information about conformational distributions and dynamics. Because most cell components are diamagnetic, EPR-based distance determination inside cells is virtually background-free. DEER spectroscopy is the commonly used technique to perform such *in-cell* distance determination[167–170]. It must be taken into account that in most cases the proteins of interest are also diamagnetic, thus it is important to find suitable tags to be added to perform EPR.

Conventionally, DEER measurements are performed between two nitroxide spin labels ($m_s = \pm 1/2$) that have been attached to biological molecules either by site-directed spin labeling or by chemical modification, however such labels have problems due to their reactivity and the strongly reducing environment to which they are subjected in natural cells[167]. Therefore, in recent years, research was focused into developing new probes where both the spin-label itself and the bond linking the probe to the protein are resistant to reduction and as rigid as possible.

Complexes of Gd(III) ($S = 7/2$) with chelating ligands are ideal candidates for in-cell EPR[171,172] at high frequencies (e.g., 95 GHz), and feature low toxicity, high stability, suitable spin relaxation times and no orientation selection. The most efficient Gd(III) spin label should exhibit a chemically stable conjugation to the protein in biological environment, a high binding constant to the metal ion to prevent leakage of Gd(III) from the complex especially in *in-cell* environments, a narrow central transition ($m_s = \pm 1/2$) and a long phase memory time for sensitivity in relation to long-distance measurements, and a rigid ligand to bind it to the protein for high-distance resolution. Considering the last point, the labeling site should be carefully chosen to ensure that the tag efficiently attaches to the protein and its rigidity does not interfere with the protein structure and

functionality. Previous research has shown that in a maleimide functional group conjugated to a cysteine residue the C-S bond between the cysteine and the maleimide group, unlike a commonly used disulphide bridge, was resistant to reduction[168,169], although the flexibility of the linker gave rise to a wide distance distribution. Different strategies have been applied to reduce the mobility of the tag, such as using more rigid attachment groups or attaching the probe via two arms linkers to the protein.[173–175]

In this study we investigated the properties of three recently developed Gd(III) labeling tags, Gd-CLaNP13i (i = a,b,c) attached to the Bacteriophage T4-Lysozyme mutant T4Lys N55C/V57C/K147C/T151C. Such tags have the characteristic of using the maleimide group for attachment to the protein and of being equipped with two arms clamping the Gd(III) to achieve higher rigidity.

In the present work, we demonstrate the excellent performance of these new tags in in-vitro, cell lysate, and *in-cell* DEER measurements. The experiments showed a stable conjugation of the labels to the protein, a high binding constant of metal even under *in-cell* conditions, and high rigidity of the clamp resulting in narrow distances distributions which agrees with expected distances.

5.2 Materials and methods

1-(2-aminoethyl)-1H-pyrrole-2,5-dione, 1-(3-aminoethyl)-1H-pyrrole-2,5-dione, 1-(4-aminoethyl)-1H-pyrrole-2,5-dione and cyclen were purchased from Abosyn Chemical Inc. and CheMatech. $Gd(OAc)_3{\cdot}4H_2O$, $Yb(OAc)_3{\cdot}4H_2O$, N-(tert-butoxycarbonyloxy)succinimide and methyl 2-bromoacetate were purchased and used without further purification. Solvents were purchased from Honeywell, BIOSOLVE or Aldrich and directly used for synthesis. ATTO-647 maleimide was obtained from ATTO-TEC GmbH. Superdex 75 columns and Sephadex G-25 PD10 desalting columns were purchased from GE Healthcare. Reactions were followed by liquid chromatography-mass spectrometry (LC-MS), TLC analysis on silica gel (F 1500 LS 254 Schleicher and Schuell, Dassel, Germany) in which compounds were visualized by UV and/or ninhydrin, $KMnO_4$. Flash chromatography was performed on Screening Devices silica gel 60 (0.04-0.063 mm). A Waters preparative RP-HPLC system, equipped with a Waters C18-Xbridge 5 µm OBD (30 x 150 mm) column and an Äkta Basic FPLC (GE Healthcare Inc.) were used for purification. A LCQ liquid chromatrography mass spectrometry system and a Finnigan LTQ Orbitrap system were used for high-resolution mass spectrometry and protein conjugation analysis. A Thermo Scientific™ NanoDrop 2000 spectrophotometer was used for protein concentration measurements.

5.2.1 Protein expression and purification

T4Lys mutant K147C/T151C was produced as described in previous work.[176] The tetra-cysteine mutant N55C/ V57C/K147C/T151C was generated by the Quikchange method (Agilent). After confirming the successful mutations by DNA sequencing, the gene was expressed in *Escherichia coli* BL21 (DE3). Transformed cells were incubated overnight at 37°C on LB agar plates with kanamycin and chloramphenicol (100 µl/mL and 34 µl/mL, respectively). Single colonies were transferred to 2 mL LB medium with kanamycin and chloramphenicol and incubated at 37°C, 250 rpm for 6 h. The precultures were used to inoculate 50 mL (1:1000) for overnight incubation in minimal medium (M9, 37°C, 250 rpm). The 50 mL culture was diluted 100 times and incubated at 37 °C, 250 rpm. When the OD_{600} value reached 1, gene expression was induced by addition of 1 mM isopropyl β-D-thiogalactoside (IPTG). The temperature was reduced to 30 °C and the cells were harvested 18h after induction by centrifugation. The T4Lys N55C/V57C/K147C/T151C was purified according to the method described by Georgieva et al.[177] with some modifications. Cells were lysed using a French press and cleared by centrifugation (45 min, 2500 rpm). The supernatant was loaded on a 5 mL HiTrap CM column equilibrated with 25 mM Tris pH 7.5. The column was washed with 0-40% linear gradient of buffer B (25 mM Tris pH 7.5, 500 mM NaCl). T4Lys N55C/V57C/K147C/ T151C was eluted in one step with 100 % buffer B.

5.2.2 DEER sample preparation

The samples for in vitro and in lysate DEER measurements contained 100-150 µM of Gd(III)-CLaNP13 labeled N55C/V57C/K147C/ T151C T4Lys in 20 mM sodium phosphate, pH = 5.5, 150 mM NaCl, 20% (*v*/*v*) glycerol, *E. coli* cell lysate (20% (*v*/*v*) glycol) and *D. discoideum* cell lysate (20% (*v*/*v*) glycol). The sample for in-cell DEER measurement was obtained by incubating *D. discoideum* medium with the protein labelled with Gd(III)-CLaNP13b at a concentration of 115 µM for 90 mins. The cells were recovered, and washed three times in medium to remove excess protein, and to concentrate the cell suspension. The sample was cooled to approx. 4 °C on ice, and 10% (*v*/*v*) DMSO was added to the medium to prepare the cells for freezing and EPR measurement.

5.2.3 EPR measurements conditions

The 95 GHz DEER measurements were recorded using an ELEXSYS E680 spectrometer (Bruker, Rheinstetten, Germany) with a home-built probe head. The measurements were done at a temperature of 10 K. The DEER spectra were recorded using the standard four pulse DEER sequence[178]. The pump pulse duration was 40 ns and the observer pulses durations were 32 and 64 ns, respectively. The pump pulse power was adjusted to invert the echo maximally. The separation between the pump and observer frequencies (Δν) was 60 MHz, with the pump pulse adjusted to irradiate at the maximum of the EPR spectrum. To make optimum use of the resonator bandwidth, the pump and observer frequencies were set to 30 MHz higher and lower frequencies respectively with respect to the resonance frequency of resonator. The full sequence for the observer was π/2–τ–π –Δt- π –echo with a delay time τ of 360 ns and time steps for DEER evolution of 8 ns. The DEER data were analyzed with the program DeerAnalysis 2018.[56]

Since the modulation depth depends strongly on the precise setup of the experiment, e.g. choice of pump and observer frequency relative to the resonance frequency of the EPR resonator, this parameter can differ from experiment to experiment on the order of 25%. The *in-cell* DEER experiment had been setup for a higher signal to noise ratio rather than for higher modulation depth due to instrumental issues. Also it must be noted that the experiment lasted less than 8 hours, instead in most cases in-cell measurements are taken for a longer period of time[179–181].

5.3 Results

5.3.1 Design and synthesis of Gd(III)-CLaNP13.

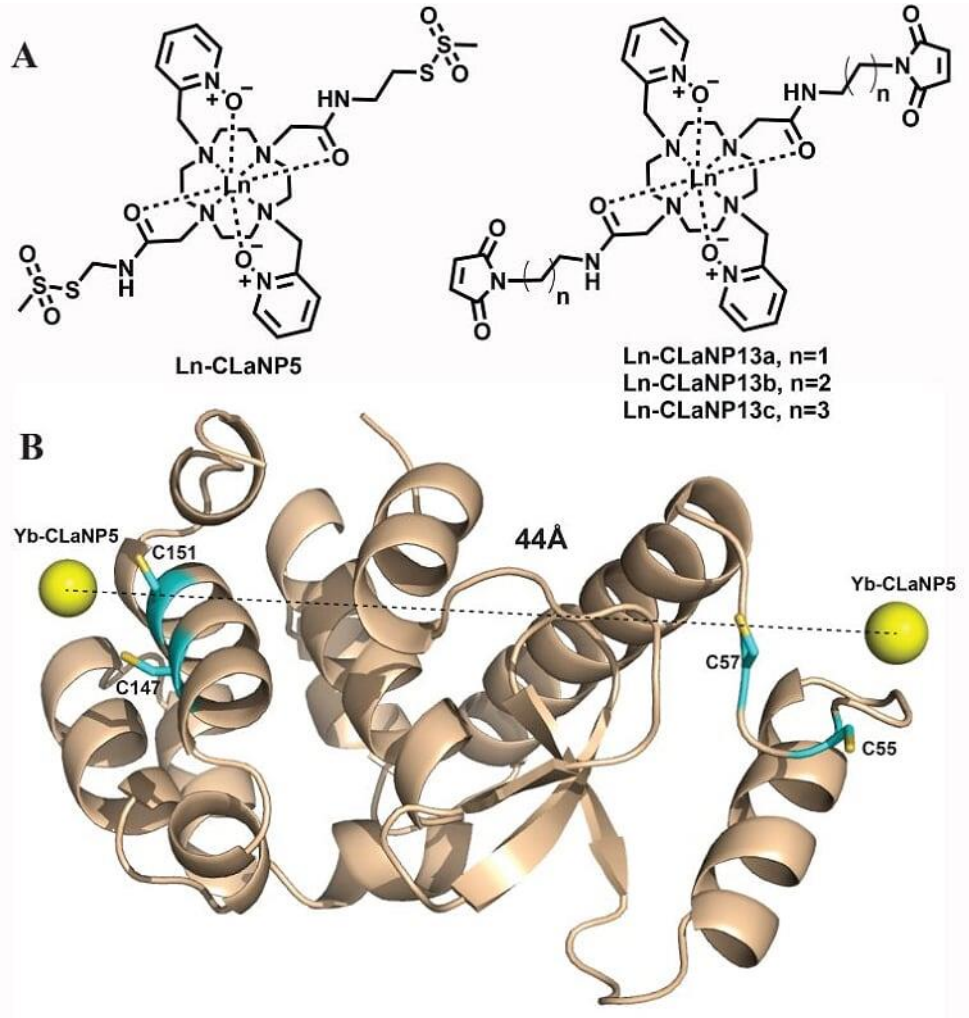

Figure 5.1 *Visual representation of the Gd(III) tags from the chemical point of view and in protein. A) Structures of Ln(III)-CLaNP5 and Ln(III)-CLaNP13; B) Model of the structure of T4Lys based on PDB entry 3dke*[182] *with two Cys pairs for the attachment of two probes. The positions of the metals are based on PCS analysis using Yb(III)-CLaNP5 as a paramagnetic probe. The backbone is drawn in ribbon representation. The Cys residues used for attachment have been modelled into the structure and are shown in sticks. The metal ions are shown as yellow spheres.*

The caged lanthanoid NMR probe CLaNP5 is a well-studied two-armed Ln(III) probe for paramagnetic NMR spectroscopy on proteins. The cyclen based molecule is equipped with two pyridine N-oxide coordination arms that reduce the arm rotation (Figure 5.1). Using CLaNP5 as a building block, Gd-CLaNP13 was designed, in which the arms for protein attachment were functionalized with maleimide groups. The length of the spacer was varied from 2 to 4 methylene groups (Figure 5.1). Maleimide can readily and

specifically react with the thiolate group of a cysteine side chain, forming a carbon-sulfur bond, which is not prone to reduction.

5.3.2 Protein labeling

A ^{15}N enriched variant of T4 lysozyme (T4Lys) with the substitutions K147C/T151C was used to determine optimal conditions for protein labeling on the basis of LC-MS and NMR results. The best result for double maleimide reactions with the cysteines was using a buffer with pH 7.8, at 4 ^{o}C for 6 h. To test whether free thiolates remained after the reaction, iodoacetamide was added. The LC-MS results, obtained by Dr. Bogdan Florea, yielded masses that match those expected for probes attached via two arms, assuming that one or two water molecules remained bound to the protein-probe complex. No peaks with additional mass of 58 Da were detected, which would be expected for protein with one-arm attached probe and an additional acetamide group linked to the second cysteine sulfur atom, and also no free protein was detected, suggesting that the protein labelling efficiency was more than 95%. In the MS spectrum of the CLaNP13a, a very small peak (20552 Da) is observed that represents T4Lys with two probes bound, each via a single arm (Figure S5.1). After attachment of the first maleimide group, the reaction with the second will generally be efficient because it is intramolecular. The limited length of the linker in CLaNP13a may allow for some competition with the second-order reaction of the protein-probe complex with a second probe molecule. However, according to the MS spectrum (Figure S5.2), the fraction of this species is not easy to detect.

5.3.3 Results of EPR experiments

For the EPR experiments, the quadruple cysteine mutant T4Lys N55C/V57C/K147C/T151C was labeled with Gd(III)-CLaNP13, variants a, b or c (Figure 5.1A), the resulting constructs are referenced to as Gd13iT4L with i:a,b,c. In the following we will describe the experimental results. We refer to investigations of the protein constructs in buffer as “in vitro”.

The echo detected EPR spectrum of the Gd13bT4L is shown as an inset in Figure 5.2 and in Figure S5.4. All three constructs have similar spectra. Specifically, they consist of a central narrow line due to the $m_s = -½ \leftrightarrow +½$ transition that is superimposed on a broad background due to all other transitions. The width of the central transitions for the different tags are shown in Table 5.1, and it can be noticed that Gd13bT4L has the narrowest central transition of the three.

Table 5.1 *Characteristics of EPR spectra of the Gd constructs used. Given is the full-width at half maximum (FWHM) from the central transition ($m_S \pm ½$) in the field-swept electron-spin echo (FSESE) of the Gd-CLaNP13a,b,c labeled proteins.*

Sample	MHz
Gd13aT4L	144.63 ± 1.91
Gd13bT4L	113.42 ± 4.41
Gd13cT4L	135.94 ± 2.42

5.3.4 Results of distance measurements

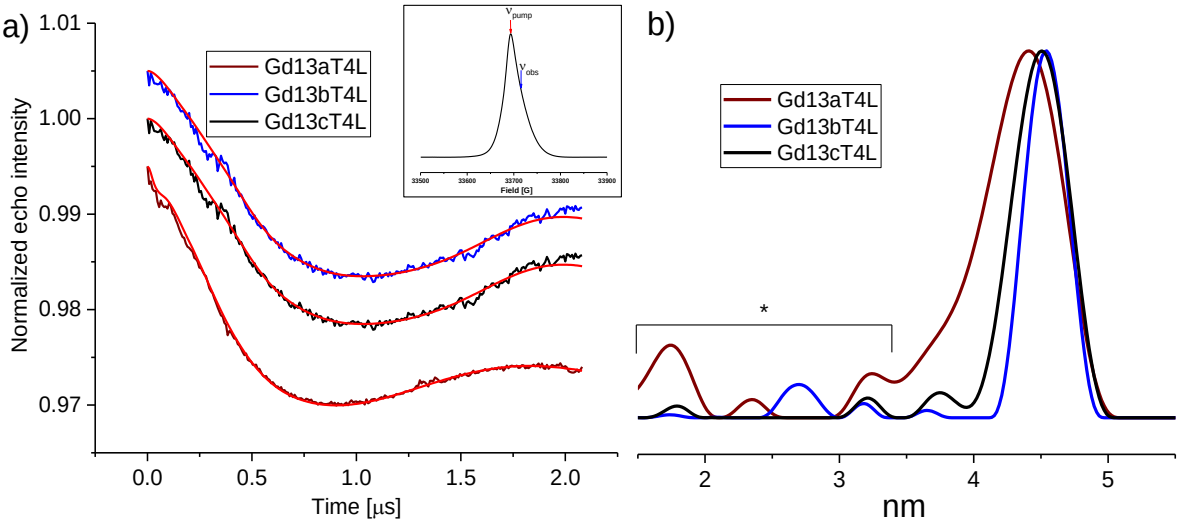

Figure 5.2 *DEER data of T4Lysozyme N55C/V57C/K147C/T151C labeled with Gd-CLaNP13a,b,c to form Gd13aT4L, Gd13bT4L, Gd13cT4L respectively. a) Background corrected DEER traces. Traces are shifted vertically for clarity. Measurements were performed at 10 K for 6 to 12 hours. Red lines: fits obtained with the distance distributions shown in b) obtained after Tikhonov regularization (α = 1000). Peaks marked with an asterisk do not contribute significantly to the data, as determined by the DeerAnalysis suppression tool*[56]*. Inset: 95 GHz field-swept electron-spin echo spectrum (FSESE) of the central transition region, position of the pump and observer frequencies are shown.*

The DEER data of all three Gd13iT4L (i = a.b.c) constructs are depicted in Figure 5.2. Figure 5.2a shows the data after background removal and Figure 5.2b the derived distance distributions. The raw DEER data are shown in Figure S5.5. The distances in all three constructs are close to 4.5 nm with Gd13aT4L having the shortest distance at 4.41 nm.

Table 5.2 *Distances between Gd-atoms in Gd13iT4L (i = a,b,c). Given is the maximum of the distance distribution and the full-width at half maximum (FWHM) of the distance distribution obtained with Tikhonov regularization (α = 1000) of the Gd-CLaNP13a,b,c labeled proteins*

Sample	Distance [nm]	Width [nm][a]
Gd13aT4L	4.41 ± 0.03	0.67 ± 0.20
Gd13bT4L	4.54 ± 0.02	0.39 ± 0.08
Gd13cT4L	4.51 ± 0.02	0.50 ± 0.11

[a] *The full-width at half maximum (FWHM)*

The modulation depth, in the order of 2 %, is typical for 95 GHz DEER on Gd (III) samples: Usually depths between 2 % and 5 % are observed.[167]

The stability of the label in more complex environmental conditions was checked by incubating Gd13iT4L in *Escherichia coli* lysate for a total time of 24 hours. The DEER traces are similar and the distance distributions are identical within the noise to the in vitro samples (see Figure S5.6). No systematic decay of the modulation depth was observed over the period of 24 hours. Since the uncertainty in the modulation depth is in the order of 25%, we cannot exclude that a decay in that order occurs over time, even though we do not find systematic changes in modulation depth.

5.3.5 In-cell DEER

To investigate whether the label is stable in the cell, we measured DEER of Gd13bT4L in *Dictyostelium discoideum (D.discoideum)* cells. Fluorescence microscopy on an ATTO-647-maleimide tagged T4Lys K147C/T151C variant shows that the protein enters the cells, and that protein outside the cell was efficiently removed by washing with PBS. The bright fluorescent spots observed within the cell indicate that the protein is likely to be contained within vesicles such as endosomes. The DEER trace in Figure 5.3, therefore, results from Gd13bT4L incorporated into the cells. Cells thawed after the DEER experiments were shown to be viable by live-cell microscopy (see SI)

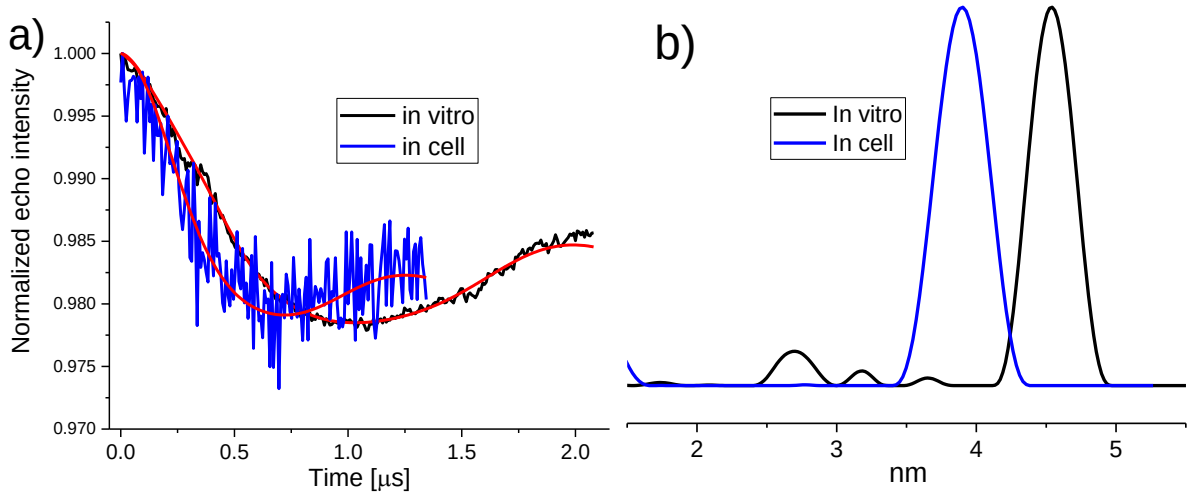

Figure 5.3 *Comparison of DEER traces of the protein in vitro and in Dictyostelium discoideum (Dicty) cells for the Gd-CLaNP13b tagged T4Lys N55C/V57C/ K147C/T151C in vitro (black) and in Dicty cells (blue). a) Background corrected DEER traces. Red lines: fits obtained with the distance-distributions calculation shown in b) obtained by Tikhonov regularization (α = 1000)*

The trace in Figure 5.3a has a clear DEER modulation, with a minimum at 0.6 µs, which is also visible in the raw data (Figure S5.5d). Also, the width of the distribution is clearly in agreement with a fully intact doubly linked CLaNP. This proves that the Gd-CLaNP13 is stable in a cellular environment. The maximum of the distance distribution (Figure 5.3b) is at a shorter distance, 3.9 nm, than observed for Gd13bT4L in vitro and in *E. coli* cell lysates (see Figure S5.6). Further DEER experiments on Gd13bT4L in *D.discoideum* cell lysate and medium (for details, see SI) resulted in distance distributions that are dominated by a distance of 4.5 nm (see Figure S5.7). The possible origins of the shorter distance between the two Gd(III) ions in Gd13bT4L in *D.discoideum* cells compared to the one in vitro are discussed below.

5.4 Discussion

Here we report the synthesis of double-armed, rigid CLaNP tags linked by maleimide linkers to a protein to generate a Gd(III) spin label that is stable under in-cell conditions. Having two arms and a rigid CLaNP design should further improve the accuracy of DEER distance measurements. Three tags were obtained, Gd-CLaNP13a,b,c, which were synthesized in good yields and had high labelling efficiencies when attached to the protein T4L. All tags show clear DEER modulations with the expected modulation depth, confirming the reliable double labelling of the protein, inferred from mass spectrometry. Partial labelling by only one tag/protein would reduce the modulation depth, tags attached by only one arm should lead to broader DEER distance distributions and less pronounced modulation, neither of which is observed to any significant degree. The differences in distances for the three linkers are not monotonous with the linker length, suggesting that the linker takes on particular conformations, or that Gd(III) interactions with the protein surface could differ for the three linkers, leading to different distances. All distances (Table 5.2) agree well with the distance of 4.4 nm inferred from paramagnetic NMR data using the CLaNP-5 probes as mimics of CLaNP-13 (see Figure 5.1). Both the DEER time traces and distance distributions are similar for the three Gd13iT4L (i = a, b, c) tags. (Figure 5.2). The width of the distance distribution, albeit small, is not exceptionally small considering the results of Gd(III) DEER experiments performed on proteins with singly linked probes.[167,173,181,183] Perhaps part of the width of the distribution observed in the present study is due to a distribution of conformations of the protein loop to which the Gd(III) at residues 55 and 57 is attached.

Having thus established that the constructs show the expected properties in vitro, we proceeded to study their resistance to cellular environments. In *E. coli* lysate, over a period of 24 h, no deterioration was detected within experimental limits, placing an upper limit of any possible decay at 25 % (see SI), a value that is largely determined by the experimental uncertainty. Prompted by the stability of the tags Gd13bT4L both in vitro and in *E. coli* lysate, in cell measurements were performed. *D. discoideum* was selected, because it is known for its high uptake of extracellular components.[184] The uptake was verified by fluorescence microscopy (see SI). The protein appears to be concentrated inside small vesicles, and there the estimated concentration is around 5 µM. The cells were shown to be viable after DEER experiments (see SI). The DEER results of Gd13bT4L in cells are promising: The modulation depth of 1.5 % is smaller than observed in vitro and in *E. coli* lysate, but the difference is close to the error margins of the data (Figure 5.3a). We attribute the lower signal-to-noise ratio of the DEER trace of Gd13bT4L in the *D.discoideum* cells (Figure 5.3a) to the lower protein concentration and the 40 % shorter accumulation time compared to in vitro experiments. The distance distribution (Figure 5.3b) has a width that is similar to that of Gd13bT4L *in vitro*, showing that the

protein has a well-defined structure, and that the spin label remains bound. The maximum of the distance distribution for the in-cell DEER measurement is 3.90 nm, which is smaller than the 4.54 nm for the in vitro sample. So, while this initial experiment confirms that the probe is stable in cells, the difference in distance is puzzling. T4Lyzozyme is known to undergo conformational changes, and by FRET experiments, distance changes in the order observed here, around 0.5 nm, were reported.[185] Also, a number of protein structures of T4L in different states are known. Nevertheless, we did not yet find any studies that match the distance of Gd13bT4L measured in *D.discoideum*. Initial experiments exposing Gd13bT4L to *D.discoideum* lysate and medium components did not pinpoint the origin of the distance change between in cell and in vitro. The protein is clearly not evenly distributed over the cytosol but rather is present in vesicles like endosomes, in which the local conditions, such as pH, could differ from those used in vitro. Although specific changes in structure of proteins in the cell have been reported before[186] and they are ultimately the motivation for in-cell work, for the present set of experiments it is too early to speculate on such an effect.

In conclusion, a new set of rigid Gd(III)-CLaNPs has been synthesized that perform well in a cellular environment, showing that the Gd(III)-CLaNP13a,b,c are promising new candidates for in-cell Gd(III) work.

5.5 Supporting information

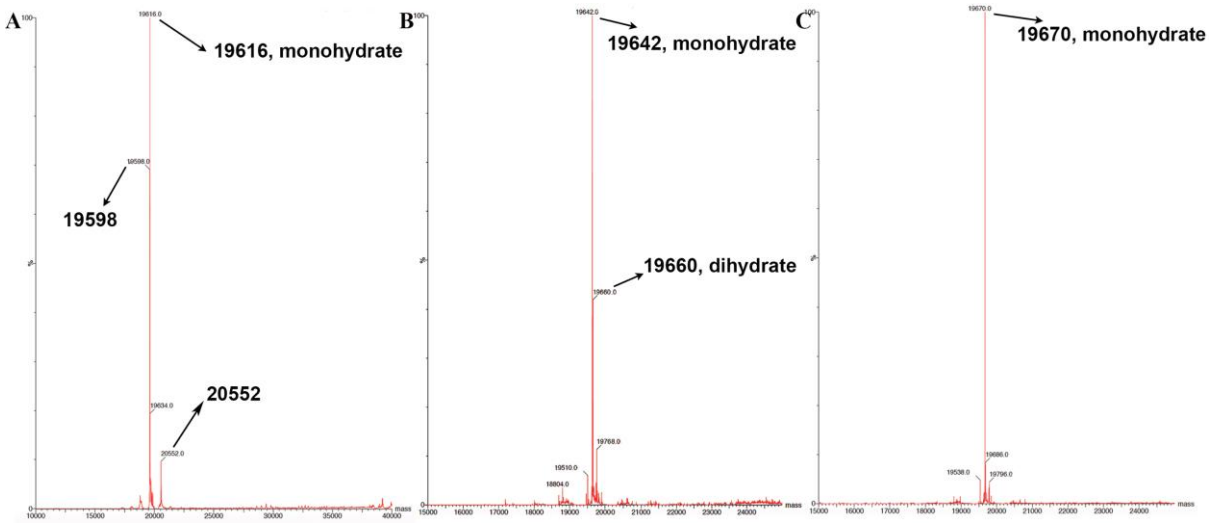

Figure S5.1 *ESI-TOF MS spectra of CLaNP13a (**A**), CLaNP13b (**B**) and CLaNP13c (**C**) linked to ^{15}N enriched T4lys T147C/N151C mutant.*

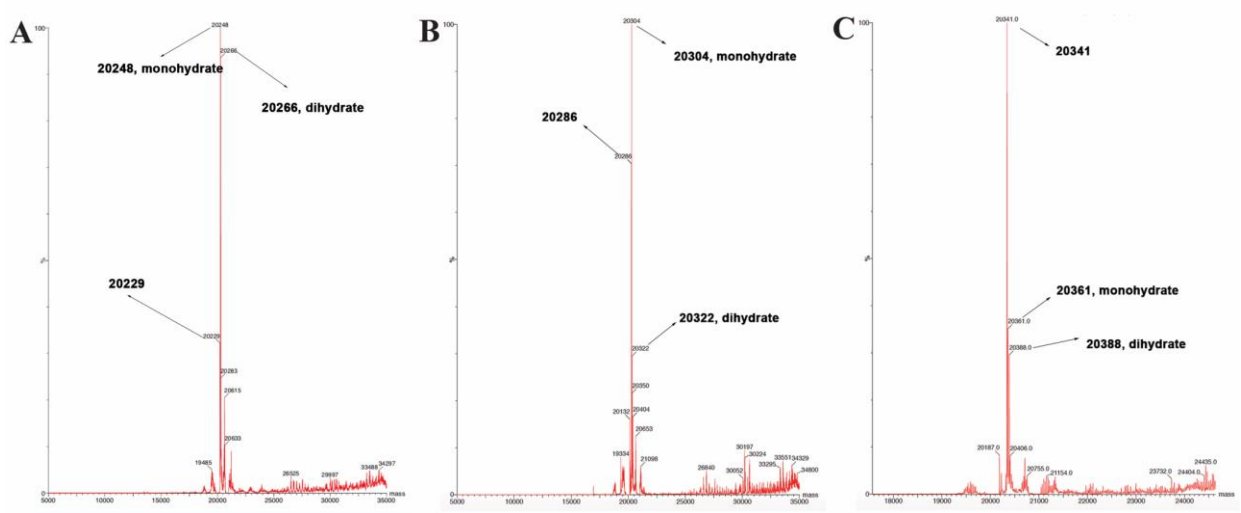

Figure S5.2 ESI-TOF MS spectra of CLaNP13a (**A**), CLaNP13b (**B**) and CLaNP13c (**C**) linked to T4Lys N55C/V57C/T147C/N151C mutant.

5.5.1 Cell culture and live cell imaging

Axenic *D. discoideum* (AX2) was obtained from Dr. Günther Gerisch (MPI for Biochemistry, Martinsried, Germany). Cells were grown at 20 °C in HL5 medium and cultured in 100 mm Petridishes (TC-treated culture dish, Corning, USA) and confluency was kept below 70%. For microscopy experiments, cells were harvested and centrifuged at 1500 rpm for 3 min followed by three successive washing steps of the cellular pellet with 17 mM K-Na-phosphate buffered saline (PBS), adjusted to pH = 6.0. After resuspension in PBS, the cells were pipetted into a 70 µL well inside a 35 mm imaging dish (Insert and dish, Ibidi GmbH, Martinsried, Germany) and left to adhere for 30 min. While imaging, T4Lys K147C/T151C linked to ATTO-647 (ATTO-TEC GmbH) was added to a final concentration of 100 µM. Cells were left to incubate for 60 min and were imaged every 30 seconds in two channels (BF and 647) with a Nikon Eclipse Ti microscope, equipped with a Yokogawa confocal spinning disk unit operated at 10,000 rpm (Nikon, Tokyo, Japan). ATTO-647 was excited by a 647 nm solid state diode laser (Coherent, Santa Clara, U.S.A.), supported in an Agilent MLC4 unit (Agilent Technologies, Santa Clara, U.S.A.). Images were captured (50 ms) by an Andor iXon Ultra 897 High Speed EM-CCD (Andor Technology, Belfast, Northern Ireland) through a Cy5 HYQ filter (Nikon, Tokyo, Japan). After incubation, the well was carefully washed with PBS five times (50 µL each step) and imaged for another 15 min.

5.5.2 Quantification of the in-cell T4-ATTO-647 concentration

Known concentrations of T4Lys K147C/T151C linked to ATTO-647 in PBS were imaged in the same condition as described for cells. The data were averaged over all images and pixels to find mean intensities per concentration. For the cell data, all pixels belonging to each cell were determined using an in-house Matlab (The Mathworks, Inc., Natick, MA, U.S.A.) script for cell edge recognition. In-cell pixels were averaged and subsequently compared to the concentration calibration measurements (see Figure S5.3) for two laser powers. All images were pre-processed by flat-field correction and dark-field subtraction.

LC-MS results showed that T4Lys K147C/T151C was tagged with only a single ATTO-647-maleimide per protein molecule. *D. discoideum* is known for its high uptake of extracellular components[184,187]. After 30 min incubation, almost all cells have taken up fluorescent protein and it was present in cell body in vesicles (Figure S5.3). The protein was washed away from the medium after 60 min, but the fluorescent protein can still be detected inside the cells and the concentration was estimated to be 5 µM. The protein can still be detected in 3 h after the start of the incubation (Figure S5.3).

Similar experiments were conducted with double CLaNP13b tagged T4Lys for in-cell DEER experiments. After incubation, the cells were washed, concentrated and cooled on

ice. Before the cells were frozen for EPR measurement, 10% (*v*/*v*) DMSO was added to the medium. DEER experiments were performed at 10 K for 48 h.

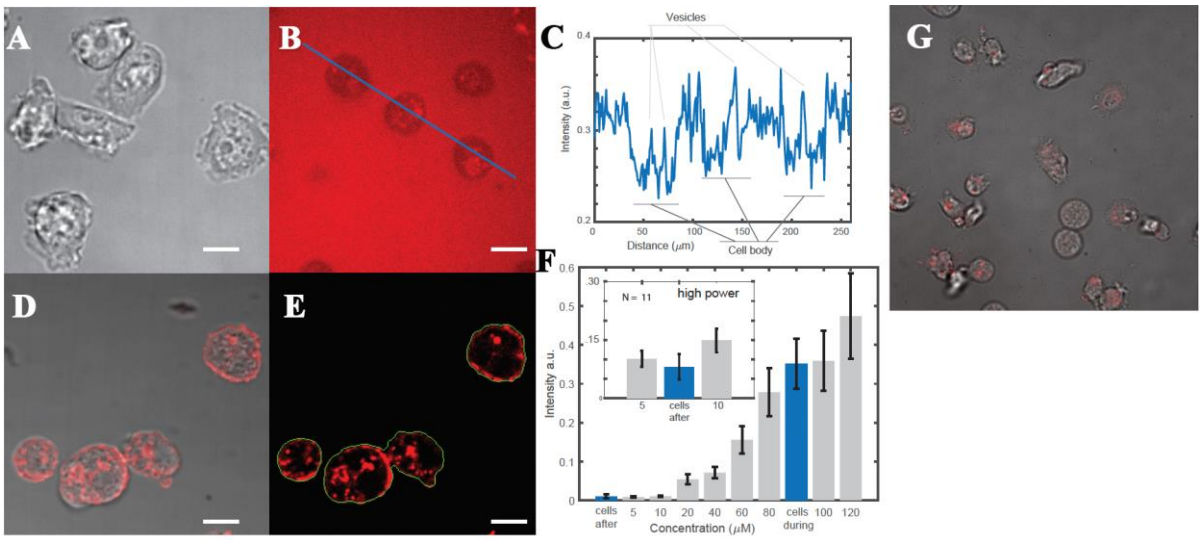

Figure S.5.3 Different stages of the internalization process of T4-ATTO-647 in vegetative *D. discoideum*. **A**) Just prior (t = 0 min) to incubation, cells are adhered to the substrate; **B**). Incubation and subsequent internalization results in a shock response and cell rounding. Confocal fluorescence image (red) shows cells after t = 30 min of incubation. Blue line marks the direction of intensity profile plot; **C**)**.** Intensity profile of three cells during incubation, the protein is ubiquitous in cell body and surrounding medium; **D**). After washing with PBS (t = 60 min) the cells recover and start spreading. T4-ATTO-647 remains in the cell after washing with phosphate buffered saline (t = 75 min); **E**). To determine the protein concentration inside cells after incubation, cell edges are recognized (green) and mean fluorescence (red) intensities measured; **F**). Mean intensity histogram of various concentrations of T4-ATTO-647 in PBS (grey) versus mean fluorescence inside cells (blue) during and after incubation, measurements were repeated with higher power (inset); **G**). State of the cells after 3 hours of incubation

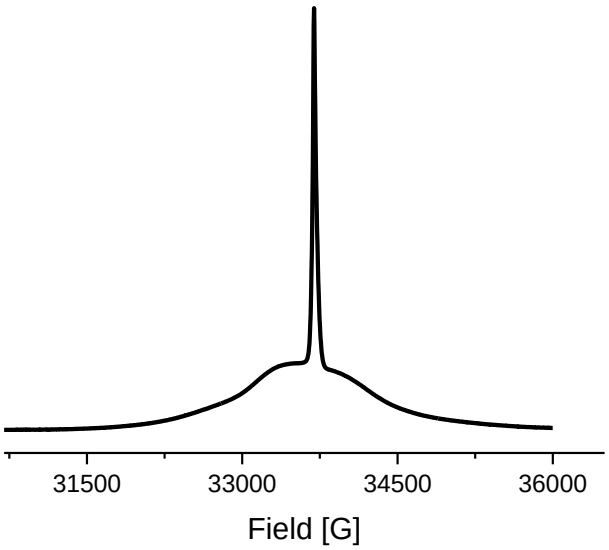

Figure S5.4 EPR of samples in vitro. Shown is a Field-Swept Electron-Spin Echo (FSESE) spectrum at 95 GHz obtained at 10 K of Gd-CLaNP13b tagged T4Lys N55C/V57C/ K147C/T151C (200 μM). The measurement was carried out by a 2-pulse {π/2-τ-π-τ-echo} sequence with t(π/2) = 32 ns, t(π) = 64 ns. and τ = 360 ns.

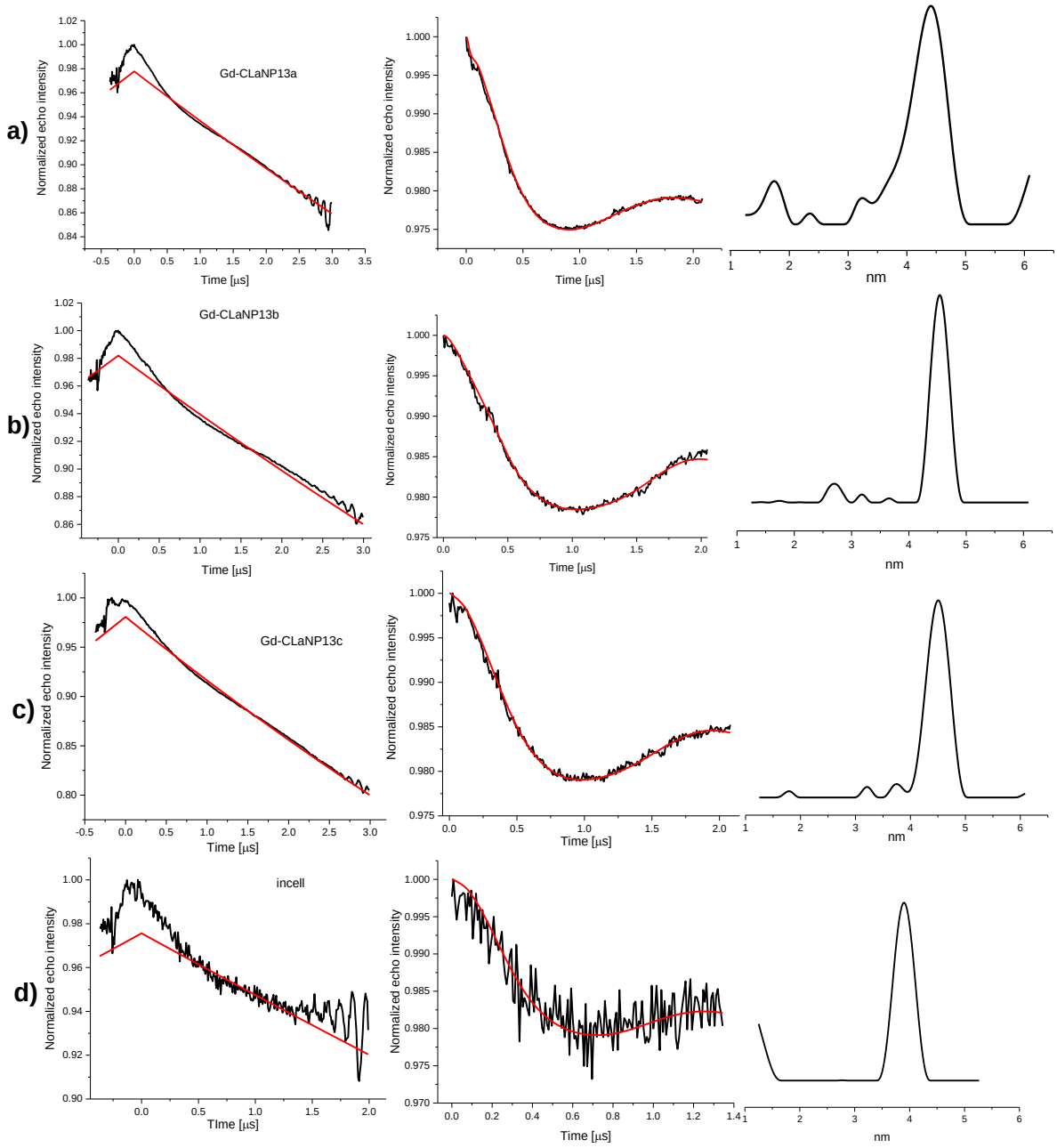

Figure S5.5 Four-pulse DEER results obtained for Gd13iT4L (with i = a,b,c) at 10 K. On the left: Normalized DEER traces fitted with appropriate background decay (in red). Middle: Same DEER traces after background removal along with the fits obtained by Tikhonov regularization (red). On the right: Distance distribution obtained by Tikhonov regularization (α = 1000) in DEER Analysis 2018. a) In vitro Gd13aT4L; b) In vitro Gd13bT4L; c) In vitro Gd13cT4L; d) In-cell Gd13bT4L.

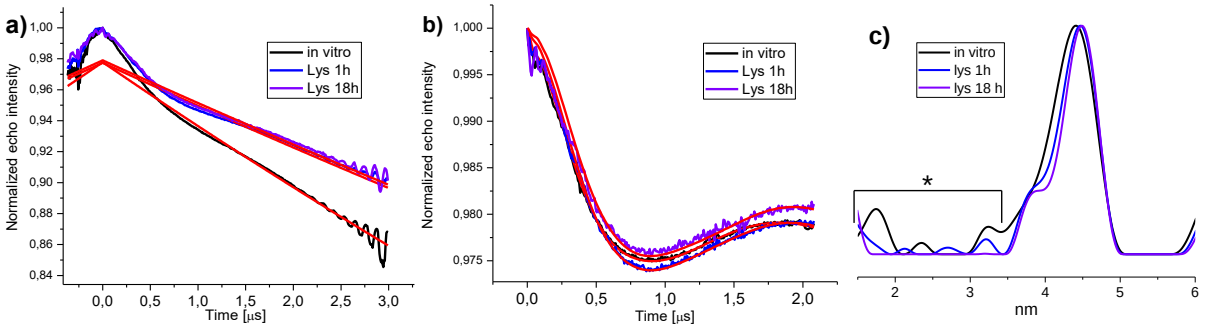

Figure S5.6. Four-pulse DEER traces and distance distributions for the Gd-CLaNP13a tagged T4Lys N55C/V57C/K147C/T151C samples in vitro (black), incubated in *E. coli* lysate for 1 hour (blue) and for 18 hours (violet). Measurements were done at 10 K. a) Normalized DEER traces fitted with appropriate background decay (red); b) Background corrected DEER traces. Red lines, fits obtained with the distance-distribution calculations shown in (c) obtained after Tikhonov regularization (α = 1000) in DEER Analysis 2018.[56] Peaks marked with an asterisk do not contribute significantly to the data.

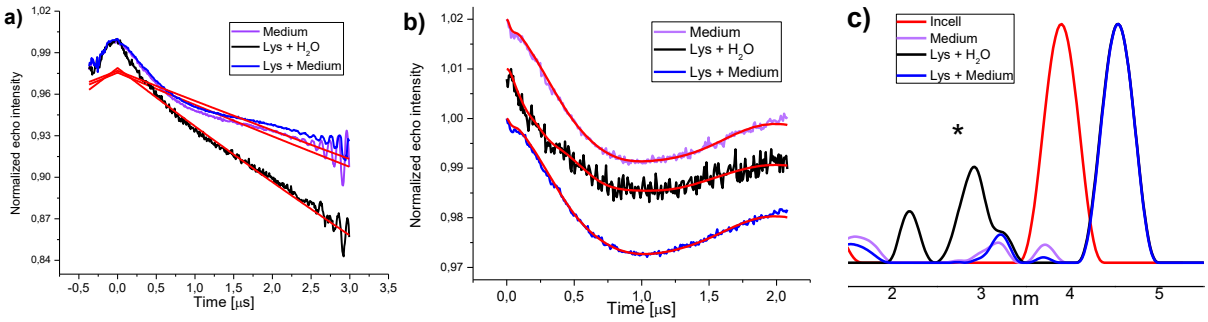

Figure S5.7. Four pulse DEER traces and distance distributions for the Gd-CLaNP13b tagged T4Lys N55C/V57C/K147C/T151C samples in *D. discoideum* cell lysate + milliQ H_2O (black), in medium (violet), in lysate + medium (blue). Measurements were done at 10 K. a) Normalized DEER traces fitted with appropriate background decay (red). b) The corresponding four pulse DEER traces along with the fits obtained by Tikhonov regularization (red). Traces are shifted vertically for clarity. c) Distance distributions obtained by Tikhonov regularization (α = 1000) in DEER Analysis 2018.[56] The distance distribution of the in-cell Gd13bT4L (red) has been added to better show the differences. Peaks marked with the asterisk do not contribute significantly to the data.