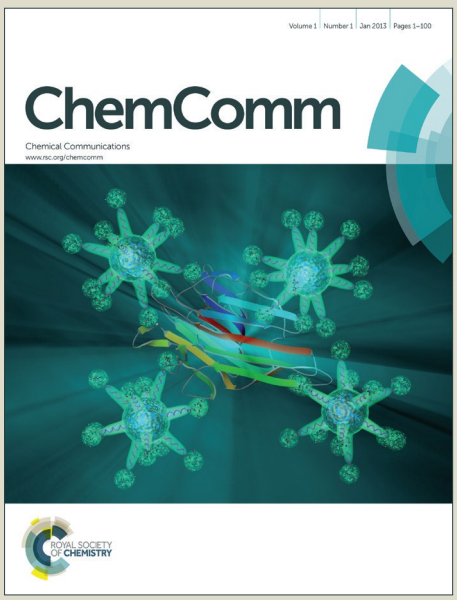
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

COMMUNICATION

Protein Conformation by EPR Spectroscopy Using Gadolinium Tags Clicked to Genetically Encoded *p*-Azido-L-Phenylalanine

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

E. H. Abdelkader,[a] A. Feintuch,[b] X. Yao,[a] L. A. Adams,[c] L. Aurelio,[c] B. Graham,[c] D. Goldfarb[b] and G. Otting[a]

Quantitative cysteine-independent ligation of a Gd^{3+} tag to genetically encoded *p*-azido-*L*-phenylalanine via Cu(I)-catalyzed click chemistry is shown to deliver an exceptionally powerful tool for Gd^{3+}-Gd^{3+} distance measurements by double electron-electron resonance (DEER) experiments, as the position of the Gd^{3+} ion relative to the protein can be predicted with high accuracy.

The DEER experiment has turned pulsed EPR into a popular tool in structural biology because of its unique capability of measuring specific distances in the range of about 2 to 8 nm, while requiring only small amounts of sample, particularly when measured at Q- and W-band frequencies, and being largely independent of molecular weight.[1-4] DEER experiments measure the dipolar interaction and, hence, the distance between two electron spins.[5-7] The electron spins have to be introduced into the (usually diamagnetic) molecule of interest by site-directed spin labelling (SDSL).[8-10] Traditionally, SDSL of proteins utilizes tags that react with cysteine residues, posing problems when naturally occurring cysteine residues are essential for the structure or function of the target protein or when the protein contains many natural cysteines. A much more general way of SDSL employs genetically encoded unnatural amino acids that can be site-specifically incorporated into proteins using orthogonal tRNA/aminoacyl-tRNA synthetase systems.[11] In a first attempt, Hubbell and co-workers incorporated *p*-acetyl-L-phenylalanine (*p*-AcPhe) into T4 lysozyme singly or in pairs and with a nitroxide tag containing a hydroxylamine group (K1 tag).[12,13] The resulting overall tag contained multiple rotatable bonds, making the location of the nitroxide group relative to the protein both uncertain and heterogeneous. Furthermore, ligation reactions with *p*-AcPhe tend to be incomplete at neutral pH. A more recent approach employed a nitroxide-containing unnatural amino acid, but this amino acid contains even more rotatable bonds between protein backbone and the nitroxide.[14] Furthermore, *in vivo* protein expression with the new nitroxide amino acid resulted in chemical reduction of one third or more of the nitroxide spin labels, which is detrimental for distance measurements.

The present work introduces an alternative approach to SDSL for EPR based distance measurements, which combines quantitative ligation yields at neutral pH with chemically stable spin labels, highly predictable distances and high absolute sensitivity. It is based on ligating a Gd^{3+} tag to two *p*-azido-L-phenylalanine (AzF) residues site-specifically incorporated into the target protein by an orthogonal tRNA/*p*-cyanophenylalanyl-tRNA synthetase system.[15,16] The spin label is provided by the C3-Gd^{3+} tag, which is attached to AzF by Cu(I)-catalysed click chemistry (Fig. 1).[17] Gd^{3+} tags offer exceptional sensitivity for DEER measurements at W-band (95 GHz), but have so far only been used in tags attached to cysteine residues.[18]

We demonstrate this new approach with the *E. coli* aspartate/glutamate binding protein (DEBP), a periplasmic 34 kDa protein containing a disulfide bond. A crystal structure of the glutamate complex of the homologous protein from *Shigella flexneri* has been determined.[19] The *S. flexneri* enzyme differs in only three residues (D127N, G130D, D200E). Glutamate binds between two rigid domains that encase the bound amino acid.

High ligation yields are important for DEER experiments, as poor yields of doubly tagged protein cannot simply be compensated for by increased sample concentration, which would increase the non-specific background of paramagnetic centres and reduce the modulation depth. This makes it difficult to subtract the correspondingly enhanced background decay prior to analysis of the remaining signal for oscillations arising from specific intra-molecular electron-electron distances. Initial attempts to attach the C3-Gd^{3+} tag to two site-specifically incorporated AzF residues by Cu(I)-catalysed click chemistry (Fig. 1) resulted in unsatisfactory

[a] *Research School of Chemistry, Australian National University, Canberra, ACT 2601 (Australia). Email: gottfried.otting@anu.edu.au*
[b] *Department of Chemical Physics, Weizmann Institute of Science, Rehovot 76100 (Israel). Email: daniella.goldfarb@weizmann.ac.il*
[c] *Monash Institute of Pharmaceutical Sciences, Monash University, Parkville VIC 3052 (Australia).*

† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

yields. First, *in vivo* protein expression lead to partial reduction of the azide group to an amino group. We solved this problem by producing the protein by cell-free synthesis. No more than 15% reduction occurred after 16 h of cell-free synthesis (depending on the site of the AzF residue in the protein) and limiting the cell-free reaction to 6 h proved to avoid the reduction without significant loss in protein yield. Cell-free synthesis had the additional benefit of allowing protein production from linear PCR-amplified DNA[20] and avoiding premature truncation at amber stop codons by facile removal of RF1.[21] Second, click reactions with the sterically demanding C3 tag are more difficult than for tags with more accessible acetylene groups.[17] Based on earlier observations, which reported incomplete ligation yields for proteins containing His_6 tags,[14] we cleaved the His_6 tag at the C-terminus of DEBP using TEV protease. Nonetheless, the ligation reaction was accompanied by severe copper-induced protein precipitation and the ligation yields remained low, until we added 150 mM NaCl to the reaction mixture (Fig. S4). The final reaction conditions included 0.05 mM protein, 0.5 mM tag, 0.3 mM $CuSO_4$, 1.5 mM BTTAA, 5 mM aminoguanidine, 0.5 mM glycerol, 5 mM ascorbic acid, 50 mM sodium phosphate (pH 8) and 150 mM NaCl at room temperature, and produced near-quantitative yields of double-tagged DEBP.

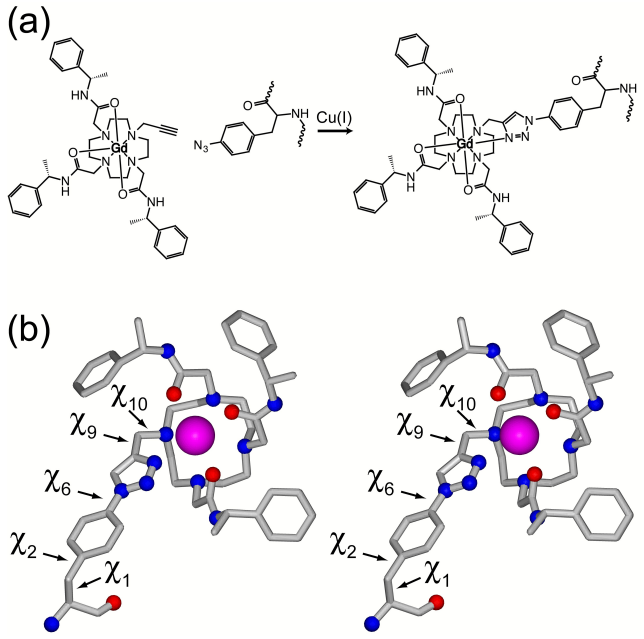


Fig. 1 Gd^{3+} tag used in the present work. The tag is positively charged owing to the charge of the Gd^{3+} ion. (a) Click chemistry reaction for attaching the C3-Gd^{3+} tag to an AzF residue. We refer to the ligation product as an AzF-C3-Gd^{3+} residue. (b) Stereoview of the AzF-C3-Gd^{3+} residue in the side chain conformation found in the present work. Nitrogen, oxygen and gadolinium atoms are shown as blue, red and magenta spheres, respectively. In the conformation shown, the dihedral angle χ_6 is 180° and coordination of the Gd^{3+} ion by a nitrogen of the triazole determines the angles χ_9 and χ_{10}.

Fifteen double-AzF mutants of DEBP ligated with C3-Gd^{3+} tags were prepared and subjected to W-band[22] Gd^{3+}-Gd^{3+} DEER distance measurements. The measurements were carried out at 10 K, using a home-built EPR spectrometer.[22] The protein solution (about 3 μL) was loaded into quartz capillaries (0.84 mm outer diameter, 0.[illegible] mm inner diameter). All mutants studied gave similar echo-detected EPR spectra (Fig. S5). The echo decay rate allowed an evolution time of up to 8 μs. DEER data were recorded using the four-pulse sequence (Fig. S6)[7] with t ranging from τ_1-200 ns to τ_1-500 ns. The pump pulse was set to the maximum of the EPR spectrum and the observer pulses 100 MHz higher. The length of the pump pulse, t_{pump}, was 15 ns, the length of the π/2 observer pulse, $t_{observer,\ \pi/2}$, was 15 ns and the length of the π observer pulse, $t_{observer,\ \pi}$, was 30 ns. The delay time τ was 350 ns and the repetition time was 800 μs. The cavity tuning was set to 94.9 GHz. Accumulation time ranged from 2 to 17 h depending on the Gd^{3+}-Gd^{3+} distance. The full transient echo traces were collected for each t value in the DEER sequence and the echo integration was carried out after data collection. The integration gate was chosen to obtain the best signal-to-noise ratio. Usually, best results were obtained for an integration width equal to the echo full width at half height. Data analysis was carried out with DeerAnalysis and Tikhonov regularization using the bending point of the L-curve.[23]

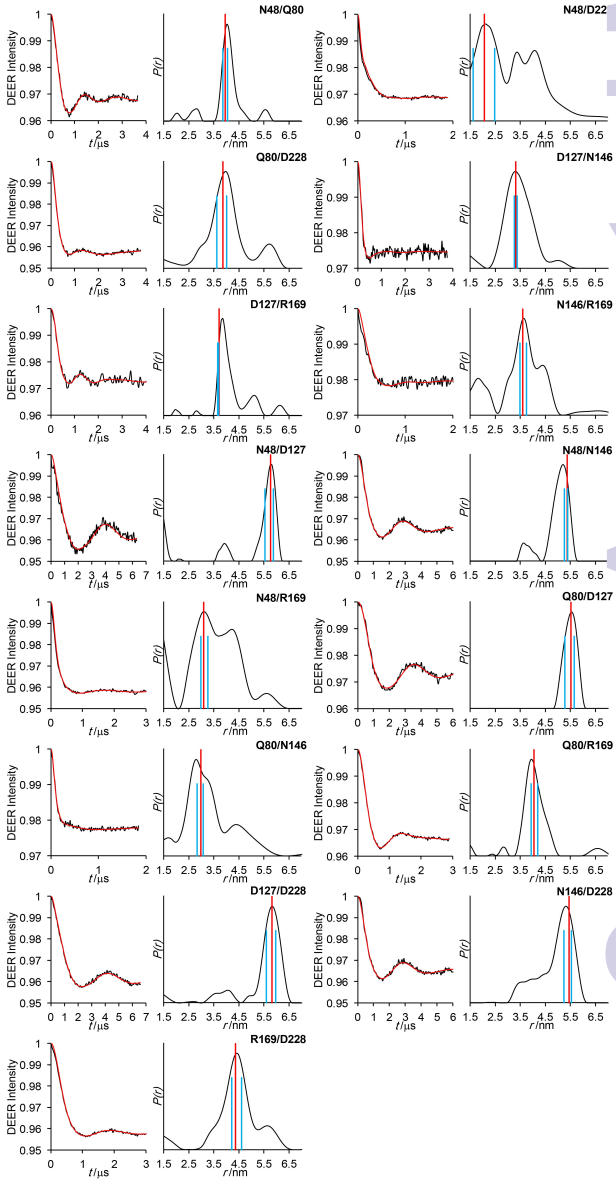


Fig. 2 DEER results at 10 K for 15 different double AzF mutants of DEBP ligated with C3-Gd^{3+} tags and with bound glutamate. The

residues replaced by AzF in each double mutant are specified in each subfigure. Left panels: Normalized DEER traces after background removal, showing the fits obtained by Tikhonov regularization[23] as smooth lines. Right panels: corresponding distance distributions. The red line indicates the distance calculated from modelling the AzF-C3-Gd^{3+} residues onto the crystal structure of the *S. flexneri* homologue (PDB ID: 2VHA).[19] The blue lines mark the distances predicted by the same models, except that the χ_1 angle of both AzF residues was changed by ±15° to explore the effect of χ_1 angle variability. (For residue 169, the χ_1 angle was varied between 0° and +15°, as more negative angles would have lead to severe van der Waals clashes.) Of the four resulting distances, the grey lines mark only those that are most different from the starting conformations. The top three rows show the intra-domain distance measurements. Protein concentration was 100 μM and sample volume 2-3 μl.

Fig. 2 shows the results (see Fig. S7 for the primary DEER data before background subtraction). The modulation depths (3-4% for all mutants) indicate a consistently high labelling efficiency. To predict the Gd^{3+}-Gd^{3+} distances, the mutation tool of PyMOL was used to determine χ_1 and χ_2 angles of the AzF-C3-Gd^{3+} residues. The angle χ_6 was set to 180° and the Gd^{3+} ion coordination was as in Fig. 1b. The agreement between the predicted distances and the maxima of the experimental distance distributions is remarkably good considering that the predictions used only single tag conformations (Fig. 3c). Only the AzF-C3-Gd^{3+} residue at site 169 displayed severe steric clashes for the χ_1 rotamer indicated by the experimental distances. They were resolved by adjusting the χ_9 and χ_{10} angles, and abandoning the coordination of the Gd^{3+} ion by the triazole ring. The non-standard tag geometry at this site seems to arise from specific binding interactions between the cyclen ring system and the protein, whereas the tags at the other sites are highly solvent-exposed (Fig. 3).

The few seemingly broad and multi-distance distance distributions may have different sources. For mutants with short distances (<3 nm as in, e.g., mutant 48/228), the weak coupling approximation used in the data analysis breaks down and, as a consequence, the distance distribution broadens and shows some artefact peaks.[24] In addition, it cannot be ruled out that mutant 48/228 includes minor populations of alternative χ_1 rotamers of the AzF residues, favoured by electrostatic repulsion between the tags, as the peaks around 3.5 nm are rather intense (Fig. 2). For mutant 48/169, the multi-distance distribution may similarly reflect the occurrence of an alternative conformation of the tag at position 169, which can, without van der Waals violations, also be modelled with an alternative χ_1 angle that would lengthen the Gd^{3+}-Gd^{3+} distance in the mutant 48/169 by about 1 nm. Most important, however, the tallest peak of every single distance distribution can be explained by single tag conformations, which, with the exception of site 169 discussed above, indicates a preferential population of the χ_6 = 180° rather than the χ_6 = 0° rotamer (Fig. 3c and d).

The widths of the distance distributions obtained with the C3-Gd^{3+} tag compare favourably with those obtained with the chemically similar C1-Gd^{3+} tag.[25] Every one of the six tagging sites yielded at least one distance distribution with a full width at half height below 1 nm (Fig. 2), indicating that increased widths contain significant information about conformational variations. Obviously, variations in χ_1 or χ_6 angles cou broaden the distributions. More interestingly, the long ar rigid tether of the tag also amplifies small variations in th backbone structure. Although we engineered all mutation si in elements of regular secondary structure, different taggi sites still exhibited different distance distribution widths. For example, mostly narrow distance distribution widths we observed for site 127, while mostly wide distributions were associated with position 146. With a greater number of taggi sites, it may thus be possible to use the AzF-C3 tag for studying the amplitudes and directions of local protein backbone flexibility.

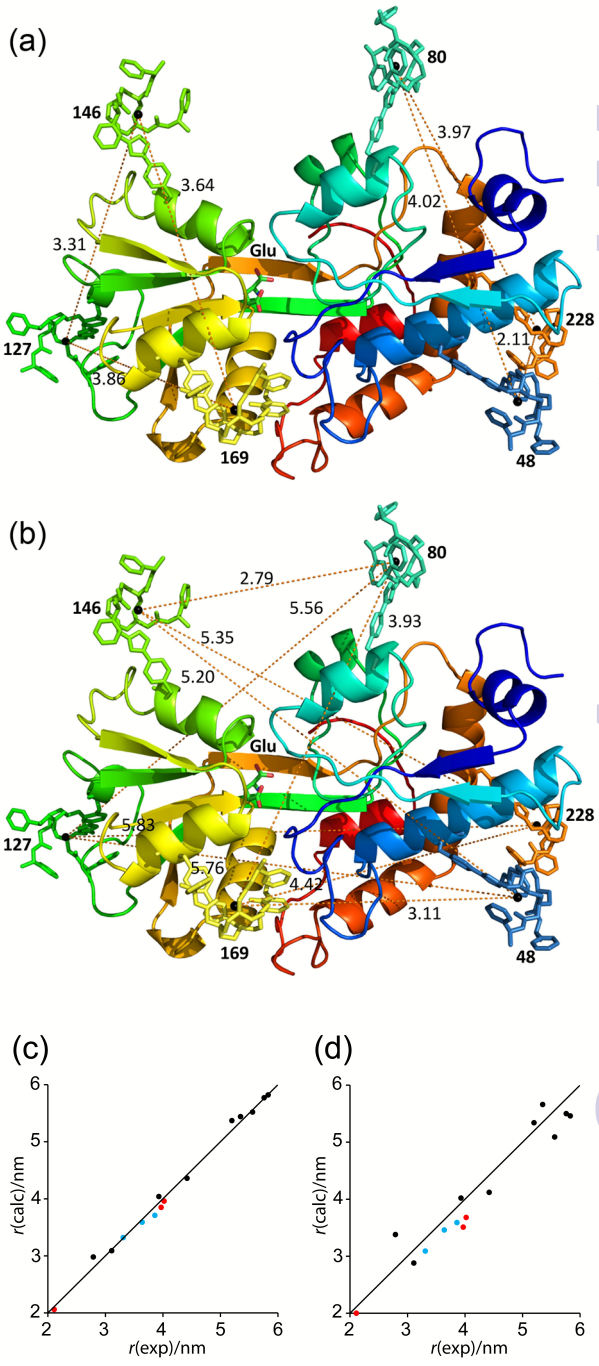


Fig. 3 Model of DEBP[16] with AzF-C3-Gd^{3+} residues. We refer to the domains on the left and right as the small and large domains, respectively. The side chains of the AzF-C3-Gd^{3+} residues were modelled as described in the main text and are shown in the colours of the backbone sites to which they are attached. Gd^{3+}-Gd^{3+} distances are identified by dashed lines. (a) Model displaying the intra-domain distances (in nm). (b) Same as (a), but showing the inter-domain distances. (c) Correlation between experimental and predicted Gd^{3+}-Gd^{3+} distances. Blue and red points mark distances in the small and large domain, respectively. Black points mark inter-domain distances. The experimental distances correspond to the maxima of the distance distributions measured by DEER experiments (Fig. 2). The predicted distances were calculated using $\chi_6 = 180^o$. (d) Same as (c), except that the predicted distances were calculated using $\chi_6 = 0^o$ (except for residue 169, for which $\chi_6 = 180^o$ was used throughout to avoid steric clashes).

The absolute maxima of the distance distributions derived from DEER data measured in the absence of glutamate were at very similar positions as those measured in the presence of glutamate, indicating preservation of the closed conformation in the frozen state of the apo-protein and highlighting the high reproducibility of the distance measurements (Fig. S8).

In conclusion, AzF-C3-Gd^{3+} residues make Gd^{3+}-Gd^{3+} distance measurements by DEER exceptionally powerful by positioning the Gd^{3+} ions at unique positions that can easily be predicted from the three-dimensional structure (or model) of the target protein. This is an important advance on conventional nitroxide tags that are attached to cysteine residues and generate distributions of nitroxide positions that need to be modelled.[26] Every one of the six sites we mutated successfully incorporated the unnatural amino acid without evidence for structural perturbation. In addition, this work established a general protocol for obtaining the requisite quantitative ligation yields at every single site using the Cu(I)-catalyzed reaction with the sterically demanding C3 tag while avoiding copper-induced precipitation of proteins. In contrast to all previous Gd^{3+} tags used with proteins,[25,27,28] the C3 tag does not depend on ligation to cysteine residues. Combined with efficient incorporation of two unnatural amino acids by cell-free synthesis, facile production of mutant proteins from PCR-amplified DNA, and the chemical stability of the resulting AzF-C3-Gd^{3+} residue, high sensitivity Gd^{3+}-Gd^{3+} distance measurements by DEER experiments offer unprecedented new opportunities in structural biology, including in-cell measurements, which require a redox-resistive conjugation to the protein.

This work was supported by the Australian Research Council (ARC), an Australia-Weizmann Making Connections grant, and the Israel Science Foundation (ISF). We thank Professor Thomas Huber for a critical reading of the manuscript, Professor Peter G. Schultz and Dr Colin C. Jackson for genes of the orthogonal tRNA/aminoacyl-tRNA synthetase system and DEBP, respectively, and Dr Ruhu Qi for help with primer design. B. G. thanks the ARC for a Future Fellowship. In part, this research was made possible by the historic generosity of the Harold Perlman family (D.G.). D. G. holds the Erich Klieger professorial chair in Chemical Physics.

Notes and references

1 G. Jeschke and Y. Polyhach, *Phys. Chem. Chem. Phys.,* 200 , **9**, 1895–1910.
2 EPR Spectroscopy: Applications in Chemistry and Biology, M Drescher and G. Jeschke, Eds., Springer: Berlin, 2012.
3 P. P. Borbat and J. H. Freed, In Structure and Bonding; C. Timmel and J. R. Harmer, Eds.; Springer: Berlin, 2013; Vo . 152, pp 1–82.
4 D. Goldfarb, *Phys. Chem. Chem. Phys*., 2014, **16**, 9685–9699.
5 A. D. Milov, K. M. Salikhov and M. D. Shirov, *Fiz. Tverd. Tel* , 1981, **23**, 975–982.
6 A. D. Milov, A. B. Ponomarev and Y. D. Tsvetkov, *Chem. Phys. Lett*., 1984, **110**, 67–72.
7 M. Pannier, S. Veit, A. Godt, G. Jeschke and H. W. Spiess, *J. Magn. Reson*., 2011, **213**, 316–325.
8 W. L. Hubbell, A. Gross, R. Langen and M. A. Lietzow, *Curr. Opin. Struct. Biol*., 1998, **8**, 649–656.
9 J. P. Klare and H.-J. Steinhoff, *Photosynth. Res*., 2009, **10** 377–390.
10 W. L. Hubbell, C. J. López, C. Altenbach and Z. Yang, *Curr. Opin. Struct. Biol*., 2013, **23**, 725–733.
11 C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem*., 2010, 413–444.
12 M. R. Fleissner, E. M. Brustad, T. Kálai, C. Altenbach, D. Cascio, F. B. Peters, K. Hideg, S. Peuker, P. G. Schultz and W. L. Hubbell, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 21637 21642.
13 L. Garbuio, E. Bordignon, E. K. Brooks, W. L. Hubbell, G. Jeschke and M. Yulikov, *J. Phys. Chem. B*, 2013, **117**, 3145–3153.
14 M. J. Schmidt, J. Borbas, M. Drescher and D. Summerer, *J. Am. Chem. Soc*., 2014, **136**, 1238–1241.
15 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.
16 D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon and P. G. Schultz, *Biochemistry*, 2011, **50**, 1894–1900.
17 C. T. Loh, K. Ozawa, K. L. Tuck, N. Barlow, T. Huber, G. Otting and B. Graham, *Bioconjugate Chem*., 2013, **24**, 260–268.
18 A. Feintuch, G. Otting and D. Goldfarb, *Methods Enzymol* , 2015, in press.
19 Y. Hu, C.-P. Fan, G. Fu, D. Zhu, Q. Jin and D.-C. Wang, *J. Mol. Biol*., 2008, **382**, 99–111.
20 P. S. C. Wu, K. Ozawa, S. P. Lim, S. Vasudevan, N. E. Dixon and G. Otting, *Angew. Chemie Int. Ed*., 2007, **46**, 3356–3358.
21 K. V. Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa and G. Otting, *Angew. Chemie Int. Ed*., 2012, **51**, 2243–2246.
22 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.
23 G. Jeschke, V. Checkik, P. Ionita, A. Godt, H. Zimmermann, J. Bahman, C. R. Timmel, D. Hilger, H. Jung, *Appl. Magn. Reson*., 2006, **30**, 473–498.
24 A. Dalaloyan, M. Qi, S. Ruthstein, S. Vega, A. Godt, A. Feintuch and D. Goldfarb, *Phys. Chem. Chem. Phys*., 201 , **17**, 18464–18476.
25 H. Yagi, D. Banerjee, B. Graham, T. Huber, D. Goldfarb and G. Otting, *J. Am. Chem. Soc*., 2011, **133**, 10418–10421.
26 G. Jeschke, *Prog. NMR Spectr*., 2013, **72**, 42–60.
27 A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting and D. Goldfarb, *J. Am. Chem. Soc*., 2010, **132**, 9040–9048.
28 E. Matalon, T. Huber, G. Hagelueken, B. Graham, A. Feintuch, V. Frydman, G. Otting and D. Goldfarb, *Angew. Chemie Int Ed*., 2013, **52**, 11831–11834.