CHEMISTRY
A European Journal

A Journal of
ChemPubSoc Europe

Accepted Article

Title: Double-Arm Lanthanide Tags Deliver Narrow Gd3+-Gd3+ Distance Distributions in DEER Measurements

Authors: Adarshi Welegedara, Yin Yang, Michael Lee, James Swarbrick, Thomas Huber, Bim Graham, Daniella Goldfarb, and Gottfried Otting

To be cited as: *Chem. Eur. J.* 10.1002/chem.201702521

Link to VoR: http://dx.doi.org/10.1002/chem.201702521

Supported by
ACES

WILEY-VCH

10.1002/chem.201702521

WILEY-VCH

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

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

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

Introduction

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

The DEER experiment (also called PELDOR) is a pulse EPR experiment for measuring specific nanometer distances in proteins and macromolecular assemblies. Importantly, the experiment measures not only distances but entire distance distributions, yielding a unique tool to assess structural variability Analysis of the distance distribution and particularly its width is complicated by the fact that the width of the distribution reflects the structural variability of both the protein and the tag. Therefore, the important structural information encoded in the width of the distance distribution can be extracted only with a rigid tag that does not perturb the protein structure and allows a faithful prediction of the location of the spin label with respect to the protein backbone.

For molecules containing two Gd^{3+} ions, the DEER experiment measures the distance between the Gd^{3+} ions and can be conducted at W-band frequency (94 GHz), where it requires only small amounts (>0.09 nmols) of sample.[1,2] Measuring Gd^{3+}–Gd^{3+} distances is advantageous for the absence of orientation selection effects that complicate corresponding measurements with nitroxide tags at this magnetic field strength.[3-6] Moreover, the redox stability of Gd^{3+} makes it attractive for in-cell DEER distance measurements.[7-11] For Gd^{3+}–Gd^{3+} distance measurements in proteins, the Gd^{3+} ions must be introduced *via* suitable tag molecules that position a Gd^{3+} ion site-specifically in a defined location relative to the protein backbone. This is not trivial to achieve. For example, the cyclen-based **C1-Gd**,[12,13] **C9-Gd**[14] and **DO3MA-3BrPy-Gd**[11] tag have been attached to single cysteine residues leading to several rotatable bonds in the tether between Gd^{3+} ion and protein backbone and correspondingly uncertain positions of the Gd^{3+} ion, resulting in broadened distance distributions. Previously, a double-arm nitroxide tag attached to two cysteine residues has been shown to restrict the rotational freedom of the tag and to produce narrow distance distributions in DEER experiments.[15] The present work demonstrates the DEER performance of different double-arm strategies to immobilize Gd^{3+} ions.

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

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

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

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

The first three authors contributed equally.

Supporting information for this article is given via a link at the end of the document.

relatively close to the protein backbone, which could facilitate the interpretation of Gd^{3+}–Gd^{3+} distances in terms of protein structure.

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

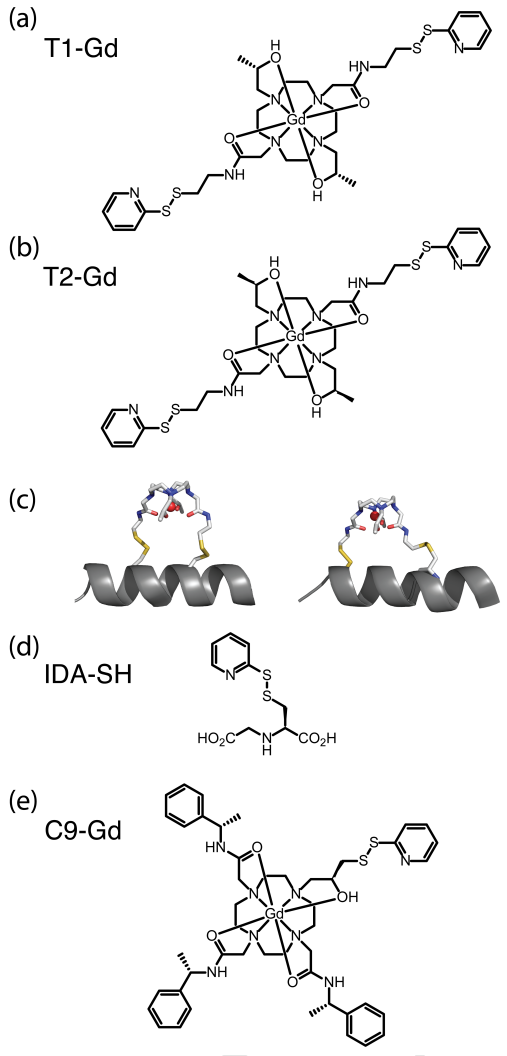


Figure 1. Structures of (a) **T1-Gd** tag (2-hydroxypropyl arms have *S* stereochemistry), (b) **T2-Gd** tag (2-hydroxypropyl arms have *R* stereochemistry), (c) **T2-Gd** tag modeled onto two cysteine residues in positions *i* and *i+4* (left) or positions *i* and *i+8* (right) of an α-helix, (d) the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

Results

Tagging strategies

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

Protein samples

Six mutants of T4 lysozyme were prepared. In all of them, the natural cysteine residues at positions 54 and 97 were mutated to threonine and alanine, respectively. These mutations are known to maintain the structure of the protein.[21] The additional mutations placed cysteine residues in α-helices. Two mutants put cysteine residues in positions 72 and 76, 127 and 131, and 131 and 135. In the following, we refer to the C54T/C97A/D72C/R76C/D127C/V131C mutant as T4L-A and the C54T/C97A/D72C/R76C/V131C/K135C mutant as T4L-B. The third mutant, referred to as T4L-C, positioned cysteines in positions 72 and 80, and 127 and 135, while positioning aspartic acid residues in positions 76 and 131 (C54T/C97A/D72C/R76D/R80C/D127C/V131D/K135C mutant). The fourth mutant, T4L-D, was prepared for the single-arm **C9-Gd** tag and contained cysteine residues at positions 72 and 131 (C54T/C97A/D72C/V131C mutant). Finally, two additional variants of T4L-C were prepared to assess the role of the aspartic acid residue at position *i*+4 in the *i*,*i*+8 attachment mode In the mutant T4L-E (C54T/C97A/D72C/R80C/D127C/K135C), the wild-type residues R76 and V131 were retained in the respective *i*+4 positions, while in the mutant T4L-F (C54T/C97A/D72C/R76D/R80C/D127C/K135C) R76 was changed to an aspartic acid but not V131.

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

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

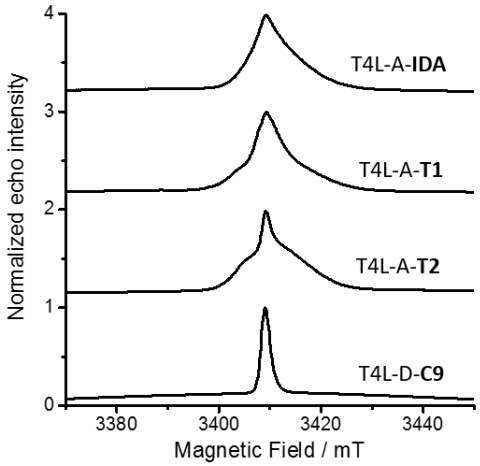


Figure 2. ED-EPR spectra of the central transition region, obtained for the Gd^{3+} ions in different tags attached to the T4 lysozyme mutants T4L-A and T4L-D. The ED-EPR spectra of the tags attached to the other mutants of T4L and MBP were similar and characteristic of the tag rather than the protein.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

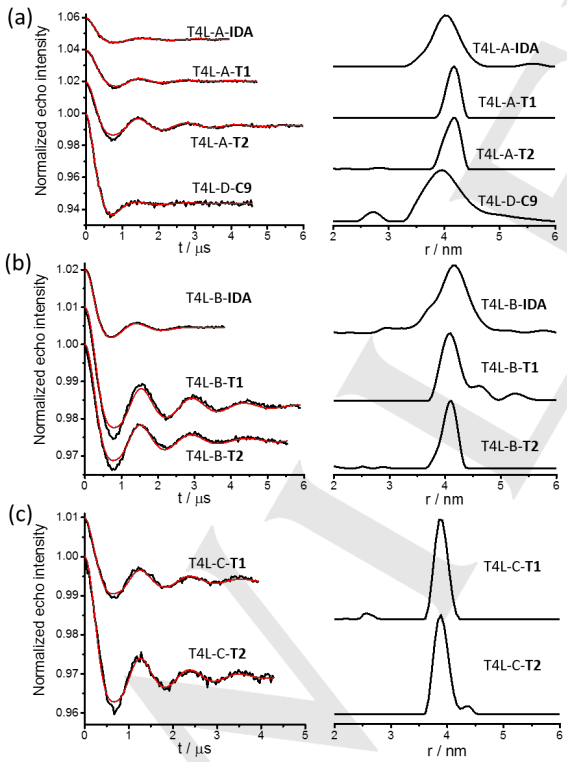


Figure 3. DEER results obtained with different Gd^{3+} tags on T4 lysozyme mutants. (a) T4L-A and T4L-D. (b) T4L-B. (c) T4L-C. (Left) Background corrected DEER traces (black) along with the fit obtained with the distance distributions shown on the right (red). The small mismatch between experimental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

ED-EPR spectra

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

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

Figure 3 shows the DEER results obtained for T4L with the four different Gd^{3+} tags used. As expected based on their small structural differences, the **T1-Gd** and **T2-Gd** tags produced very similar distance distributions for all protein constructs designed for double-arm tag attachment. The **IDA-SH** and **C9-Gd** tags produced significantly broader distance distributions with maxima within 0.3 nm of the maxima observed with the **T1-Gd** or **T2-Gd** tags. As the zero-field splitting of the **C9-Gd** tag is small and the central line narrow, distance distributions obtained with this tag are prone to artificial broadening arising from neglecting the pseudo-secular terms of the dipolar interaction.[23] We checked this hypothesis by carrying out RIDME experiments, which produce dipolar evolution traces that are much less sensitive to the effects of the pseudo-secular terms than the DEER experiments.[24] The RIDME data, recorded of T4L-B with **T1-Gd** tag and of T4L-D with **C9-Gd** tag, showed a clear contribution from a second harmonic when analyzed with DeerAnalysis[25] (Figure 4) and also a third weaker peak, the position of which does not quite match the expected third harmonic and therefore cannot be assigned unambiguously. Most importantly, however, the sample tagged with **C9-Gd** showed a distance distribution for the first harmonic that was significantly narrower than in the DEER experiment (full width at half amplitude of about 0.5 nm versus 0.8 nm in the DEER results), whereas the RIDME and DEER measurements of T4L-B with **T1-Gd** tag showed very similar distance distributions apart from a minor (0.1 nm) shift in the maximum (Figure 4). RIDME measurements of T4L-C with **T2-Gd** tag did not reveal any narrowing with respect to DEER either (Figure S5.) We attribute the conservation in distribution width to the significantly broader EPR line width associated with the **T1-Gd** and **T2-Gd**

tags, which allows the dipolar pseudo-secular terms to be neglected without risking the artificial broadening associated with a data analysis carried out under the weak coupling approximation.[23] The comparison between RIDME and DEER data also suggests that the narrow component observed in some of the ED-EPR spectra measured with the **T1-Gd** and **T2-Gd** tags (Figures 2, S1, and S2) has no significant impact on the distance distribution width.

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

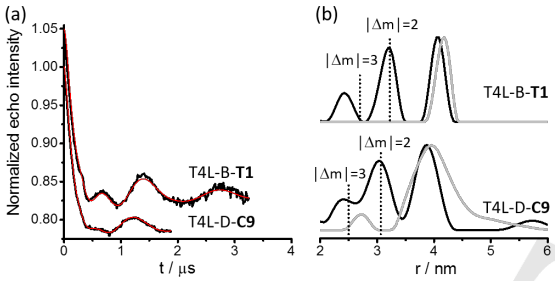


Figure 4. RIDME results obtained with T4L-B-**T1-Gd** and T4L-D-**C9-Gd**. (a) Background corrected RIDME traces (black) along with the fit obtained with the distance distributions shown on the right (red). The uncorrected traces are shown in Figure S4. (b) Comparison of the distance distributions obtained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

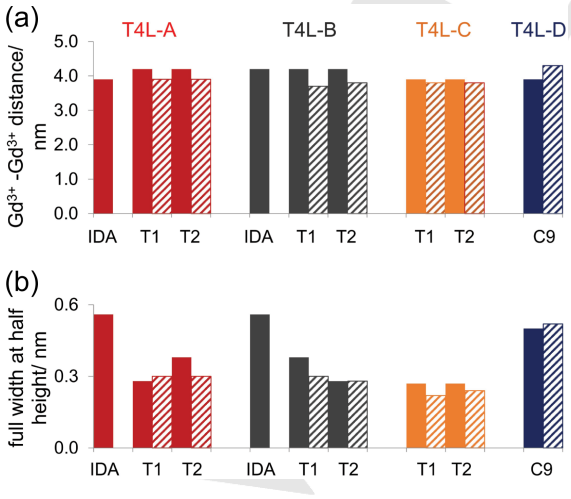


Figure 5. Bar graphs of the experimental and modeled values of distance and width of the distance distributions in different T4 lysozyme mutants. The distance was read from the maximum of a distance distribution and the width is reported as the full width at half-height. Experimental and modeled data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

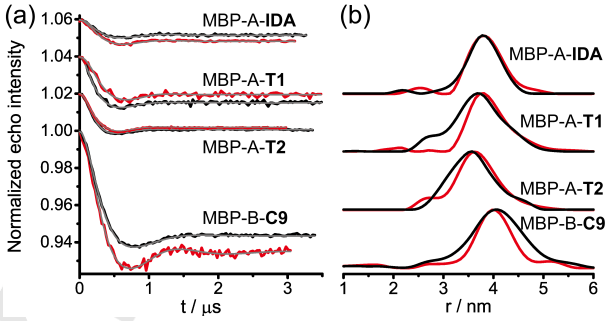


Figure 6. DEER results obtained with different tags on MBP mutants with (red) and without (black) one equivalent of maltose. (a) Background corrected DEER traces along with the fit obtained with the distance distributions shown on the right (grey). The uncorrected traces are shown in Figure S10 and corresponding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

Except for the samples with **IDA-SH** tag, where the precise complexation mode is uncertain, we benchmarked the experimental results against theoretical predictions by modeling the tags on the crystal structures with different rotamer states of the linker segments between the Gd^{3+}-complex and protein backbone, and used the rotamer libraries to compute distance distributions. Figure 5 shows that, with a single exception, the maxima of the experimentally determined distance distributions were all within 0.3 nm of the simulated distance distributions. The exception is T4L-B tagged with **T1-Gd** or **T2-Gd**, where the experimental distance was up to 0.5 nm longer than the distance predicted by modeling. In contrast, the C9-Gd tag attached to T4L-D showed a shorter experimental distance than predicted by modeling (Figure 5a). Excellent agreement, within 0.1 nm, between experimental and predicted distance and distribution width was obtained for the **T1-Gd** and **T2-Gd** tags attached to T4L-C in the *i*,*i*+8 attachment mode. This indicates that a smaller conformational space available for the tag not only delivers narrow distance distributions, but also improves the accuracy with which Gd^{3+}–Gd^{3+} distance distributions can be predicted.

To assess the importance of an aspartic acid residue in position *i*+4 in the *i*,*i*+8 attachment mode, we also measured DEER and RIDME data of the mutants T4L-E and T4L-F with **T1-Gd** and **T2-Gd** tags. T4L-E contains wild-type residues in both *i*+4 positions, while T4L-F contains an aspartic acid mutation at one of the *i*+4 positions and a wild-type residue

(valine) at the other. Both mutants produced consistently broader distance distributions than the T4L-C mutant (Figures S6-S9), highlighting the value of an aspartic acid residue in position *i*+4 when using **T1-Gd** or **T2-Gd** tags in the *i*,*i*+8 attachment mode in an α-helix.

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

The DEER results obtained for MBP with and without maltose are shown in Figure 6. The distance distributions were significantly broader than for T4 lysozyme and of similar width for the IDA-SH, T1-Gd, and T2-Gd tags. The C9-Gd tag also produced a broad distance distribution and RIDME measurements (shown in Figure S4) suggest that the width is intrinsic and due to a multitude of conformations and not due to pseudo-secular term contributions. The comparison with the calculated distance distributions show a better agreement between experimental and modeled distances than for T4 lysozyme, with no difference greater than 0.1 nm (Figure 7a). The addition of maltose did not change the distances significantly, but the widths of the distance distributions became narrower (except for the sample tagged with IDA-SH, where the change in distance distribution width was very small). As in the case of T4 lysozyme, the predicted distribution widths were consistently narrower for the double-arm tags than the C9-Gd tag, but they notably underestimated the experimental distribution widths for the double-armed T1-Gd and T2-Gd tags (Figure 7b), although the computations accepted tag conformations with inter-atomic distances slightly shorter than the sum of van der Waals radii. This result suggests that the MBP domain tagged in the present work is more flexible than the C-terminal domain of T4 lysozyme and that the binding of maltose rigidifies the protein to some degree but not completely. In the case of the C9-Gd tag, the presence of maltose resulted in a more substantial narrowing of the distribution than for any of the double-arm tags. While it is difficult to pinpoint the molecular mechanism underlying this observation, the flexibility of the tether of a single-arm tag obviously makes it more prone to finding a preferential binding site on the protein. Such effects are known[27] and would change the distance distributions.

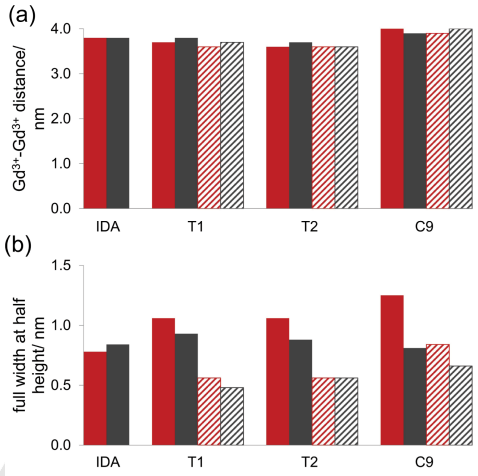


Figure 7. Bar graphs of the experimental and modeled distance and width of the distance distributions in MBP in the absence and presence of maltose. The distance was read from the maximum of a distance distribution and the width is reported as the full width at half-height. Experimental and modeled data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

Tag-specific contribution to the widths of distance distributions

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

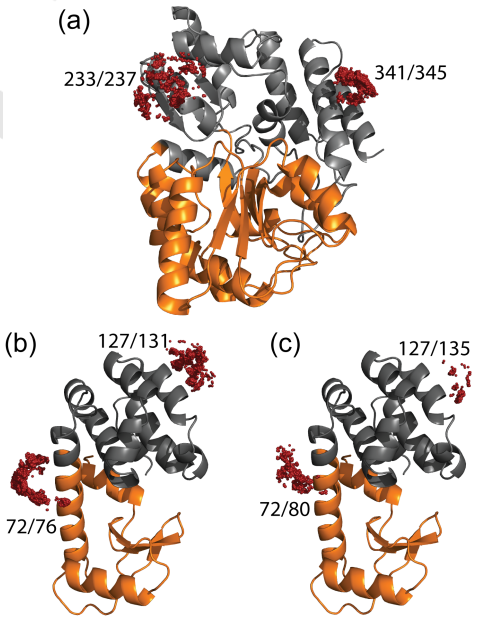


Figure 8. Coordinates found for the Gd^{3+} ion in the **T2-Gd** tag attached to MBP-A, T4L-A and T4L-C. Both proteins assume a two-domain structure, which is highlighted by coloring the N-terminal and C-terminal domains in orange and grey, respectively. The metal positions predicted by modeling are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

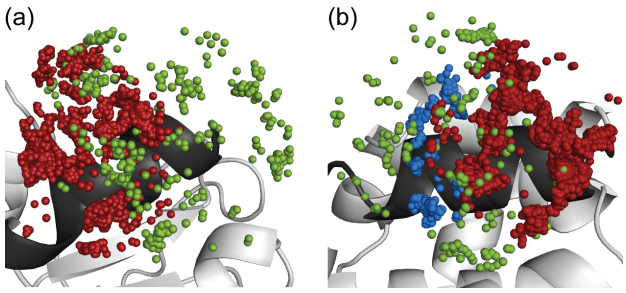


Figure 9. Comparison of Gd^{3+} ion coordinates predicted by modeling of **T2-Gd** and **C9-Gd** tags. Predicted metal positions are indicated by spheres. The helix carrying the tags is shown in dark grey. (a) Metal positions predicted for the site C341/C345 in MBP-A tagged with T2-Gd (red) and site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

Discussion

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

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

Considering that the **IDA-SH** tag has only two chemical bonds between the sulfur atom and the IDA moiety, two **IDA-SH** tags coordinated to a single Gd^{3+} ion would be expected to yield even better immobilization of the metal ion than the **T1-Gd** and **T2-Gd** tags. It is thus surprising that the widths of the distance distributions obtained with the **IDA-SH** tags were much larger than expected. A possible explanation for this effect may be multiple coordination modes for the metal ion. The modulation depths in the DEER traces obtained with the **IDA-SH** tags, about 60% of those obtained for the **T1-Gd** and **T2-Gd** tag (Figures 3a, 3b, and 6a), are somewhat shallower than expected from the differences in ED-EPR line widths, suggesting a lower binding affinity of the metal ion. To improve the binding affinity, we also attempted the attachment of **NTA-SH** tags[32] but consistently failed to install more than three tag molecules in the same protein sample (data not shown), possibly due to electrostatic repulsion between two negatively charged NTA moieties in close proximity. Quite generally, the **IDA-SH** tags are more difficult to use than the cyclen-based tags, because they need to be titrated with Gd^{3+} ions after their ligation to the target protein. In view of the small sample volumes associated with DEER measurements at W-band, it is easy to over- or under-titrate the samples. In the present work, the **IDA-SH** tags delivered no distribution widths much smaller than 0.6 nm, whereas we achieved distribution widths below 0.3 nm with the **T1-Gd** and **T2-Gd** tags. These widths approach those reported previously for the nitroxide double-arm tag HO-1944 attached to two different proteins (0.15-0.35 nm), where no attempts were made to model the distance distributions obtained.[15] They are narrower even than those reported from DEER measurements of chemically synthesized rigid Gd^{3+}–Gd^{3+} rulers (which ranged from 0.32-0.48 nm for distances of 3.1-4.8 nm)[23] and close to the width of 0.28 nm measured by RIDME for another synthetic Gd^{3+}–Gd^{3+} ruler, where the Gd^{3+} ions are separated by only 2.35 nm.[24]

Distribution widths are governed both by protein dynamics and tag flexibility. While the tag's contribution to the distribution width can readily be assessed by modeling of a rotamer library, assessing the range of protein conformations is less straightforward as molecular dynamics runs are sensitive to the accuracy of the force fields used. In the present work, we positioned the tags in single domains of two-domain proteins (Figure 8). In the case of T4 lysozyme, many crystal structures have been determined, which indicate that the C-terminal domain is structurally conserved even when the protein undergoes a hinge-bending motion that changes the relative

position of the N-terminal relative to the C-terminal domain. Indeed, exceptionally narrow distance distributions were obtained with the double-arm tags. In the present work, we used the crystal structure 2LZM of the wild-type protein to model the tag conformations. Using instead the crystal structures 150L and 1SSY of multi-site mutants of T4 lysozyme,[33,34] which have previously been proposed to present better representations of the global conformation of the enzyme in solution,[35,36] the distances simulated with the **C9-Gd** tag varied by ±0.1 nm. Crystal packing effects could thus explain small discrepancies between the experimental and predicted distances.

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

Conclusions

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

Experimental Section

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

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

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

Protein Ligation with C9-Gd, T1-Gd, T2-Gd and IDA Tags: To ensure that all cysteine residues were reduced, DTT was added to a 0.1 mM solution of protein in buffer C to a final concentration of 5 mM DTT. Excess DTT was removed using Amicon ultracentrifugation tubes (MWCO 10 kDa). The reduced protein was added slowly into 5 equivalents of **C9-Gd**, **T1-Gd**, **T2-Gd** or 10 equivalents of **IDA-SH** tag in the same buffer. Reaction mixtures were shaken overnight at room temperature. The completion of the ligation reactions was confirmed by mass spectrometry.

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

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

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

remove instrumental artifacts and to compensate for DC offset. The accumulation time ranged from 2 to 10 h.

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

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

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

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

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

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

Acknowledgements

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

Keywords: distance distribution • double electron-electron resonance • double-arm gadolinium tag • EPR spectroscopy • maltose binding protein • T4 lysozyme

[1] G. Jeschke, *Annu. Rev. Phys. Chem.* **2012**, *63*, 419–446.
[2] G. Jeschke, Y. Polyhach, *Phys. Chem. Chem. Phys.* **2007**, *9*, 1895–1910.
[3] A. M. Raitsimring, C. Gunanathan, A. Potapov, I. Efremenko, J. M. Martin, D. Milstein, D. Goldfarb, *J. Am. Chem. Soc.* **2007**, *129*, 14138–14139.
[4] A. Feintuch, G. Otting, D. Goldfarb, *Methods Enzymol.* **2015**, *563*, 415–457.
[5] D. Goldfarb, *Phys. Chem. Chem. Phys.* **2014**, *16*, 9685–9699.
[6] A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting, D. Goldfarb, *J. Am. Chem. Soc.* **2010**, *132*, 9040–9048.
[7] A. Martorana, G. Bellapadrona, A. Feintuch, E. Di Gregorio, S. Aime, D. Goldfarb, *J. Am. Chem. Soc.* **2014**, *136*, 13458–13465.
[8] M. Qi, A. Gross, G. Jeschke, A. Godt, M. Drescher, *J. Am. Chem. Soc.* **2014**, *136*, 15366–15378.
[9] F. C. Mascali, H. Y. V. Ching, R. M. Rasia, S. Un, L. C. Tabares, *Angew. Chem. Int. Ed.* **2016**, *55*, 11041–11043.
[10] F. X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb, P. Selenko, *Nature* **2016**, *530*, 45–50.
[11] Y. Yang, F. Yang, Y. J. Gong, J. L. Chen, D. Goldfarb, X.-C. Su, *Angew. Chem. Int. Ed.* **2017**, *56*, 2914–2918.
[12] B. Graham, C. T. Loh, J. D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, G. Pintacuda, T. Huber, G. Otting, *Bioconjugate Chem.* **2011**, *22*, 2118–2125.
[13] H. Yagi, D. Banerjee, B. Graham, T. Huber, D. Goldfarb, G. Otting, *J. Am. Chem. Soc.* **2011**, *133*, 10418–10421.
[14] E. H. Abdelkader, M. D. Lee, A. Feintuch, M. R. Cohen, J. D. Swarbrick, G Otting, B. Graham, D. Goldfarb, *J. Phys. Chem. Lett.* **2015**, *6*, 5016–5021.
[15] M. R. Fleissner, M. D. Bridges, E. K. Brooks, D. Cascio, T. Kálai, K. Hideg W. L. Hubbell, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16241–16246.
[16] M. D. Lee, M. L. Dennis, J. D. Swarbrick, B. Graham, *Chem. Commun.* **2016**, *52*, 7954–7957.
[17] J. D. Swarbrick, P. Ung, X.-C. Su, A. Maleckis, S. Chhabra, T. Huber, G. Otting, B. Graham, *Chem. Commun.* 2011, 47, 7368–7370.
[18] J. D. Swarbrick, P. Ung, M. L. Dennis, M. D. Lee, S. Chhabra, B. Graham, *Chem. Eur. J.* **2016**, *22*, 1228–1232.
[19] P. H. J. Keizers, J. F. Desreux, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.* **2007**, *129*, 9292–9293.
[20] P. H. J. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand, M. Ubbink, *J. Am. Chem. Soc.* **2008**, *130*, 14802–14812.
[21] M. Matsumura, B. W. Matthews, *Science* **1989**, *243*, 792–794.
[22] M. D. Lee, C.-T. Loh, J. Shin, S. Chhabra, M. L. Dennis, G. Otting, J. D. Swarbrick, B. Graham, *Chem. Sci.* **2015**, *6*, 2614–2624.
[23] A. Dalaloyan, M. Qi, S. Ruthstein, S. Vega, A. Godt, A. Feintuch, D. Goldfarb, *Phys. Chem. Chem. Phys.* **2015**, *17*, 18464–18476.
[24] A. Collauto, V. Frydman, M. D. Lee, E. H. Abdelkader, A. Feintuch, J. D. Swarbrick, B. Graham, G. Otting, D. Goldfarb, *Phys. Chem. Chem. Phys.* **2016**, *18*, 19037–19049.
[25] S. Razzaghi, M. Qi, A. I. Nalepa, A. Godt, G. Jeschke, A. Savitsky, M. Yulikov, *J. Phys. Chem. Lett.* **2014**, *5*, 3970–3975.
[26] L. H. Weaver, B. W. Matthews, *J. Mol. Biol.* **1987**, *193*, 189–199.
[27] E. H. Abdelkader, A. Feintuch, X. Yao, L. A. Adams, L. Aurelio, B. Graham, D. Goldfarb, G. Otting, *Chem. Commun.* **2015**, *51*, 15898–15901.
[28] A. J. Sharff, L. E. Rodseth, J. C. Spurlino, F. A. Quiocho, *Biochemistry* **1992**, *31*, 10657–10663.
[29] F. A. Quiocho, J. C. Spurlino, L. E. Rodseth, *Structure* **1997**, *5*, 997–1015.
[30] A. Doll, M. Qi, N. Wili, S. Pribitzer, A. Godt, G. Jeschke, *J. Magn. Reson.* **2015**, *259*, 153–162.
[31] M. A. S. Hass, P. H. J. Keizers, A. Blok, Y. Hiruma, M. Ubbink, *J. Am. Chem. Soc.* **2010**, *132*, 9952–9953.
[32] H. Yagi, A. Maleckis, G. Otting, *J. Biomol. NMR* **2013**, *55*, 157–166.
[33] M. M. He, Z. A. Wood, W. A. Baase, H. Xiao, B. W. Matthews, *Protein Sci.* **2004**, *13*, 2716–2724.

[34] S. Daopin, T. Alber, W. A. Baase, J. A. Wozniak, B. W. Matthews, *J. Mol. Biol*. **1991**, *221*, 647–667.
[35] N. K. Goto, N. R. Skrynnikov, F. W. Dahlquist, L. E. Kay, *J. Mol. Biol*. **2001**, *308*, 745–764.
[36] J.-L. Chen, Y. Yang, L.-L. Zhang, H. Liang, T. Huber, X.-C. Su, G. Otting, *Phys. Chem. Chem. Phys.* **2016**, *18*, 5850–5859.
[37] C. Tang, C. D. Schwieters, G. M. Clore, *Nature* **2007,** *449*, 1078–1082.
[38] C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love, N. E. Dixon, *Biochemistry* **2000**, *39*, 11989–11999.
[39] P. S. C. Wu, K. Ozawa, S. P. Lim, S. G. Vasudevan, N. E. Dixon, G. Otting, *Angew. Chem. Int. Ed*. **2007**, *46*, 3356–3358.
[40] M. Apponyi, K. Ozawa, N. E. Dixon, G. Otting, *Methods in Molecular Biology*, 2008, **426**, Structural proteomics: high-throughput methods; B. Kobe, M. Guss and T. Huber Eds., Humana Press, Totowa, USA, pp. 257–268.
[41] D. Goldfarb, Y. Lipkin, A. Potapov, Y. Gorodetsky, B. Epel, A. M. Raitsimring, M. Radoul, I. Kaminker, *J. Magn. Reson.* **2008**, *194*, 8–15.
[42] F. Mentink-Vigier, A. Collauto, A. Feintuch, I. Kaminker, V. Tarle, D. Goldfarb, *J. Magn. Reson.* **2013**, *236*, 117–125.
[43] M. Pannier, S. Veit, A. Godt, G. Jeschke, H. W. Spiess, *J. Magn. Reson.* **2000**, *142*, 331–340.
[44] S. Milikisyants, F. Scarpelli, M. G. Finiguerra, M. Ubbink, M. Huber, *J. Magn. Reson.* **2009**, *201*, 48–56.
[45] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger, H. Jung, *Appl. Magn. Reson.* **2006**, *30*, 473–498.
[46] M. Stanton-Cook, X.-C. Su, G. Otting, T. Huber, http://compbio.anu.edu.au/mscook/PPT.

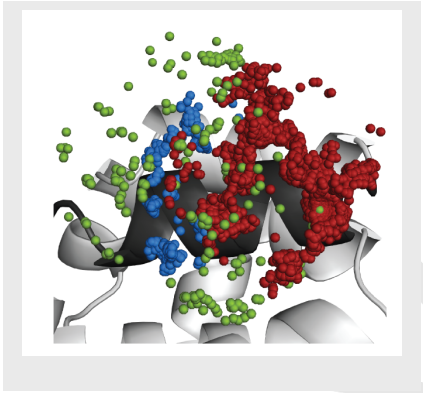
10.1002/chem.201702521

WILEY-VCH

Entry for the Table of Contents

FULL PAPER

Metal on a tether: Double-arm cyclen-based Gd^{3+} tags confine the space accessible to the metal ion and produce accurate nanometer scale Gd^{3+}–Gd^{3+} distance measurements by double electron-electron resonance (DEER) experiments, which show excellent agreement with predictions both for the maximum and width of the measured distance distributions.



Adarshi P. Welegedara, Yin Yang, Michael D. Lee, James D. Swarbrick, T. Huber, Bim Graham, Daniella Goldfarb,* Gottfried Otting**

Page No. – Page No.

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