Journal Name

COMMUNICATION

3D structure determination of a protein in living cells using paramagnetic NMR spectroscopy

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

DOI: 10.1039/x0xx00000x

www.rsc.org/

Bin-Bin Pan,[a] Feng Yang[a], Yansheng Ye[c], Qiong Wu[c], Conggang Li*[c], Thomas Huber*[b] and Xun-Cheng Su*[a]

Determining the three-dimensional structure of a protein in living cells remains particularly challenging. We demonstrated that the integration of site-specific tagging proteins and GPS-Rosetta calculations provides a fast and effective way of determining the structures of proteins in living cells, and in principle the interactions and dynamics of protein-ligand complexes.

It is still an open question whether structural and dynamic behavior of proteins in a cellular environment is truthfully reproduced in *in-vitro* experiments, where the conditions differ greatly in viscosity, molecular crowding and redox potential. Therefore, dissection of the interactions, dynamics and structures of biomolecules in living cells is highly desirable to understand living processes. NMR spectroscopy offers an attractive way of analysing the dynamics, interactions and structures of proteins at atomic resolution in living cells,[1-19] however, its effectiveness is severely limited by the adverse experimental conditions. Crowding conditions generally produce non-specific associations of proteins with each other and also other cellular components, resulting in broader NMR signals. Efforts towards shortening the acquisition time of NMR spectra have been made, however, determining the structure of a protein in living cells is still challenging.[15,17-19] Up to date, only one three-dimensional (3D) protein structure has been experimentally determined in living *E. coli* cells[7] and the 3D structure of a protein in eukaryotic cells has not yet been reported.‡

Paramagnetic effects, including pseudocontact shift (PCS) and paramagnetic relaxation enhancement (PRE), have proven to be important tools in structural biology by NMR spectroscopy,[21-24] which can significantly shorten the NMR time requirement for structure determination. As many proteins do not have paramagnetic centers, generation of paramagnetic effects generally relies on site-specific labeling of proteins with a paramagnetic radical or metal ion. The reducing environment of the cell cytoplasm makes the commonly used disulfide bond modifications of proteins unfeasible for in-cell analysis.[15,18] Maleimide derivative tags have instead been used for in-cell EPR and PRE measurements,[19,25,26] however, the maleimide reaction with a cysteine introduces a new chiral center that leads to diastereomeric protein–tag complexes, and therefore is not suitable for PCS analysis.[27] 4-VPyMTA tag avoids this problem and is suitable for NMR assay in crowding conditions and in-cell EPR analysis,[27,26] but its reaction with a protein thiol results in a long and flexible linker between the protein and a paramagnetic center.[27-30] A better paramagnetic tag for PCS analysis in living cells should be stable and rigid, and not produce multiple paramagnetic resonances in the NMR spectrum.

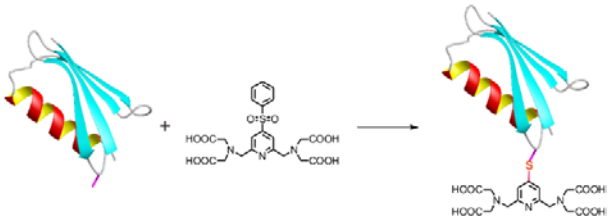
Herein, we used Streptococcal β1 immunoglobulin binding domain of protein G (GB1) as a model protein and site-specifically labeled this protein with a recently developed paramagnetic tag, 4PhSO_2-PyMTA,[29] to form an adduct which is stable under physiological conditions (Scheme 1). The GB1-PyMTA adducts complexed with paramagnetic lanthanide ions were evaluated by ^{15}N-HSQC spectra recorded in aqueous buffer and in living cells respectively. With the GPS-Rosetta program,[31,32] we then assessed the feasibility of determining the structure of GB1 from experimental PCSs that were collected in *Xenopus laevis* oocytes.

The reaction of single cysteine mutant of GB1 (T11C or V21C) with 4PhSO_2-PyMTA tag generated a short and stable thioether bond between the sidechain of the cysteine in GB1 mutant and PyMTA (see Supplementary Information). MALDI-TOF mass spectrometry analyses showed that the GB1 T11C and

[a] *State Key Laboratory and Research Institute of Elemento-Organic Chemistry, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Nankai University, Tianjin 300071, China.* E-mail: xunchegnsu@nankai.edu.cn
[b] *Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia.* E-mail: t.huber@anu.edu.au
[c] *Key Laboratory of Magnetic Resonance in Biological Systems, State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, National Center for Magnetic Resonance in Wuhan, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan 430071, China.* E-mail: conggangli@wipm.ac.cn

† 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

V21C mutants were modified with only one tag molecule (Fig. S1).



Scheme 1 Site specific tagging a protein via formation of a stable thioether bond between target protein and a functional tag for in cell spectroscopic analysis.

The reactivity of V21C to 4PhSO$_2$-PyMTA was lower than that of T11C mutant, but labeling proceeded to near completion with both mutant proteins (see Supplementary Information). ^{15}N-HSQC spectra indicated significant chemical shift changes for residues close to the ligation site (Fig. S2), and some heterogeneity was observed in the ^{15}N-HSQC spectra for the structural segments including the termini of β2, loop β2/α1 and the beginning of α1, which contain residues 18-29. These heterogeneities manifested themselves in two or more cross-peaks present in both the GB1 T11C-PyMTA and V21C-PyMTA spectra, and are likely due to the introduction of the overall negatively net-charged PyMTA tag. The chemical shift heterogeneity of GB1 was first observed in solid state NMR for the protein samples prepared in different ways.[33] The heterogeneity was greatly attenuated when GB1-PyMTA adduct was complexed with diamagnetic Y^{3+} ion, thus neutralizing the negatively charged tag (Fig. S3).

The interaction of GB1-PyMTA with paramagnetic lanthanide ion was investigated by monitoring the chemical shift changes upon titration with metal ion. The protein complex with Y^{3+} was used for the diamagnetic reference, since Y^{3+} shares a similar ion radius with Ho^{3+} that resides in the middle of the late lanthanide series. Addition of paramagnetic ions Tb^{3+}, Tm^{3+} or Yb^{3+} into the solution of ^{15}N-GB1-PyMTA adducts in 20 mM 2-(N-morpholino) ethanesulfonic acid (MES) buffer at pH 6.5 generated large chemical shift changes (Fig. S4 and S5). The cross-peak corresponding to the paramagnetic species increased in intensity with addition of paramagnetic lanthanide ion. Exchange between the free protein and its lanthanide bound complex is slow in ^{15}N-HSQC spectra. One paramagnetic species was observed for most residues of GB1 in complex with paramagnetic lanthanide ion, and only few residues close to the ligation site showed more than one cross-peak. Excess of lanthanide ion causes non-specific interaction between the metal ion and GB1, which manifests itself in significant PRE effects for acidic residues on the surface of the protein. However, the effect from an excess of free lanthanide can readily be removed by addition of one equivalent of EDTA, which removes the non-specific associated ions while still guaranteeing maximum lanthanide loading of PyMTA.

Most cross-peaks were clearly visible in the ^{15}N-HSQC spectra and were assigned except that for the residues with less than 12 Å distance to the paramagnetic center (Tb^{3+} and Tm^{3+}) were broadened beyond detection due to strong PRE effects caused by Curie-spin relaxation and electron-nucleus dipolar interactions.[20] PCSs of backbone amide protons were determined as the chemical shift differences between the paramagnetic and diamagnetic samples. The anisotropic magnetic susceptibility tensors (Δχ-tensors) were calculated from the experimental PCSs and the crystal structure of GB1 (PDB code:2QMT)[33] using the Numbat program.[34] The determined Δχ-tensor parameters are listed in Table S1. In general, the complex of V21C-PyMTA showed larger tensors in magnitude. These differences in Δχ-tensor magnitudes can be attributed to differences in mobility of the PyMTA tag in the two GB1-PyMTA adducts, yielding distinct paramagnetic averaging despite both T11C and V21C being located in loop segments of GB1. Compared with a ubiquitin G47C-PyMTA construct,[29] both GB1-PyMTA adducts displayed generally smaller Δχ-tensors, but were still larger than those observed with a 4-vinyl-PyMTA tag,[28] suggesting that a shorter linker facilitates more rigid attachment of the tag.

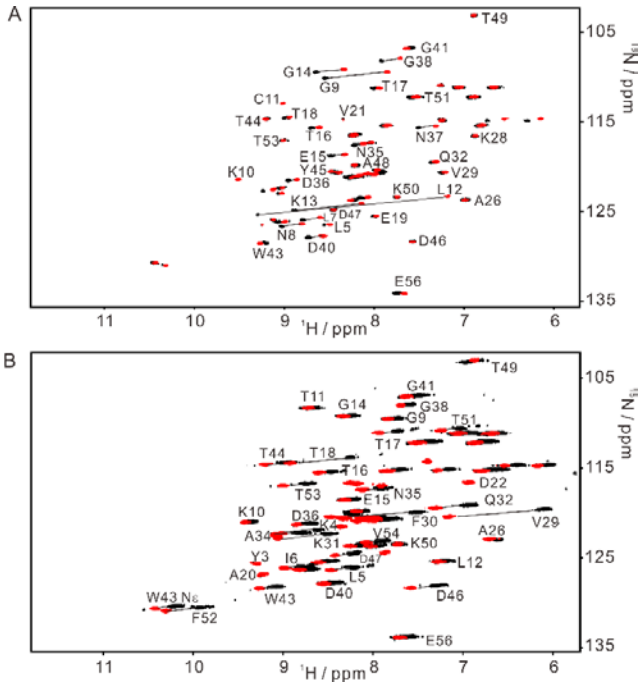


Fig. 1 Superimposition of ^{15}N-HSQC spectra of GB1-PyMTA complexed with diamagnetic Y^{3+} (red) and paramagnetic lanthanide ion (black), respectively, in living *Xenopus laevis* oocytes. A) GB1-T11C-PyMTA-Y^{3+} (red) and GB1-T11C-PyMTA-Yb^{3+} (black). B) GB1-V21C-PyMTA-Y^{3+} (red) and GB1-T11C-PyMTA-Tb^{3+} (black). The cross-peak of the same residue corresponding to the diamagnetic and paramagnetic samples were connected by solid line. The molar ratio of [Ln^{3+}]/[protein] is about 0.9. All the NMR spectra were recorded at 298K with a proton frequency of 600 MHz.

Following the established protocols of in-cell NMR,[35,36] the physiological stability of GB1-PyMTA and its lanthanide complex was subsequently investigated in living *Xenopus laevis* oocytes

(see Supplementary Information). The ^{15}N-HSQC spectrum from *Xenopus laevis* oocytes showed similar chemical shift dispersions for the complex of GB1-PyMTA and Y^{3+} but has broader NMR signals (Fig. S6). The heterogeneity observed in the in-vitro HSQC spectra was significantly reduced in the in-cell spectrum. The crowding conditions in living cells caused additional interesting NMR features. Residue 21, which displayed heterogeneity in both GB1-PyMTA adducts in the *in-vitro* ^{15}N-HSQC spectra, was visible only as a weak cross-peak in the GB1 T11C-PyMTA complex, and was entirely absent in GB1 V21C-PyMTA (Fig. S6). Chemical shift differences for these two protein constructs were measured between in-vitro and in-cell, but their magnitudes are small and suggest no significant structure changes have occurred (Fig. S7), which is in line with the previous analysis.[35]

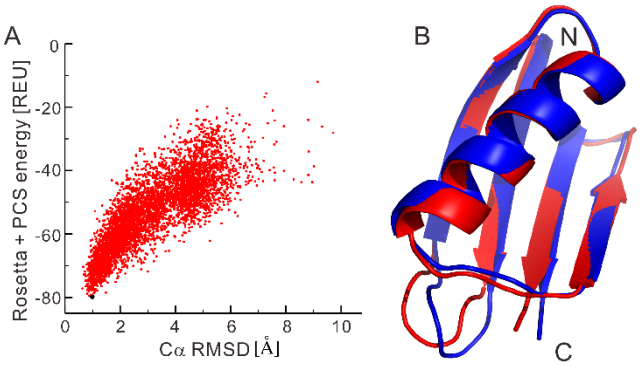


Fig. 2 High-resolution structure calculation from in-cell PCS data using GPS-Rosetta. A) Combined Rosetta and PCS energy (in Rosetta energy units; REU) is plotted against the Cα RMSD of 5000 generated model structures to the crystal structure of GB1 (PDB ID: 2QMT).[33] The structure with the lowest combined energy has a RMSD of 1.0 Å and is highlighted in black. B) Comparison of 3D representations of the structure with lowest combined energy (blue) and the crystal structure (red).

We then performed in-cell analysis of paramagnetic samples formed by GB1-PyMTA and paramagnetic ions (Supplementary Information). Fig. 1 shows the ^{15}N-HSQC spectra recorded on GB1-PyMTA complexed with diamagnetic Y^{3+} and paramagnetic lanthanide ion, respectively, in living *Xenopus laevis* oocytes. Compared with GB1-PyMTA with Y^{3+}, large chemical shift perturbations were observed for the complex of GB1-PyMTA in complex with paramagnetic lanthanide ion (Fig. S8 and S9). In-cell ^{15}N-HSQC spectra gave a similar number of observable cross-peaks compared with the in-vitro paramagnetic NMR spectra, suggesting structural restraints from PCSs can be reliably determined despite molecular crowding in the cellular environment. These results indicated that the pyridine-2-ylthio bridged protein adduct is stable in the intracellular environment. Similar to the Gd^{3+} complexes formed by protein conjugates of DOTA-derivatives and 4-vinyl-PyMTA,[19,25,26] the protein-PyMTA adduct retains its coordination with the lanthanide and is not out-competed by other natural occurring lanthanide chelators like phosphate and nucleotides, which are present in high concentrations in living cells. However, the disulfide bond linked GB1-4MMDPA[37a] adduct and a lanthanide binding peptide tag (YIDTNNDGWYEGDELLA)[37b] fused GB1 (GB1-LBT) both failed to reproduce observable PCSs in *Xenopus laevis* oocytes. This is likely due to the instability of disulfide bond in GB1-4MMDPA and a limited binding affinity for lanthanide ion in GB1-LBT for in-cell measurement (data not shown).

In-cell and in-vitro PCSs were generally of comparable quality and values (Fig. S10), suggesting the averaged orientation of PyMTA with respect to the protein frame changed little in living cells. To quantify tag poses, $\Delta\chi$-tensors were calculated by fitting the PCSs of backbone amide protons to the crystal structure of GB1 (PDB code:2QMT).[33] The differences in paramagnetic tensors shown in Table 1 may arise from the different dynamic averaging of paramagnetic tags between the in vitro and in-cell environment. The calculated paramagnetic centers from the in-vitro and in-cell data are within a distance of 2.0 Å (Fig. S11). Excellent correlations between the experimental and calculated PCSs were obtained (Figure S12 and S13) and the consistent low Q-values of less than 0.15 with all in-cell and in-vitro PCS data sets highlight the high quality of paramagnetic data that can be achieved with in-cell PCS measurements (Fig. S14). This demonstrates for the first time that PCSs can be determined accurately in living cells.

To demonstrate that the quality of PCS data obtained from living *Xenopus laevis* oocytes is sufficient for protein structure determination, we modelled the structure of GB1 using GPS-Rosetta.[31,38] The GPS-Rosetta approach is based on Rosetta's ab-initio structure calculation,[39] but additionally takes explicit advantage of PCS data from multiple labeling sites in a protein. Using fragment libraries that explicitly excluded homologs of GB1, and using a total of 201 PCSs measured for the backbone amide protons including 42 and 39 from T11C-PyMTA with Tm^{3+} and Yb^{3+}, 38, 40 and 42 from V21C-PyMTA with Tb^{3+}, Tm^{3+} and Yb^{3+}, respectively, we computed a total of 5000 structures according to the GPS-Rosetta protocol.[40] Fig. 2A shows the combined Rosetta and PCS energy plotted against the Cα RMSD (root mean squared deviation) with the crystal structure (PDB code:2QMT)[33] for all calculated structures. A pronounced energy funnel is observed, which is partly generated by the PCS score that showed a nearly linear trend towards the crystal structure. The marked drop of energy close to the crystal structure indicated that the structure calculation converged to a solution that is in good agreement with both the physical energy terms in the Rosetta force field as well as the experimental PCSs from in-cell measurements, therefore demonstrating that the in-cell structure of a protein can be reliably obtained by PCSs collected in living cells. The structure with the best combined PCS and Rosetta energy has an RMSD of 1.0 Å from the crystal structure (Fig. 2B) and the 25 lowest energy structures form a tightly clustered ensemble that deviates less than 0.15 Å (RMS) from the lowest energy structure (all compared for Cα). The low RMSD value between the crystal and in-cell structure and the excellent agreement of the structure with experimental PCS implied that the structure of GB1 is generally unchanged in the cellular environment

despite the notable structural variations for residues 8-12 in the β1/β2 loop (Fig. S15).

In summary, we have presented an efficient way to determine the structure of a protein in living cells by employing paramagnetic restraints from PCSs. PCSs are readily measured by the chemical shift differences observed in ^{15}N-HSQC spectra. The high sensitivity of the experiment allows accurate PCS data to be recorded in living cells where the limited lifetime of the cells under the condition of the NMR measurement prohibits long measurement times and/or protein concentration can be a limiting factor. Moreover, low protein concentration (~0.05 mM) was sufficient for recording ^{15}N-HSQC spectra within 2 hours.

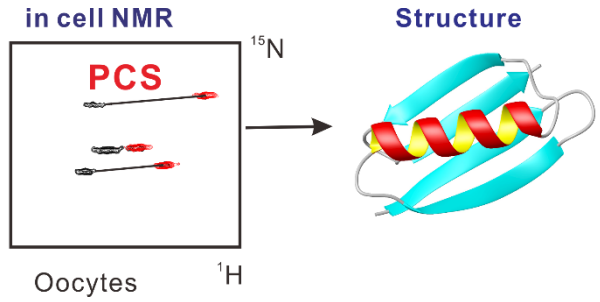
The combination of paramagnetic labeling technique, NMR spectroscopy and GPS-Rosetta is a powerful tool to characterize structure and interactions of proteins in living cells. With techniques to tag macromolecules with paramagnetic labels[41] and new advanced techniques to record NMR spectra in living cells[16,18] coming of age, paramagnetic NMR spectroscopy will become a suitable and powerful tool in the dissection of the structures, dynamics and interactions of proteins and protein-ligand complexes in living cells,

This work was supported by Major National Scientific Research Projects (2013CB910200), Natural Science Foundation of China (21473095 and 21273121) and the Australian Research Council (DP150100383).

Notes and references

‡ During revision of the manuscript, similar work has been reported by T. Müntener, D. Häussnger, P. Selenko and F. –X. Thellet in *J. Phys. Chem. Lett.*, 2016, **7**, 2821-2925.

1 Z. Serber, A. T. Keatinge-Clay, R. Ledwidge, A. E. Kelly, S. M. Miller and V. Dötsch, *J. Am. Chem. Soc.*, 2001, **123**, 2446–2447.
2 Z. Serber, W. Straub, L. Corsini, A. M. Nomura, N. Shimba, C. S. Craik, P. O. de Montellano and V. Dötsch, *J. Am. Chem. Soc.*, 2004, **126**, 7119–7125.
3 T. Sakai, H. Tochio, T. Tenno, Y. Ito, T. Kokubo, H. Hiroaki and M. Shirakawa, *J. Biomol. NMR*, 2006, **36**, 179–188.
4 C. Li, L. M. Charlton, A. Lakkavaram, C. Seagle, G. Wang, G. B. Young, J. M. Macdonald and G. J. Pielak, *J. Am. Chem. Soc.*, 2008, **130**, 6310–6311.
5 D. S. Burz, K. Dutta, D. Cowburn, A. Shekhtman, *Nat. Protoc.*, 2006, **1**, 146–152.
6 Z. Serber, P. Selenko, R. Hänsel, S. Reckel, F. Löhr, J. E. Ferrell, G. Wagner and V. Dötsch, *Nat. Protoc.*, 2006, **1**, 2701–2709.
7 D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Wälchli, B. O. Smith, M. Shirakawa, P. Güntert and Y. Ito, *Nature*, 2009, **458**, 102-105.
8 J. Hamatsu, D. O'Donovan, T. Tanaka, T. Shirai, Y. Hourai, T. Mikawa, T. Ikeya, M. Mishima, W. Boucher and B. O. Smith, *J. Am. Chem. Soc.*, 2013, **135**, 1688–1691.
9 K. Bertrand, S. Reverdatto, D. S. Burz, R. Zitomer and A. Shekhtman, *J. Am. Chem. Soc.*, 2012, **134**, 12798–12806.
10 K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hiroaki and M. Shirakawa, *Nature*, 2009, **458**, 106–109.
11 C. Li, G. Wang, Y. Wang, R. Creager-Allen, E. A. Lutz, H. Scronce, K. M. Slade, R. A. Ruf, R. A. Mehl and G. J. Pielak, *J. Am. Chem. Soc.*, 2010, **132**, 321–327.
12 Y. Ye, X. Liu, Z. Zhang, Q. Wu, B. Jiang, L. Jiang, X. Zhang, M. Liu, G. J. Pielak and C. Li, *Chem. Eur. J.*, 2013, **19**, 12705–12710.
13 L. Banci, L. Barbieri, I. Bertini, E. Luchinat, E. Secci, Y. Zhao and A. R. Aricescu, *Nat. Chem. Biol.*, 2013, **9**, 297-299.
14 E. Luchinat and L. Banci, *J. Biol. Chem.*, 2016, **291**, 3776-3784.
15 A. E. Smith, Z. Zhang, G. J. Pielak and C. Li, *Curr. Opin. Struct. Biol.*, 2015, **30**, 7-16.
16 I. C. Felli, L. Gonnelli and R. Pierattelli, *Nat. Protoc.*, 2014, **9**, 2005-2016.
17 Y. Ito and P. Selenko, *Curr. Opin. Struct. Biol.*, 2010, **20**, 640-648.
18 R. Hänsel, L. M. Luh, I. Corbeski, L. Trantirek and V. Dötsch, *Angew. Chem. Int. Ed. Engl.*, 2014, **53**, 10300-10314.
19 F. X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb and P. Selenko, *Nature*, 2016, **530**, 45-50.
20 I. Bertini, C. Luchinat and G. Parigi, *Prog. NMR Spectr.*, 2002, **40**, 249-273.
21 G. Pintacuda, M. John, X. C. Su and G. Otting, *Acc. Chem. Res.*, 2007, **40**, 206-212.
22 G. Otting, *J. Biomol. NMR*, 2008, **42**, 1-9.
23 M. A. Hass and M. Ubbink, *Curr. Opin. Struct. Biol.*, 2014, **24**, 45-53.
24 G. M. Clore and J. Iwahara, *Chem. Rev.*, 2009, **109**, 4108-4139.
25 A. Martorana, G. Bellapadrona, A. Feintuch, E. Di Gregorio, S. Aime and D. Goldfarb, *J. Am. Chem. Soc.*, 2014, **136**, 13458-13465.
26 M. Qi, A. Gross, G. Jeschke, A. Godt and M. Drescher, *J. Am. Chem. Soc.*, 2014, **136**, 15366-15378.
27 Y. Yang, Q. F. Li, C. Cao, F. Huang, X. C. Su, *Chem. Eur. J.*, 2013, **19**, 1097-1103.
28 Q. F. Li, Y. Yang, A. Maleckis, G. Otting and X. C. Su, *Chem. Commun.*, 2012, **48**, 2704-2706.
29 Y. Yang, J. T. Wang, Y. Y. Pei and X. C. Su, *Chem. Comm.*, 2015, **51**, 2824-2827.
30 A. Martorana, Y. Yang, Y. Zhao, Q. F. Li, X. C. Su and D. Goldfarb, *Dalton Trans.*, 2015, **44**, 20812-20816.
31 C. Schmitz, R. Vernon, G. Otting, D. Baker, T. Huber, *J. Mol. Biol.*, 2012, **416,** 668-677.
32 K. B. Pilla, G. Otting and T. Huber, *J. Mol. Biol.*, 2016, **428**, 522-532.
33 H. L. Schmidt, L. J. Sperling, Y. G. Gao, B. J. Wylie, J. M. Boettcher, S. R. Wilson and C. M. Rienstra, *J. Phys. Chem. B*, 2007, **111**, 14362-14369.
34 C. Schmitz, M. J. Stanton-Cook, X. C. Su, G. Otting and T. Huber, *J. Biomol. NMR*, 2008, **41**, 179–189.
35 P. Selenko, Z. Serber, B. Gadea, J. Ruderman and G. Wagner, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 11904-11909.
36 Y. Ye, X. Liu, G. Xu, M. Liu and C. Li, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 5328-5330.
37 a) X. C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B. A. Messerle and G. Otting, *J. Am. Chem. Soc.*, 2008, **130**, 10486-10487. b) J. Wöhnert, K. J. Franz, M. Nitz, B. Imperiali, H. Schwalbe, *J. Am. Chem. Soc.*, 2003, **25**, 13338-13339.
38 H. Yagi, K. B. Pilla, A. Maleckis, B. Graham, T. Huber and G. Otting, *Structure*, 2013, **21,** 883-890.
39 K. T. Simons, C. Kooperberg, E. Huang and D. Baker, *J. Mol. Biol.*, 1997, **268**, 209-225.
40 K. B. Pilla, J. K. Leman, G. Otting and T. Huber**,** *PLoS One*, 2015, **10,** e0127053.
41 (a) X. C. Su and G. Otting, *J. Biomol. NMR*, 2010, **46**, 101-112; (b) J. Koehler and J. Meiler, *Prog. NMR Spectr.*, 2011, **59**, 360-389; (c). W. M. Liu, M. Overhand and M. Ubbink, *Coord. Chem. Rev.*, 2014, **273–274**, 2–12.



The integration of site-specific labeling of proteins with a stable lanthanide binding tag, paramagnetic NMR spectroscopy and GPS-Rosetta program presents an effective and fast way of determining the three-dimensional structure of a protein in living cells.